The fluctuation behavior of heart and respiratory system signals as a quantitative tool for studying long-term environmental exposures and chronic diseases

INAUGURALDISSERTATION

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel

Von

Delphine Meier

aus Paris, Frankreich

Basel, 2017

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Marcel Tanner, Fakultätsverantwortlicher

Prof. Dr. Nicole Probst-Hensch, Dissertationsleiterin

Prof. Dr. Urs Frey, Dissertationsleiter

Prof. Dr. Philippe Cattin, Korreferent

Basel, den 17.10.2017

Dekan Prof. Dr. Martin Spiess

“Education is the most powerful weapon which you can use to change the world.”

Nelson Mandela

Table of contents

List of Abbreviations and Definitions

Abbreviation	Full term
ACE inhibitors	Angiotensin-converting-enzyme inhibitor
ACQ	Asthma Control Questionnaire
BIOAIR	Longitudinal Assessment of Clinical Course and BIOmarkers in Severe Chronic AIRway Disease
BMI	Body mass index
CD40 L	CD 40 ligand
$_{95\%}$CI	95% confidence interval
COPD	Chronic obstructive pulmonary disease
CRP	Serum C-reactive protein
DFA	Detrended fluctuation analysis
D_{LCO}	Diffusing capacity of the lung for carbon monoxide
ECG	Electrocardiogram
ETS	Environmental Tobacco Smoke
FBC	Fluctuation based clustering
FEV_1	Forced expiratory volume in one second
FeNO	Fraction of exhaled nitric oxide
FRC	Forced residual volume
FVC	Forced vital capacity
GM	Geometric means
GST	Glutathione S-transferase
HF	Power in the high frequency range
HRD	Heart rate dynamics
HRV	Heart rate variability
IVC	Inspiratory vital capacity
LF	Power in the low frequency range
MD	Missing data
MSE	Multiscale entropy
NA	Not applicable
NN intervals	Cardiac interbeat intervals = RR intervals

OCS	Oral corticosteroid
PM_{10}	Particulate matter of less than 10 micrometers in diameter
PSD	Power spectral density
QoL	Quality of life
RR intervals	Cardiac interbeat intervals = NN intervals
RV	Residual volume
SampEn	Sample entropy
SAPALDIA	Swiss Cohort Study on Air Pollution and Lung and Heart Disease in Adults
SD	Standard deviation
SDNN	Standard deviation of all NN intervals
SE	Standard error
SGRQ	St George's Respiratory Questionnaire
sRAGE	Soluble receptor for advanced glycation end products
TLC	Total lung capacity
TPM_{10}	Traffic-related particulate matter of less than 10 micrometers in diameter

Acknowledgments

I am indebted to many people who have contributed to this work in various ways and to whom I express my sincere words of thanks.

I am deeply grateful to my PhD committee Prof. Dr. Marcel Tanner, Prof. Dr. Urs Frey, Prof. Dr. Nicole Probst-Hensch, Prof. Dr. Philippe Cattin, and Dr. Edgar Delgado-Eckert for initiating this PhD thesis, guiding and supporting me along, always within a collegial atmosphere.

A very special thank you goes to Urs Frey. Your guidance has helped throughout the progress of this work, by your invaluable assistance with scientific thinking, but also through your encouragement and motivation of me, and your dynamism. I have appreciated your availability very much, and your efforts to make my PhD a productive and stimulating experience. I am also very thankful for your support as I worked to balance my PhD work with my new role as a mother.

Special words of thanks are given to Nicole Probst-Hensch and her team at the Swiss Tropical and Public Health Institute, especially to PD Dr. Christian Schindler, and Emmanuel Schaffner. Our meetings, as part of our work on the SAPALDIA cohort, were always an enjoyable time; instructive and helpful. I am very grateful that I have had the opportunity to work with the three of you, and I benefitted very much from your experience.

My sincere thanks also go to Edgar Delgado-Eckert, for opening the door to the fascinating world of computational science. Thank you for the enjoyable teamwork, our brainstorming meetings, in which we often had challenging issues to tackle, and for your support along the way.

I am also grateful to Prof. Dr. Sven-Erik Dahlen and Dr. Maciek Kupcyk for the enjoyable collaboration we had, your insightful comments, and for our stimulating meetings, always at a different place in the world!

Acknowledgments

I would like to make a special mention of lic. phil. Eva-Maria Häusler. I have appreciated your availability and efficiency very much, as well as your team spirit and all your kind attention. We started working at the Universitäts-Kinderspital beider Basel the same day, we met in the lift within our first minutes there, and from that moment on you were a pillar for me.

I am indebted to Karine Hugentobler for her very valuable English editing on manuscripts. You have such a way of communicating your joy and enthusiasm, that every moment with you was very funny and motivational.

I would also like to thank all group members with whom I have had the pleasure to work, and, in particular, thank you to my officemates, Olga, Sylvia, Ema, Binbin, and Anne.

I also must express my very profound gratitude to the friends and colleagues who have contributed to my personal life in Switzerland. I am particularly grateful for the "Mittagspartys" with Sylvia and Insa, our "Mamiesnachmittage" with Loretta, the daily chats with Aline, the Apéros, hikes, and nice meetings with Millie and Jean, the Tuesdays climbing with Sebastian and Anne, and for all the nice moments and good meals with Kai, Birgit, and Karin.

I would like to thank my family for all their love and encouragement, and for supporting me in all my pursuits. I thought fondly of my grandparents and my uncle. I would also like to thank Véronique & Laurent, Anna & Lukas & Valentin, Bettina & Michel & Jonas & Aaron, for having welcomed me so warmly into their family, for all the lovely moments together, and the traditional wonderful holidays in Mallorca.

Lastly, I would like to thank my husband, Nik, for being such an exceptional life partner, and our little boy Eliott, who brought us so much love and happiness.

Summary

Background: Several studies over the last decades have suggested that a wide range of disease states, as well as the aging process itself, are marked by progressive impairment of the involved physiological processes to adapt, resulting in a loss of complexity in the dynamics of physiological functions. Therefore, measuring complexity from physiological system signals holds enormous promise for providing a new understanding of the mechanisms underlying physiological systems and how they change with diseases and aging. Furthermore, since physiological systems are continuously exposed to environmental factors, measuring how physiological complexity changes during exposure to environmental elements might also provide new insights into their effects. Indeed, this approach may be able to unveil subtle but important changes in the regulatory mechanisms of physiological systems not detectable by traditional analysis methods.

Objectives: The overall objective of this PhD thesis was to quantify the complexity of the dynamics of heart and respiratory system signals, in order to investigate how this complexity changes with long-term environmental exposures and chronic diseases, using data from large epidemiological and clinical studies, in order to control for most potential confounders of the fluctuation behavior of systems signals (e.g., demographic, environmental, clinical, and lifestyle factors). We specifically aimed (1) at assessing the influence, first, of long-term smoking cessation, and second, of long-term exposure to traffic-related particulate matter of less than 10 micrometers in diameter (TPM_{10}), on the regulation of the autonomic cardiovascular system and heart rate dynamics in an aging general population, using data from the SAPALDIA cohort study; (2) to assess whether the subgrouping of patients with recurrent obstructive airway diseases, including mild-to-moderate asthma, severe asthma, and COPD, according to their pattern of lung function fluctuation, allows for the identification of phenotypes with specific treatable traits, using data from the BIOAIR study.

Methods: In the SAPALDIA cohort, a population-based Swiss cohort, 1608 participants ≥ 50 years of age underwent ambulatory 24-hr electrocardiogram monitoring and reported on lifestyle and medical history. In each participant, heart rate variability and heart rate dynamics were characterized by means of various quantitative analyses of the inter-beat interval time series generated from 24-hour electrocardiogram recordings. Each parameter obtained was then used as the outcome variable in multivariable linear regression models in order to evaluate the association with (1) smoking status and time elapsed since smoking cessation; (2)

long-term exposure to TPM_{10}. The models were adjusted for known confounding factors. In the BIOAIR study, we conducted a time series clustering analysis based on the fluctuation of twice-daily FEV_1 measurements recorded over a one year period in a mixed group of 134 adults with mild-to-moderate asthma, severe asthma, or COPD from the longitudinal Pan-European BIOAIR study.

Results: In the SAPALDIA cohort, our findings indicate that smoking triggers adverse changes in the regulation of the cardiovascular system, even at low levels of exposure since current light smokers exhibited significant changes as compared to lifelong non-smokers. Moreover, there was evidence for a dose-response effect. Furthermore, full recovery was achieved in former smokers (i.e., normalization to the level of lifelong non-smokers). However, while light smokers fully recovered within the 15 first years of cessation, heavy former smokers might need up to 15-25 years to fully recover. Regarding long-term exposure to TPM_{10}, we did not observe an overall association with heart rate variability/heart rate dynamics in the entire study population. However, significant changes in the heart rate dynamics were found in subjects without cardiovascular morbidity and significant changes, both in the heart rate dynamics and in the heart rate variability, were found in non-obese subjects without cardiovascular morbidity. Furthermore, subjects with homozygous GSTM1 gene deletion appeared to be more susceptible to the effects of TPM_{10}. In the BIOAIR study, we identified five phenotypes, of those three distinct phenotypes of severe asthma, in which the progressive functional alteration of the lung corresponded to a gradually increasing clinical severity and translated into specific risks of exacerbation and treatment response features.

Conclusions: This thesis hopes to demonstrate the importance of multidimensional approaches to gain understanding in the complex functioning of the human physiological system and of disease processes. Characterization of the complexity in the fluctuation behavior of system signals holds enormous promise for providing new understandings of the regulatory mechanisms of physiological systems and how they change with diseases. However, it is important to combine this kind of approach with classical epidemiological approaches in order to disentangle the various contributions of the intrinsic physiological dynamics, aging, diseases and comorbidities, lifestyle, and environment. In the SAPALDIA cohort study, we were able to disentangle the influence of specific environmental exposures, such as particulate matter air pollution and smoking exposure, on the heart rate variability and heart rate dynamics, and thus to unveil long-term alterations in former heavy smokers, as well

as adverse effects of low level, but long-term, exposure to TPM_{10} in healthy subjects and in subjects with homozygous GSTM1 gene deletion. In the BIOAIR study, we provide evidence that airway dynamics contain substantial information, which enables the identification of clinically meaningful phenotypes, in which the functional alteration of the lung translates into specific treatable traits.

List of Figures

List of Figures

List of Tables

List of Tables

1. Introduction

1.1. Complexity of the dynamics of physiological systems

Physiological systems generally exhibit complex dynamics which result from the interference, cooperation or competition of their constituent parts with one another (Schumacher 2004). These properties allow the physiological processes involved to continually adapt to extrinsic and intrinsic stimuli. Several studies over the last decades have suggested that a wide range of disease states, as well as the aging process itself, are marked by progressive impairment of these physiological processes to adapt, resulting in a loss of complexity in the dynamics of physiological functions (Lipsitz and Goldberger 1992, Goldberger 1997, Costa, Goldberger et al. 2002, Goldberger, Peng et al. 2002). Therefore, measuring complexity from physiological system signals holds enormous promise for providing a new understanding of the regulatory mechanisms of physiological systems and how they change with diseases and aging (Goldberger, Peng et al. 2002). Furthermore, since physiological systems are continuously exposed to environmental factors, measuring how physiological complexity changes during exposure to environmental elements might also provide new insights into their effects. Indeed, this approach may unveil subtle but important changes in the regulatory mechanisms of physiological systems not detectable by traditional analysis methods.

1.2. Measuring the complexity of physiological system signals

1.2.1. Nonlinear dynamic systems theory

Measuring the complexity of physiological system signals is a major contemporary challenge. This complexity arises from the interaction of a myriad of structural units and regulatory feedback loops which translate into non-random fluctuation behaviors over multiple temporal and spatial scales (Costa, Goldberger et al. 2002, Goldberger, Peng et al. 2002). As a result, dynamics of most physiological outputs are generally marked by a combination of nonstationarity and nonlinearity. A stationary process is a process whose probability distribution does not change when shifted in time. Consequently, parameters such as mean and variance do not change over time. The term nonlinear applies to systems whose components interact in a non-additive way (Goldberger, Peng et al. 2002). In other words, a nonlinear system is a system in which the change of the output is not proportional to the change of the input (Manor and Lipsitz 2013).

Consequently, in order to describe and quantify these complex dynamics, analysis techniques borrowed from the nonlinear dynamic systems theory have been applied (Costa, Goldberger et al. 2002). These techniques allow for the calculation of measures that probe different aspects of fluctuation behaviors, and which can be classified into three categories:

(1) **Fractal measures**, which assess whether signals exhibit similar kinds of fluctuations at different temporal resolutions.

(2) **Entropy measures**, which assess the regularity/irregularity or randomness of fluctuations.

(3) **Phase space methods**, which assess long-term predictability of fluctuations, as well as the overall dynamic properties of fluctuations.

No single measure is sufficient to capture the properties of complex signals. Instead, an ensemble of measures is required in order to probe signals of interest for different attributes.

1.2.2. Fractal measures

Fractal forms are composed of subunits (and sub-subunits, etc.) that resemble the structure of the overall object (**Figure 1.1** ***left***) (Goldberger, Amaral et al. 2002). This property is known as self-similarity (or scale-invariance). A number of complex anatomic structures display fractal-like geometry, such as the arterial and venous trees, the ramifying tracheobronchial tree and the His-Purkinje network. This fractal-like geometry enables a rapid and efficient transport over complex spatially-distributed systems. Analogous to scale-invariant objects that have a branching structure across multiple length scales, the fractal concept can also be applied to fluctuation across multiple time scales (**Figure 1.1** ***right*****)** (Goldberger, Amaral et al. 2002). Such processes exhibit similar kinds of fluctuations at different temporal resolutions.

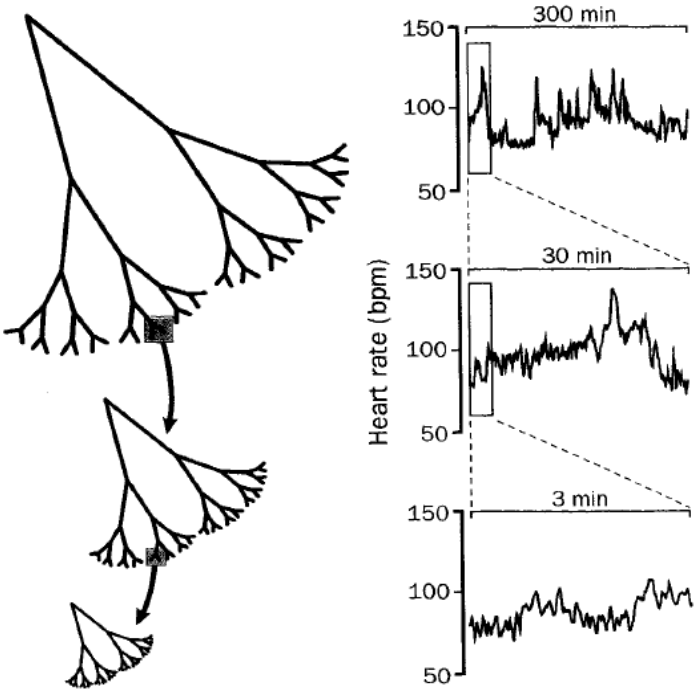

Figure 1.1. Schematic representation of self-similar structure (left) and self-similar dynamics (right). Source: Lancet. 1996 May 11;347(9011):1312-4.

Fractals scaling and related correlation properties of an on object (or of a physiologic time series) can be quantified by computing a so-called fractal dimension. A fractal dimension is a scaling rule comparing how a pattern's detail changes with the scale at which it is considered. In other words, it measures the degree of complexity by evaluating how fast the number of pieces of an object (or the number of data points of a time series) increases or decreases as the scale becomes larger or smaller (**Figure 1.2**).

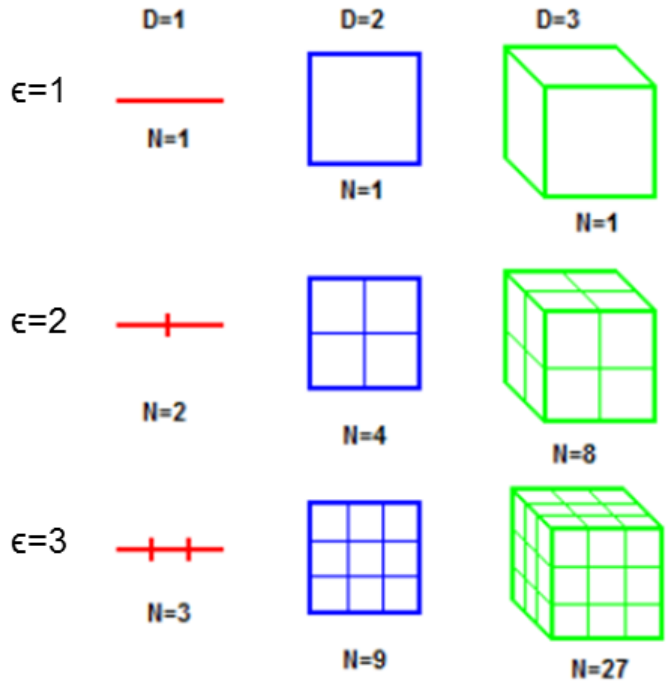

Figure 1.2. Traditional notions of geometry for defining scaling and dimension. The scaling rule or fractal dimension D is defined by the relationship $N \propto \epsilon^{-D}$, where the N stands for number of pieces, and ϵ for the scale used to get the new pieces. For instance, when scaling a filled square by 1/2, there will always be 4 new pieces, each 1/4 the area of the original, and D would be equal to 2 (e.g., $4=(1/2)^{-2}$).
Source: https://en.wikipedia.org/wiki/Fractal_dimension

The fractal dimension does not have to be an integer (**Figure 1.3**).

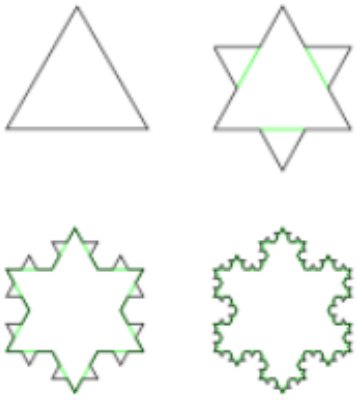

Figure 1.3. The first four iterations of the Koch snowflake, which has an approximate fractal dimension of 1.2619. Source: https://en.wikipedia.org/wiki/Fractal_dimension

In the context of biological signals, fractal analysis had to take into account the nonstationarity of signals. Thus, a specific fractal analysis methods has been introduced, the *detrended fluctuation analysis* (DFA) (Peng, Havlin et al. 1995). This method measures the presence or absence of fractal correlation properties in signals (namely the "memory effect"). The fractal long-range correlations are characterized by a scaling exponent α. A fractal-like signal results in α=1 (i.e., information-rich signal). White Gaussian noise (totally random signal) results in a value of 0.5.

1.2.3. Entropy measures

Entropy measures are measures of order/disorder. They have been used to assess the regularity/irregularity or randomness of fluctuations (Voss, Schulz et al. 2009). A typical measure of entropy is the sample entropy (SampEn). It quantifies the conditional probability that two sequences of consecutive data points that are similar to each other will remain similar when one consecutive point is included. A limitation of such measures is that order/disorder is not systematically associated with complexity. Indeed, an increase in the entropy (disorder) of a system is not necessarily always associated with an increase in its complexity (e.g., white noise). To help distinguish uncorrelated random signals from more complex (information-rich) signals, the multiscale entropy (MSE) algorithm was developed (Costa, Goldberger et al. 2005). This approach is founded on the observation that complex signals encode information over multiple time scales, whereas uncorrelated random signals or very periodic signals do not.

1.2.4. Phase space methods

Some physical or physiological systems may require several independent magnitudes in order to fully describe the state of the system. These magnitudes constitute the dimensions of a space, called the phase space. Therefore, in the phase space, all possible states of a system are represented, with each possible state of the system corresponding to one unique point in the phase space. Generally, it is not possible to measure all magnitudes that define a system, and commonly, in many scientific studies, only one magnitude/signal is measured. However, Takens' theorem shows that if a proper phase space embedding (i.e., time delay (or time lag) embedding into phase space) is performed, into a space of sufficiently high dimension, the system's dynamics can be reconstructed from a single signal (Takens 1981) (**Figure 1.4**). An example of the time delay embedding procedure is provided in **Appendix 2**.

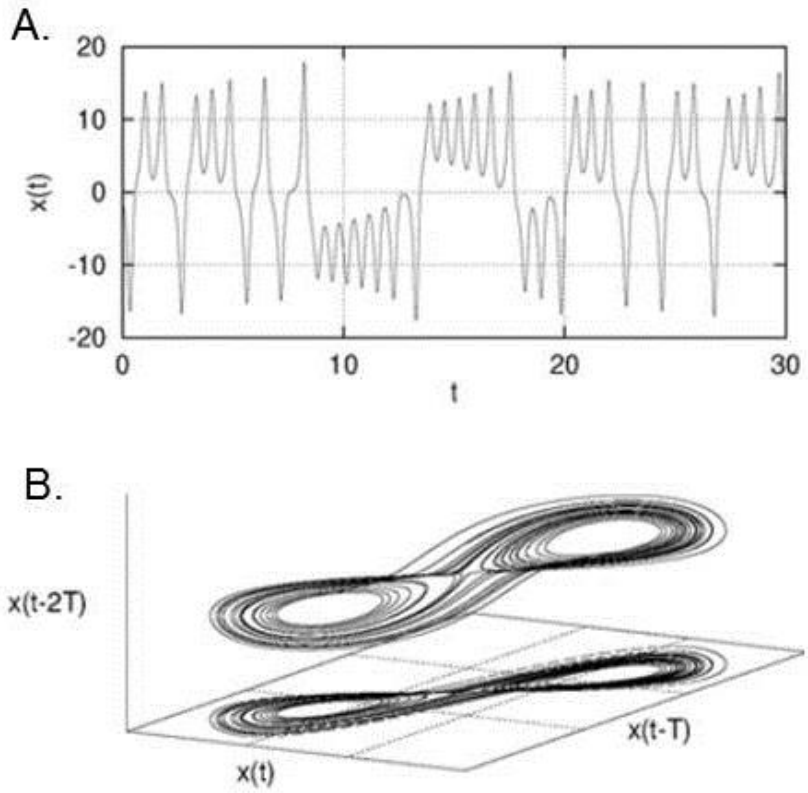

Figure 1.4. Examination of phase space embedding (A) time series formed by x coordinate, (B) 2- and 3-dimensional phase space representation. The structure or geometry of the set of system states becomes visible after embedding it into a space of proper dimension.
Source: http://www.scholarpedia.org/article/Attractor_reconstruction

1.2.4.1. Largest Lyapunov exponent

Detection of chaotic behaviour (i.e., deterministic chaos) in a time series can be done by measuring the largest Lyapunov exponent in an appropriate phase space embedding (Rosenstein, Collins et al. 1993). Deterministic chaotic systems display dynamics that appear to be random in the complete absence of randomness. They have a very sensitive dependence on initial conditions, and may be very simple, yet, in the long term, they produce completely unpredictable and rapidly divergent behaviour. Such fluctuations cannot be adequately measured with statistics based simply on mean and variance. Indeed, it is possible for two processes with very different dynamics, for example deterministic chaos and randomness, to have outputs with nearly identical means and variances (**Figures 1.5 and 1.6**) (Boeing 2016). The the largest Lyapunov exponent quantifies the exponential divergence of initially close state-space trajectories and estimates the amount of chaos in a system. The extent to which

chaos relates to physiological or pathological dynamics is a subject of active investigation and some controversy (Goldberger, Amaral et al. 2000).

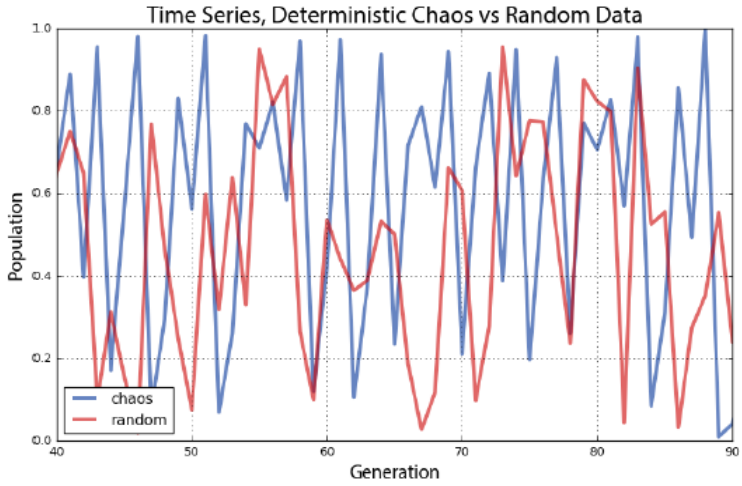

Figure 1.5. Plot of two time series, one chaotic (blue), and one random (red).
Source: Systems 2016, 4, 37; 10.3390/systems4040037

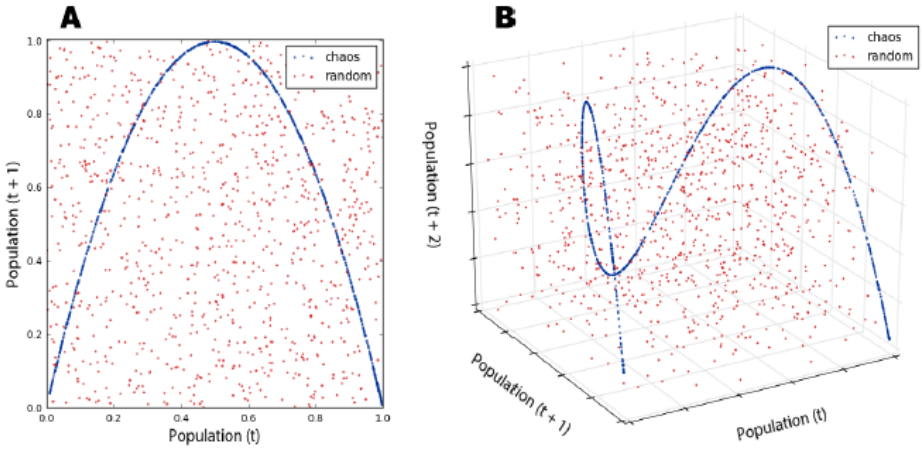

Figure 1.6. Phase space representation of the time series in Figure 3. (A) is a 2-dimensional phase space, (B) is a 3-dimensional phase space.
Source: Systems 2016, 4, 37; 10.3390/systems4040037

1.2.4.2. Poincaré plot

A specific application of the phase space representation is the Poincaré plot. Poincaré plot is widely used in the analysis of cardiac interbeat interval (RR) dynamics, where each RR interval is plotted against the next RR interval (**Figure 1.7**) (i.e., 2-dimensional phase-space representation of RR intervals). The shape of the plot provides information on the behaviour

of the system (Woo, Stevenson et al. 1992). The quantitative analysis of the shape can be done by calculating the standard deviations SD_1 (dispersion of points perpendicular to the axis of line-of-identity) and SD_2 (dispersion of points along to the axis of line-of-identity).

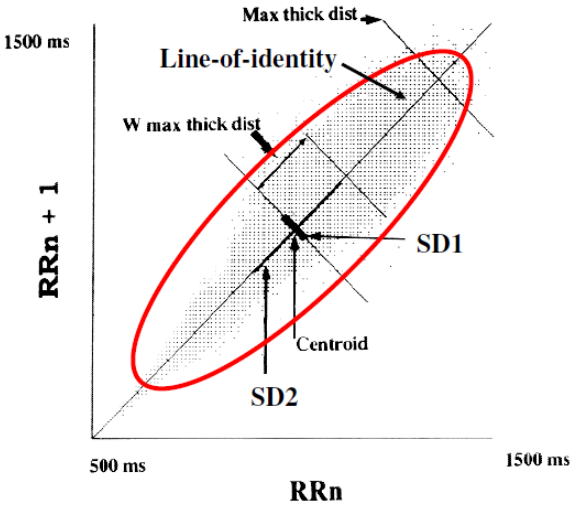

Figure 1.7. Poincaré plot of RR intervals

Source: https://www.physionet.org/events/hrv-2006/yang.pdf

1.3. Comparing the complexity between physiological systems signals

Another approach to investigate the dynamics complexity of a given physiological system is to assess the similarity between signals of different individuals. If diseases are marked by a loss of complexity in the dynamics of physiological functions, the stage or the clinical severity of a given disease might be related to the degree of loss of complexity in the dynamics of the involved physiological functions. Thus, it can be relevant to identify groups of patients with a similar (loss of) complexity, namely, a similar fluctuation behavior.

1.3.1. Time series clustering

The clustering analysis is an approach which allows for the identification of structure in an unlabeled data set by objectively organizing data into homogeneous groups where the within-group-object similarity and the between-group-object dissimilarity are maximized (Warren-Liao 2005). Time series clustering analysis allows for the consideration of dynamic behavior of the object while generating the groups.

1.3.2. Quantification of similarity between time series

Quantification of similarity between time series can be done by measuring the distance between them (Moeckel and Murray 1997). However, measuring distance between time series generated by dynamical systems requires specific metrics. For instance, a chaotic system exhibits sensitive dependence on initial conditions, so that two time series, x and y, generated by the same system, but with slightly different initial conditions, will soon diverge from one another, producing a large value of distance between x and y. Therefore, using here a traditional measure of distance, such as the Euclidean distance, would be too strict. Thus, for chaotic dynamical systems, the distance should rather be related to the attractor (small distance if similar attractor). An attractor is a set of numerical values toward which a system tends to evolve, for a wide variety of starting conditions of the dynamical system. Similarly, for stochastic processes (i.e., situations containing a random element, hence unpredictable and without a stable pattern or order; all natural events are stochastic phenomenon), distance should be related to the probability distributions (small distance if nearby distributions).

Among the different approaches developed to measure the distance for dynamical systems, the transportation distance is particularly interesting in the context of biological signals, since it is less sensitive to outliers, perturbations and discretization errors (Moeckel and Murray 1997). The transportation distance corresponds to the minimal transportation cost to move points from an initial distribution to match a final distribution. In the present work, we used a transportation distance called Earth mover's distance (or Kantorovich–Rubinstein distance) (Muskulus and Verduyn-Lunel 2011) to measure the distance between pairs of probability distribution of daily lung function measurements recorded over a predetermined window of observation (**Chapter 3**).

1.3.3. Grouping of individuals into clusters

There are five major categories of clustering methods (Han and Kamber 2001): partitioning methods, hierarchical methods, density-based methods, grid-based methods, and model-based methods. In the present work, we used an agglomerative hierarchical clustering method that groups data objects into a tree of clusters (Warren-Liao 2005) (**Chapter 3**). This method uses the Ward's minimum variance algorithm which starts by placing each subject in its own cluster, then merges the two clusters with the minimum distance (i.e., smallest increase in the value of the sum-of-squares variance), and repeats the merging process until all the subjects are merged to form one cluster.

1.3.4. Describing the clusters

The clustering approach generates several clusters and their pair-wise comparisons are affected by the multiple testing issue. Post-hoc tests for pair-wise multiple comparisons can be performed using the Tukey's test or the Nemenyi test for continuous variables, as appropriate. For categorical variables, we recommend a resampling method to address the multiple testing issue, setting the family-wise error rate at the 5% level, instead of the commonly used Bonferroni correction, which is known to be very conservative. A more detailed description of this resampling method is provided in **Chapter 3**.

1.4. Combining the analysis of the dynamics of physiological system signals with classical epidemiological approaches

The fluctuation behavior of physiological system signals is influenced by several main components: the intrinsic physiological dynamics (e.g., circadian rhythm), aging, underlying health condition (e.g., obesity, diseases), lifestyle (e.g., physical activity), and environmental factors. To disentangle effects of these components, and thus to be able to investigate the effect of a specific factor (e.g., disease process, environmental exposure), the analysis of fluctuation behavior of physiological system signals should be combined with classical epidemiological approaches in order to account for these multitude of influences. In this thesis, we exemplify this combination of both approaches in three studies, by using the fluctuation behavior of heart and respiratory system signals as a quantitative tool for studying long-term environmental exposures and chronic diseases, using data from large epidemiological and clinical studies.

1.4.1. Effect of long-term environmental exposures on heart rate variability and heart rate dynamics

Is it possible to fully recover after long-term smoking cessation?

To the best of our knowledge, the effect of long-term smoking cessation has only been investigated in terms of risk of coronary heart disease, and it is not clear when or even whether the risk of coronary heart disease reverts to that of lifelong non-smokers. We investigated, in an aging general population, whether long-term smoking cessation results in normalization of the parameters describing the heart rate variability (HRV) to the level of lifelong non-smokers, and whether this normalization is associated with the amount previously smoked (**Chapter 1**). The parameters used to describe the HRV were standard

measures of HRV (i.e., time- and frequency-domain measures of HRV), as well as parameters calculated with methods from nonlinear dynamics. While standard measures of HRV have traditionally been used, the increasing evidence that the regulation of the cardiovascular system involves nonlinear control mechanisms has encouraged the quantitative assessment of HRV using methods from nonlinear dynamics. These methods have shown new insights into HRV changes under various physiological and pathological conditions, providing additional prognostic information and complementing traditional time- and frequency-domain analyses.

Are there any adverse effects of long-term exposure to traffic-related PM_{10}?

To the best of our knowledge, effects of long-term particulate matter (PM) exposure has essentially been investigated in terms of risk of coronary heart disease, and there is limited or weak available epidemiological evidence that HRV is altered by low-level, but long-term, exposure (years). Furthermore, previous studies have provided evidences that population, such as the elderly, patients with preexisting cardiovascular disease, diabetes, obese subjects, ever smokers, females, or people with reduced antioxidative defenses might be particularly susceptible to the adverse effects of air pollution. The American Heart Association recently stated that studies on the long-term effects of air pollution on HRV and cardiovascular health are a major unresolved issue. We investigated the influence of low-level, but long-term (10 years), exposure to traffic-related particulate matter (TPM_{10}) on the regulation of the autonomic cardiovascular system and heart rate dynamics in an aging general population, as well as the a priori selected effect modifiers sex, smoking status, obesity, and gene variation in selected glutathione S-transferases (GSTs) (**Chapter 2**). In the same way as for the investigation of smoking exposure, we used standard measures of HRV, as well as parameters calculated with methods from nonlinear dynamics.

1.4.2. Investigation of lung function fluctuation behavior in chronic obstructive airway diseases for disease phenotyping purposes

Phenotyping appears especially relevant in severe asthma, COPD and the transition forms between these entities, in which the heterogeneity of response to drug therapy and the unpredictable nature of exacerbations are a major clinical challenge. For clinicians, identification of phenotypes related to specific treatable traits is of primary concern. However, to date, clustering approaches to asthma phenotyping have not enabled the identification of strong relationship between specific pathological features and particular clinical patterns or

treatment responses. The clustering approaches were mainly based on cross-sectional information related to demographic, clinical, and biological characteristics, and did not considered the fluctuation behavior of the lung function. Airway function dynamics are at the intersection between pathophysiological mechanisms and the expression of particular clinical patterns or treatment responses. Consequently, investigation of lung function fluctuation might give new insight into the relationship between specific pathological features and clinically meaningful outcomes. As part of the present work, we conducted a lung function fluctuation based clustering (FBC) analysis in a mixed group of 134 adults with mild-to-moderate asthma, severe asthma, or COPD, with a unique one-year collection of twice-daily lung function data, from the longitudinal European BIOAIR (Longitudinal Assessment of Clinical Course and BIOmarkers in Severe Chronic AIRway Disease) study (**Chapter 3**). We investigated whether the subgrouping of patients with chronic obstructive airway diseases, including mild-to-moderate asthma, severe asthma, and COPD, according to their pattern of lung function fluctuation, allows for the identification of phenotypes with specific treatable traits.

Pre-requisite for analyzing time series of lung function measurements

A common issue in a cohort study, and in a telemonitoring setting, is the handling of incomplete data sets. Within the BIOAIR study, especially, patients were asked to perform daily lung function measurements over one-year-period. Consequently, a pre-requisite of our work was the examination of missing data; whether data were missing at random, or related to a specific clinical state (e.g., exacerbation, hospitalization), and whether data imputation was necessary. The approach we used is described in **Appendix 1**.

2. Objectives

The overall objective of this PhD thesis was to quantify the complexity of the dynamics of heart and respiratory system signals, in order to investigate how this complexity changes with long-term environmental exposures and chronic diseases, using data from large epidemiological and clinical studies, in order to control for most potential confounders of the fluctuation behavior of systems signals (e.g., demographic, environmental, clinical, lifestyle factors).

We specifically aimed at:

1. Assessing the long-term influence of smoking cessation on the regulation of the autonomic cardiovascular system and on the heart rate dynamics in an aging general population, using data from the SAPALDIA cohort study:
    a. Whether long-term smoking cessation results in normalization of heart rate dynamics (as compared to lifelong non-smokers)
    b. Whether this normalization and the waiting time for it to set in are associated with the amount previously smoked

2. Evaluating the influence of low-level, but long-term (10 years), exposure to traffic-related particulate matter (TPM_{10}) on the regulation of the autonomic cardiovascular system and heart rate dynamics (HRD) in an aging general population, using data from the SAPALDIA cohort study:
    a. How is the overall TPM_{10}–HRV/HRD relationship in the entire study population?
    b. How that relationship is modified by both the underlying cardiovascular condition and the related drug treatments in subjects with cardiovascular morbidity?
    c. Is there a modification of effect by sex, smoking status, obesity, and gene variation in selected GSTs?

3. Assessing whether the subgrouping of patients with chronic obstructive airway diseases, including mild-to-moderate asthma, severe asthma, and COPD, according to their pattern of lung function fluctuation, allows for the identification of phenotypes with specific treatable traits, using data from the BIOAIR study:

a. How the fluctuation behavior of airway function dynamics varies between patients with mild-to-moderate asthma, severe asthma, and COPD
b. Whether clusters based on lung function fluctuation are related to specific pathophysiological features
c. Whether clusters based on lung function fluctuation are related to particular clinical patterns or treatment responses

3. Chapter 1: Long-term smoking cessation and heart rate dynamics in an aging healthy cohort: is it possible to fully recover?

Delphine Girard[1,2], Edgar Delgado-Eckert[1], Emmanuel Schaffner[2,3], Christoph Häcki[1], Martin Adam[2,3], Georgette L. Stern[1], Nitin Kumar[1], Denise Felber Dietrich[4], Alexander Turk[5], Marco Pons[6], Nino Künzli[2,3], Jean-Michel Gaspoz[7], Thierry Rochat[8], Christian Schindler[2,3], Nicole Probst-Hensch[2,3], Urs Frey[1]

[1]University of Basel, University Children's Hospital (UKBB), Basel, Switzerland; [2]Swiss Tropical and Public Health Institute, Basel, Switzerland; [3]University of Basel, Basel, Switzerland; [4]Federal Office for the Environment FOEN, Air Quality Management Section, Bern, Switzerland; [5]Zürcher Höhenklinik Wald, Faltigberg-Wald, Switzerland; [6]Regional Hospital of Lugano, Division of Pulmonary Medicine, Lugano, Switzerland; [7]University Hospital of Geneve, Health and Community Medicine, Geneve, Switzerland; [8]University Hospital of Geneve, Pneumology, Geneve, Switzerland

Corresponding author: Delphine Girard, University of Basel, University Children's Hospital (UKBB), Computational Physiology and Biostatistics Group, Spitalstrasse 33, 4056 Basel, Switzerland. Phone: +41 61 704 28 47, Fax: +41 61 704 12 13. E-mail address: delphine.gd@gmail.com

This article has been published

in Environmental Research (2015), **143**:39–48

Award: best paper of 2015 in Environmental Research

3.1. Abstract

Aim: To evaluate the long-term influence of smoking cessation on the regulation of the autonomic cardiovascular system in an aging general population, using the subpopulation of lifelong non-smokers as control group.

Methods: We analyzed 1481 participants aged ≥ 50 years from the SAPALDIA cohort. In each participant, heart rate variability and heart rate dynamics were characterized by means of various quantitative analyses of the inter-beat interval time series generated from 24-hour electrocardiogram recordings. Each parameter obtained was then used as the outcome variable in multivariable linear regression models in order to evaluate the association with smoking status and time elapsed since smoking cessation. The models were adjusted for known confounding factors and stratified by the time elapsed since smoking cessation.

Results: Our findings indicate that smoking triggers adverse changes in the regulation of the cardiovascular system, even at low levels of exposure since current light smokers exhibited significant changes as compared to lifelong non-smokers. Moreover, there was evidence for a dose-response effect. Indeed, the changes observed in current heavy smokers were more marked as compared to current light smokers. Furthermore, full recovery was achieved in former smokers (i.e., normalization to the level of lifelong non-smokers). However, while light smokers fully recovered within the 15 first years of cessation, heavy former smokers might need up to 15-25 years to fully recover.

Conclusion: This study supports the substantial benefits of smoking cessation, but also warns of important long-term alterations caused by heavy smoking.

Keywords: heart rate variability; nonlinear dynamics; smoking cessation; recovery of function

3.2. Introduction

The risk of coronary heart disease in current smokers is increased by a factor of 2.5 to 4 compared to lifelong non-smokers (Shaper, Pocock et al. 1985, Wannamethee, Shaper et al. 1995, Health 2004, Teo, Ounpuu et al. 2006, Shields and Wilkins 2013). Smoking cessation decreases cardiovascular morbidity and mortality and improves quality of life (Doll and Peto 1976, Novello 1990, Ockene, Kuller et al. 1990, Lightwood and Glantz 1997, Health 2004, Teo, Ounpuu et al. 2006). However, the magnitude of the risk reduction and the length of cessation required remain poorly understood. While the risk seems to decrease immediately after smoking cessation (Novello 1990, Ockene, Kuller et al. 1990, Dobson, Alexander et al. 1991, Tverdal, Thelle et al. 1993, Negri, La Vecchia et al. 1994, Wannamethee, Shaper et al. 1995, Doll, Peto et al. 2004, Teo, Ounpuu et al. 2006, Honjo, Iso et al. 2010, Mannan, Stevenson et al. 2010, Shields, Garner et al. 2013, Shields and Wilkins 2013), it is not clear when or even whether the risk reverts to that of lifelong non-smokers. While some studies have shown that the risk of coronary heart disease reverts to that of lifelong non-smokers within 3-5 years (Novello 1990, Dobson, Alexander et al. 1991, Tverdal, Thelle et al. 1993, Mannan, Stevenson et al. 2010) or within 10-20 years (Honjo, Iso et al. 2010, Shields and Wilkins 2013), other studies have identified a remaining risk in former smokers after 10 or even 20 years of continuous smoking cessation (Negri, La Vecchia et al. 1994, Wannamethee, Shaper et al. 1995, Teo, Ounpuu et al. 2006). A remaining risk was exclusively identified in former heavy, but not in former light smokers. These findings led us to the hypothesis that repeated exposure to tobacco smoke over years could trigger an irreversible change in the regulation of the autonomic cardiovascular system.

Heart rate variability (HRV) is a useful non-invasive measure to assess the autonomic regulation of cardiac rhythm (1996). Lower HRV is associated with higher cardiovascular morbidity and mortality and has proved itself as an important prognostic tool for several cardiovascular conditions (Kleiger, Miller et al. 1987, Bigger, Fleiss et al. 1992, 1996, Tsuji, Larson et al. 1996). HRV has been found to increase immediately after smoking cessation (Yotsukura, Koide et al. 1998, Minami, Ishimitsu et al. 1999, Munjal, Koval et al. 2009, Harte and Meston 2013), to reach a peak after 2 to 7 days, and to gradually decline thereafter (Harte and Meston, 2014; Lewis et al., 2010; Minami et al., 1999; Yotsukura et al., 1998). The increase in HRV persisted 1 month after smoking cessation (Stein, Rottman et al. 1996, Yotsukura, Koide et al. 1998, Harte and Meston 2013). However, the long term evolution of HRV after smoking cessation has, to our best knowledge, only been investigated by Gac et al.

(Gac and Sobieszczanska 2014). Based on a cross-sectional study including 145 hypertensive subjects the authors reported that former smokers with cessation periods of over five years had increased HRV compared to those who actively smoked cigarettes, but decreased HRV compared to those who had never smoked. Therefore, a more thorough investigation in a larger sample from the general population, and for a longer period of time, is in order.

While HRV has traditionally been measured using time- and frequency-domain measures, there is increasing evidence that the regulation of the cardiovascular system involves nonlinear control mechanisms (1996, Rajendra Acharya, Paul Joseph et al. 2006). Thus, a quantitative assessment of the inter-beat interval time series generated from 24-hour electrocardiogram recordings, using nonlinear time series analysis techniques, appears promising (Goldberger and West 1987, Meyer and Stiedl 2003, Rajendra Acharya, Paul Joseph et al. 2006, Vandeput, Verheyden et al. 2012), and may help to unveil subtle, but important changes in the heart rate dynamics (Goldberger and West 1987, Pincus 1991, Pikkujamsa, Makikallio et al. 2001). Only one pilot study has so far examined the influence of smoking cessation over a 30-day period on heart rate dynamics using multifractal analysis (Lewis, Balaji et al. 2010). Multifractality of cardiac time-series was found to be similar for smokers and non-smokers, and seemed unchanged by smoking abstinence or nicotine replacement therapy.

The objective of the present study was to evaluate the long-term influence of smoking cessation on the regulation of the autonomic cardiovascular system in an aging general population, using the subpopulation of lifelong non-smokers as control group. We investigated whether smoking cessation resulted in long-term normalization of the parameters describing the HRV and heart rate dynamics to the level of lifelong non-smokers, and whether this normalization was associated with the amount previously smoked.

3.3. Methods

3.3.1. Ethics statement

The study was approved by the central Ethics Committee of the Swiss Academy of Medical Sciences and the Cantonal Ethics Committees for each of the study areas. Each subject was informed in detail about the health examinations and signed an informed consent before any of the health examinations was conducted.

3.3.2. Study population

This study is part of the SAPALDIA (Swiss Cohort Study on Air Pollution and Lung and Heart Disease in Adults) study which was designed to assess the health effects from long-term exposure to air pollutants in the Swiss adult population. The study design has been described in detail elsewhere (Martin, Ackermann-Liebrich et al. 1997, Ackermann-Liebrich, Kuna-Dibbert et al. 2005). In brief, the SAPALDIA cohort (n=9651) was enrolled in 1991, and consisted of a random sample of the Swiss population aged 18 to 60 years, recruited from the local registries of inhabitants in eight areas featuring distinct geographical and environmental conditions.

In 2002, the follow-up study included 8047 (83.4%) participants. A random sample of 1846 out of 4417 participants aged ≥ 50 years underwent a 24-hour electrocardiogram (ECG) Holter recording to assess HRV, as previously described in detail (Felber Dietrich, Schindler et al. 2006). Exclusion criteria were general or spinal anaesthesia within 8 days before the ECG recording (n=5), a myocardial infarction within 3 months prior to the examination (n=2), taking digitalis (n=6), and an artificial internal pacemaker (n=0). Participants with recordings showing atrial fibrillation (n=12), ECG duration lower than 18 hours (n=73), or of insufficient quality (n=6), non-valid data on HRV (n=96) were also excluded (Felber Dietrich, Schindler et al. 2006). Participants who smoked pipe, cigars and/or cigarillos, but not cigarettes were excluded as well (n=38). Participants who smoked pipe, cigars and/or cigarillos in addition to cigarettes were not excluded. Finally, 127 subjects were excluded due to missing data on smoking status. Thus, the current study includes 1481 subjects.

3.3.3. Data collection

Data were collected using an electronic Case Report Form (eCRF) developed specifically for the SAPALDIA study. Information about the questionnaires and the measurements can be found in the **Online Supplement**.

3.3.4. Computational methods

Time series analysis parameters of heart rate variability were calculated for each individual time series of inter-beat intervals (RR series) generated from the 24-hour ECG recordings.

Traditional time and frequency domain measures were calculated in agreement with the standards of measurement proposed by the Task Force of the European Society of Cardiology

and the North American Society of Pacing and Electrophysiology (1996). The time domain measure used was the standard deviation of normal interbeat intervals (SDNN). For the frequency domain measures, Fast Fourier Transform procedures were used to derive the spectral distribution, which resulted in the calculation of total power, low frequency (LF) power (0.04–0.15 Hz), high frequency (HF) power (0.15–0.40 Hz), and the ratio between LF and HF (LF/HF). Moreover, we utilized the Power Spectral Density and its integral over different frequency intervals ($PSD_{1\ to\ 6}$).

The nonlinear time series analysis methods utilized to quantify and characterize the heart rate dynamics can be classified into three categories (Voss, Schulz et al. 2009): (1) **Fractal measures**, which assess heartbeat fluctuations over multiple time scales; (2) **Entropy measures**, which assess the regularity/irregularity or randomness of heartbeat fluctuations; and (3) **Phase space methods**, which assess long-term predictability of the heartbeat as well as the overall dynamic properties of the heartbeat. For the first category, we utilized Detrended Fluctuation Analysis (α). For the second category, the methods of choice were the Sample Entropy (SampEn), and the Multiscale Entropy. For the third category, we used the Largest Lyapunov Exponent, the Correlation Dimension (CD), and two standard deviation parameters derived from Poincaré Plots (SD_1 and SD_2). More details about the choice, implementation, and properties of the aforementioned time series analysis methods can be found in the **Online Supplement**.

3.3.5. Definition of smoking status

To assess the joint impact of the amount smoked and the current smoking status, the participants were classified as lifelong non-smokers (total lifetime amount smoked <0.1 pack-years), former light smokers, current light smokers, former heavy smokers, and current heavy smokers. Smokers were defined as heavy smokers if they had smoked ≥20 pack-years. The smoked pack-years were calculated by multiplying the number of years smoked by the average number of packs smoked per day.

3.3.6. Statistical analysis

Results are expressed as numbers and percentages for categorical variables and as a mean (± standard deviation) or median [25^{th}quartile;75^{th}quartile] for continuous variables, according to their distribution. Differences in distributions according to the smoking status were assessed using Chi2 tests for categorical variables, and using one-way ANOVA (if

normal distribution) or Kruskal-Wallis test (if non-normal distribution) for continuous variables.

Each parameter describing the HRV, or heart rate dynamics, was used as the outcome variable in multivariable linear regression models in order to evaluate the association with smoking status. The models were stratified by the time elapsed since cessation (0 years in current smokers, within 0 and 15 years, within 15 and 25 years, and ≥ 25 years in formers smokers). These time intervals were defined according to the literature, but also to ensure a balanced sample size of the resulting strata. Initial inspection of the outcome variable showed a skewed distribution of the residuals for the traditional time and frequency domain measures and for some of the other time series analysis parameters. These variables were therefore log-transformed. Results of these analyses are therefore presented as geometric means and percent changes in geometric means. All the models were adjusted for known confounding factors (Felber Dietrich, Schindler et al. 2006, Adam, Felber Dietrich et al. 2012). These factors were: sex (male as reference), age (for an increase of 1 year), alcohol consumption (<1 glass/day as reference, ≥ 1 glass/day), weekly physical activity – to the point of getting out of breath or sweating – (never as reference, between 0.5h and 2h, ≥ 2h/week), daily exposure to environmental tobacco smoke (for an increase of 1 hour/day), diabetes (no as reference, yes), body mass index (BMI, for an increase of 1 kg/m^2), BMI^2, average annual NO2 (for an increase of 1 $\mu g/m^3$), number of cardiovascular medications (0 as reference, 1, ≥ 2). This last variable was computed using the information on the cardiovascular medication intake (beta-blockers, angiotensin-converting-enzyme (ACE) inhibitors, angiotensin II receptor antagonists, calcium channel blockers, diuretics, antiarrhythmic drugs class I + III, sympathomimetics). The number of cardiovascular medications allowed us to summarize the information about cardiovascular medication in one variable and to gradually represent the severity of the cardiovascular disease.

The linear interaction between the total lifetime amount smoked (pack-years) and time elapsed since cessation (years) was assessed for each outcome. All the models were adjusted for the same confounders mentioned above.

Finally, we performed sensitivity analyses. First, the random effect of the study areas was included using multivariable linear mixed models. Second, for each of the parameters describing the HRV or heart rate dynamics, we excluded participants with a value lower than the 1^{th} percentile or higher than the 99^{th} percentile of the distribution of the parameter. Then, we excluded the participants taking at least one cardiovascular medication or with missing

information on the number of cardiovascular medications. Finally, since a strong interaction between air pollution and ACE inhibitors has been reported in the SAPALDIA cohort (Adam, Felber Dietrich et al. 2012), we looked for interaction between smoking status and ACE inhibitors.

All tests were two-sided with a significance level of 0.05. Statistical analysis was performed using R, Version 2.10 (2008). We used the packages Lattice (Sarkar 2008), gmodels, gplots, ggplot2 (Wickham 2009), prettyR, VIF, nlme (Pinheiro, Bates et al. 2014), lm4.

3.4. Results

3.4.1. Characteristics of the study population

The study sample consisted of 1481 subjects. The mean age of the subjects was 60.4±6.2 years. Anthropometric parameters, characteristics related to lifestyle, smoking habits, and cardiovascular health are summarized in **Table 3.1**. There were 699 (47.2%) lifelong non-smokers, 307 (20.7%) former light smokers, 57 (3.8%) current light smokers, 207 (14.0%) former heavy smokers and 211 (14.2%) current heavy smokers.

Table 3.1. Characteristics of the study population according to the smoking status

Characteristic	Lifelong non-smokers (n=699)	Former light smokers (n=307)	Current light smokers (n=57)	Former heavy smokers (n=207)	Current heavy smokers (n=211)	P-value	All (n=1481)	Missing Data
Sex, Men	236 (33.8)	139 (45.3)	15 (26.3)	162 (78.3)	126 (59.7)	<0.001	678 (45.8)	-
Age, years	61.1±6.4	59.6±6.1	58.1±5.4	61.2±6.0	59.2±5.9	<0.001	60.4±6.2	-
Lifestyle factors								
Alcohol, ≥ 1 glass/day	173 (24.8)	86 (28.1)	19 (33.3)	98 (47.3)	95 (45.0)	<0.001	471 (31.8)	1
Physical activity						0.01		13
None	336 (48.3)	139 (45.4)	28 (49.1)	95 (46.8)	127 (61.7)		725 (49.4)	
[0.5-2h[/week	212 (30.5)	84 (27.5)	15 (26.3)	60 (29.6)	41 (19.9)		412 (28.0)	
≥ 2h/week	148 (21.3)	83 (27.1)	14 (24.6)	48 (23.6)	38 (18.4)		331 (22.5)	
ETS exposure						<0.001		1
None	597 (85.5)	252 (82.1)	39 (68.4)	164 (79.2)	117 (55.5)		1169 (79.0)	
< 3h/day	73 (10.5)	33 (10.7)	14 (24.6)	24 (11.6)	54 (25.6)		198 (13.4)	
≥ 3h/day	28 (4.0)	22 (7.2)	4 (7.0)	19 (9.2)	40 (19.0)		113 (7.6)	
Smoked packyears	0.0 [0.0;0.0]	5.3 [1.5;12.0]	8.7 [3.3;14.7]	38.9 [25.4;54.0]	40.6 [31.1;56.9]	<0.001[a]	21.8 [7.0;40.7]	-
Smoking duration, years	0.0 [0.0;0.0]	10.0 [5.0;17.0]	36.1 [30.5;40.7]	28.0 [21.0;34.0]	40.1 [36.4;43.5]	<0.001[a]	27.0 [13.0;38.0]	7
Age at smoking initiation, years	-	19.0 [18.0;21.0]	20.0 [18.0;25.0]	18.0 [16.0;20.0]	18.0 [16.0;20.0]	<0.001[a]	19.0 [17.0;20.0]	-
Age at smoking cessation, years	-	30.0 [25.0;37.8]	-	46.0 [40.0;53.0]	-	<0.001[b]	37.0 [29.0;47.0]	7
Time after smoking cessation						<0.001[b]		7
0 year	699 (100)	0 (0.0)	57 (100)	0 (0.0)	211 (100)		967 (65.3)	
]0-15[years	0 (0.0)	36 (11.9)	0 (0.0)	112 (54.6)	0 (0.0)		148 (10.0)	
[15-25[years	0 (0.0)	80 (26.5)	0 (0.0)	62 (30.2)	0 (0.0)		142 (9.6)	
≥ 25 years	0 (0.0)	186 (61.6)	0 (0.0)	31 (15.1)	0 (0.0)		217 (14.7)	
Cardiovascular health and diabetes								
BMI, kg/m^2	26.4±4.5	26.3±4.3	26.4±4.4	28.6±4.2	26.2±3.9	<0.001	26.7±4.4	2

Table 3.1. Characteristics of the study population according to the smoking status (continued)

Characteristic	**Lifelong non-smokers (n=699)**	**Former light smokers (n=307)**	**Current light smokers (n=57)**	**Former heavy smokers (n=207)**	**Current heavy smokers (n=211)**	**P-value**	**All (n=1481)**	**Missing Data**
Hypertension	325 (46.5)	132 (43.0)	21 (36.8)	122 (58.9)	91 (43.1)	0.002	691 (46.7)	-
Number of cardiovascular medications						0.007		33
0	487 (71.4)	224 (74.7)	41 (73.2)	126 (62.4)	162 (77.9)		1040 (71.8)	
1	145 (21.3)	58 (19.3)	10 (17.9)	48 (23.8)	38 (18.3)		299 (20.6)	
2-3-4	50 (7.3)	18 (6.0)	5 (8.9)	28 (13.9)	8 (3.8)		109 (7.5)	
Diabetes	30 (4.3)	9 (2.9)	2 (3.5)	13 (6.3)	7 (3.3)	0.44	61 (4.1)	-
Heart rate variability								
SDNN	137.8 [117.5;161.9]	136.5 [114.3;163.4]	131.1 [112.4;148.3]	124.6 [108.1;153.2]	122.8 [103.1;144.1]	<0.001	133.6 [112.5;157.4]	-
Total power	4077 [2826;6128]	4370 [2805;6802]	3708 [2249;5244]	4043 [2714;5766]	3415 [2043;4717]	<0.001	4006 [2693;5927]	-
HF	275.4 [155.4;461.2]	283.2 [158.2;494.1]	236.7 [141.5;348.7]	248.6 [140.8;467.0]	232.0 [137.1;415.9]	0.06	269.4 [148.9;459.7]	-
LF	984.8 [606.8;1576.0]	1071.0 [635.5;1734.0]	801.7 [530.2;1432.0]	980.5 [599.3;1742.0]	858.7 [491.0;1176.0]	<0.001	697.0 [594.0;1578.0]	-
Ratio HF/LF	3.6 [2.4;5.2]	3.8 [2.5;5.4]	3.7 [2.5;5.3]	3.9 [2.4;5.5]	3.4 [2.3;5.0]	0.38	3.6 [2.4;5.3]	-

ETS, Environmental Tobacco Smoke; BMI, body mass index

Values shown are mean ± standard deviation, median [25thquartile; 75thquartile] and numbers (percentages)

Differences in distributions according to the smoking status were assessed using Chi2 tests for categorical variables, and using one-way ANOVA (if normal distribution) or Kruskal-Wallis test (if non-normal distribution) for continuous variables

[a] tested in ever smoker, [b] tested in former smoker

Lifestyle factors

All lifestyle factors differed significantly depending on the smoking status of the subjects. The current heavy smokers were the least physically active and the most exposed to environmental tobacco smoke. Both current and former heavy smokers consumed more alcohol than other groups. Former light smokers did not significantly differ from lifelong non-smokers with regard to alcohol consumption, physical activity, and environmental tobacco exposure (p=0.26, p=0.12 and p=0.10 respectively). Heavy smokers started smoking earlier than the light smokers (18 [16;20] years *vs.* 19 [18;21] years, p<0.001). Former heavy smokers ceased smoking later than the former light smokers (46.0 [40.0;53.0] years *vs.* 30.0 [25.0;37.8] years, p<0.001).

Cardiovascular health, obesity, and diabetes

The prevalence of diabetes did not significantly differ depending on smoking status (p=0.44), unlike the other factors related to cardiovascular health. Former heavy smokers exhibited a higher BMI, higher prevalence of hypertension, and were undergoing cardiovascular treatment more frequently. The other groups did not differ regarding these factors (p=0.92, p=0.41, p=0.47 respectively for BMI, hypertension and number of cardiovascular medication).

3.4.2. Exploration of the association between current smoking and heart rate dynamics

Using standard parameters of HRV

Table 3.2 shows the associations between the smoking status and the time- and frequency-domain measures of HRV in multivariable analysis, stratified by the time elapsed since cessation. Irrespective of smoking intensity, current smokers showed a significantly decreased HRV for SDNN, total power and LF. The HF and ratio LF/HF was significantly decreased only in the current heavy smokers. Moreover, there is evidence for a dose-response effect, given that SDNN, TP and LF were more markedly decreased in current heavy smokers as compared to current light smokers.

Table 3.2. Association between smoking status and time-domain and frequency-domain measures of HRV in multivariable analysis, stratified by time elapsed since cessation (n=1420 due to missing data on co-variables)

	SDNN %GM, $_{95\%}$CI	p-value	**Total power** %GM, $_{95\%}$CI	p-value	**HF** %GM, $_{95\%}$CI	p-value	**LF** %GM, $_{95\%}$CI	p-value	**Ratio LF/HF** %GM, $_{95\%}$CI	p-value
Time after cessation: 0 year										
Smoking status (ref.=Lifelong non-smoker)		*<0.001*		*<0.001*		*0.04*		*<0.001*		*0.005*
Current light smoker	-7.3% [-13.6;-0.6]	0.03	-21.7% [-33.5;-7.9]	0.003	-19.2% [-35.6;1.5]	0.07	-20.1% [-33.1;-4.6]	0.01	-1.1% [-15.0;15.1]	0.88
Current heavy smoker	-11.7% [-15.4;-7.9]	<0.001	-27.1% [-34.0;-19.5]	<0.001	-12.9% [-24.2;-0.01]	0.05	-25.3% [-32.9;-16.7]	<0.001	-14.2% [-21.8;-5.9]	0.001
Time after cessation:]0-15[years										
Smoking status (ref.=Lifelong non-smoker)		*0.009*		*0.06*		*0.99*		*0.11*		*0.04*
Former light smoker	-2.2% [-10.2;6.4]	0.60	-11.2% [-27.3;8.3]	0.24	1.2% [-23.5;33.9]	0.93	-19.0% [-34.8;0.7]	0.06	-20.0% [-34.0;-3.0]	0.02
Former heavy smoker	-8.0% [-12.9;-2.9]	0.003	-13.0% [-23.5;-1.2]	0.03	0.8% [-15.8;20.7]	0.93	-6.7% [-18.9;7.3]	0.33	-7.4% [-18.2;4.7]	0.22
Time after cessation: [15-25[years										
Smoking status (ref.=Lifelong non-smoker)		*0.43*		*0.19*		*0.25*		*0.17*		*0.93*
Former light smoker	2.6% [-3.2;8.7]	0.39	10.4% [-3.8;26.7]	0.16	15.8% [-4.4;40.4]	0.13	14.6% [-1.5;33.4]	0.08	-1.0% [-13.1;12.7]	0.87
Former heavy smoker	-3.1% [-9.5;3.9]	0.38	-8.2% [-22.1;8.0]	0.30	-6.5% [-25.5;17.4]	0.56	-4.0% [-19.8;14.9]	0.66	2.6% [-12.0;19.7]	0.74
Time after cessation: ≥ 25 years										
Smoking status (ref.=Lifelong non-smoker)		*0.16*		*0.53*		*0.64*		*0.90*		*0.42*
Former light smoker	-4.0% [-7.9;0.1]	0.06	-5.3% [-14.1;4.4]	0.28	-5.5% [-17.5;8.1]	0.41	-1.8% [-11.7;9.3]	0.74	4.0% [-5.1;13.9]	0.40
Former heavy smoker	0.07% [-8.9;10.0]	0.99	-4.3% [-23.2;19.2]	0.70	5.2% [-22.4;42.4]	0.75	-4.3% [-24.7;21.8]	0.72	-9.0% [-25.9;11.8]	0.37

All the models are adjusted for gender, age, ETS exposure, alcohol consumption, physical activity, diabetes, BMI, BMI squared, number of cardiovascular medication, average annual NO_2

SDNN, standard deviation of all NN intervals; HF, power in the high frequency range; LF, power in the low frequency range

Values shown are as percent changes in geometric means (GM) and 95% confidence interval ($_{95\%}$CI)

Participants were classified as never smokers if the total lifetime amount smoked was <0.1 pack-years. Smokers were defined as heavy smokers if the total lifetime amount smoked was ≥ 20 pack-years. Pack-years were calculated by multiplying the number of years smoked by the average number of packs smoked per day

Using non-standard parameters

Table 3.3 shows the associations between smoking status and the non-standard parameters in multivariable analysis, stratified by the time elapsed since cessation. The first category includes the parameters PSD_5, exponent α short-term time scale (α_3), Multiscale entropy low and Largest Lyapunov Exponent, which reflected significant changes in current heavy smokers. Compared to lifelong non-smokers, PSD_5 was significantly increased in the current heavy smokers (0.2±0.08, p=0.02) and α short-term time scale, Multiscale entropy low and Lyapunov Largest Exponent were significantly decreased (-0.1±0.03, p<0.001, -0.02±0.003, p<0.001 and -0.04±0.006, p<0.001 respectively). The second category includes the parameters SD_1 and SD_2 derived from the Poincaré Plot, which were significantly decreased both in light and heavy current smokers compared to lifelong non-smokers.

The parameters α long-term time scale (α_4), PSD_2, Multiscale entropy high and $SampEn_1$ did not detect changes in the regulation of the cardiovascular system as a response to current tobacco smoke exposure (**Online Supplement**).

3.4.3. Exploration of the association between long-term smoking cessation and heart rate dynamics

Using standard parameters of HRV

SDNN, total power and LF showed a full recovery (i.e., normalization to the level of the lifelong non-smokers) in former light smokers within the first 15 years of cessation (**Table 3.2**).

While HF, LF and the ratio LH/HF also showed a full recovery in former heavy smokers within the first 15 years of cessation, SDNN and total power remained significantly decreased, and the normalization to the level of lifelong non-smokers appeared in the group of subjects who had ceased smoking 15-25 years prior. Finally, we found a significant interaction between the total lifetime amount smoked (pack-years) and the time elapsed since cessation (years) for SDNN, total power, and LF (**Online Supplement**) which provides evidence that the former smokers recovered differently according to the number of packyears they had smoked, as suggested by the later results.

Table 3.3. Association between smoking status and non-standard parameters in multivariable analysis, stratified by time elapsed since cessation (n=1420 due to missing data on co-variables)

	Category 1								**Category 2**			
	PSD_5		**α short-term time scale**		**Multiscale entropy low**		**Largest Lyapunov exponent**		**Poincaré SD_1**		**Poincaré SD_2**	
	coefficient±se	p-value	coefficient±se	p-value	coefficient±se	p-value	coefficient±se	p-value	%GM, $_{95\%}$CI	p-value	coefficient± se	p-value
Time after cessation: 0 year												
Intercept	-3.3±0.8		3.0±0.3		0.1±0.03		0.4±0.07				258.9±42.7	
Smoking status (ref.=Lifelong non-smoker)		0.03		<0.001		<0.001		<0.001		0.004		<0.001
Current light smoker	-0.1±0.1	0.37	-0.03±0.05	0.48	-0.0004±0.005	0.93	0.01±0.01	0.33	-14.8% [-23.8;-4.8]	0.005	-13.7±6.7	0.04
Current heavy smoker	0.2±0.08	0.02	-0.1±0.03	<0.001	-0.02±0.003	<0.001	-0.04±0.006	<0.001	-7.1% [-13.3;-0.6]	0.03	-23.2±4.1	<0.001
Time after cessation:]0;15[years												
Intercept	-3.8±0.8		3.0±0.3		0.1±0.03		0.5±0.07				281.6±43.9	
Smoking status (ref.=Lifelong non-smoker)		0.19		0.14		0.03		0.06		0.87		0.008
Former light smoker	-0.08±0.2	0.63	-0.07±0.06	0.26	-0.01±0.006	0.06	-0.02±0.01	0.17	-0.06% [-13.0;14.8]	0.99	-3.5±8.3	0.67
Former heavy smoker	0.2±0.1	0.09	-0.06±0.04	0.09	-0.009±0.004	0.04	-0.02±0.008	0.05	-2.3% [-10.6;6.8]	0.60	-16.6±5.3	0.002
Time after cessation: [15;25[years												
Intercept	-4.8±0.8		2.8±0.3		0.1±0.03		0.4±0.07				287.5±44.1	
Smoking status (ref.=Lifelong non-smoker)		0.20		0.64		0.76		0.07		0.35		0.25
Former light smoker	-0.04±0.1	0.70	-0.03±0.04	0.48	-0.003±0.005	0.57	0.001±0.009	0.91	7.2% [-2.7;18.1]	0.16	6.2±5.8	0.29
Former heavy smoker	0.2±0.1	0.09	0.03±0.05	0.58	-0.003±0.005	0.62	-0.02±0.01	0.02	-0.4% [-11.2;11.7]	0.94	-8.0±6.8	0.24
Time after cessation: ≥ 25 years												
Intercept	-3.3±0.8		3.0±0.3		0.1±0.03		0.4±0.07				280.4±42.7	
Smoking status (ref.=Lifelong non-smoker)		0.92		0.90		0.90		0.22		0.63		0.23
Former light smoker	-0.0005±0.07	0.99	0.01±0.03	0.70	-0.0001±0.003	0.96	-0.008±0.006	0.21	-3.2% [-9.5;3.5]	0.34	-7.0±4.1	0.09
Former heavy smoker	-0.07±0.2	0.68	0.02±0.06	0.79	-0.003±0.007	0.65	-0.02±0.01	0.19	0.1% [-13.9;16.5]	0.99	0.4±9.3	0.97

All the models are adjusted for gender, age, ETS exposure, alcohol consumption, physical activity, diabetes, BMI, BMI squared, number of cardiovascular medication, average annual NO2

PSD, Power Spectral Density

Values shown are as percent changes in geometric means (GM) and 95% confidence interval ($_{95\%}$CI) or coefficient ± standard error (se)

Participants were classified as never smokers if the total lifetime amount smoked was <0.1 pack-years. Smokers were defined as heavy smokers if the total lifetime amount smoked was ≥ 20 pack-years. Pack-years were calculated by multiplying the number of years smoked by the average number of packs smoked per day

Using non-standard parameters

In former light smokers, according to Poincaré parameters SD_1 an SD_2, a full recovery appeared within the first 15 years of smoking cessation.

All parameters showed a full recovery in former heavy smokers as well. However, while this recovery appeared within the first 15 years of cessation for the PSD_5, α short-term time scale and Poincaré SD_1, it appeared later for the Multiscale entropy low, the Largest Lyapunov Exponent and the Poincaré SD_2. Indeed, the normalization to the level of lifelong non-smokers appeared in the group of subjects who had ceased smoking 15-25 years prior for Multiscale entropy low and Poincaré SD_2, and after 25 years of cessation for the Largest Lyapunov Exponent.

Finally, as well as with the standard parameters of HRV, we found a significant positive interaction between the total lifetime amount smoked (pack-years) and the time elapsed since cessation (years) for α short-term time scale, Multiscale entropy low, and Poincaré SD_2 (**Online Supplement**).

3.4.4. Sensitivity analyses

Exclusion of participants taking at least one cardiovascular medication or with missing information on the number of cardiovascular medications showed that LF normalized to the level of lifelong non-smokers between 15 to 25 years of smoking cessation in former heavy smokers (5.5% [-15.4%;31.6%], p=0.63). The other parameters showed similar results when compared to the analysis of the whole cohort (**Online Supplement**).

Exclusion of the outliers for each outcome variable showed (i) that former heavy smokers needed up to 15 to 25 years to fully recover, when the HRV was assessed by the ratio LF/HF, and (ii) that the Largest Lyapunov Exponent normalized to the level of lifelong non-smokers between 15 and 25 years of cessation in former heavy smokers (-0.004±0.009, p=0.07). Thus, the remaining decrease in former heavy smokers after 15-25 years of cessation in the analysis of the whole cohort might be due to outliers (**Online Supplement**).

After exclusion of both outliers and participants taking at least one cardiovascular medication, or with missing information on the number of cardiovascular medications, our results suggested a full normalization of the Largest Lyapunov Exponent within the first 15 years of smoking cessation (-0.01±0.009, p=0.17). There were no changes in the association between the standard parameters of HRV and smoking status (**Online Supplement**).

3.5. Discussion

3.5.1. Main results

This study evaluates the long-term influence of smoking cessation on heart rate dynamics using standard parameters, Power Spectral Density parameters, and nonlinear time series analysis parameters. Our findings provide evidence supporting the following conclusions:

1. Smoking triggers adverse changes in the regulation of the cardiovascular system, even at low levels of exposure. Indeed, the SDNN, total power, LF, and Poincaré SD_1 and SD_2 parameters were decreased in light and heavy current smokers compared to lifelong non-smokers. Moreover, there is evidence for a dose-response effect.
2. After cessation, light smokers fully recover within the first 15 years of cessation. Indeed, both standard and non-standard parameters normalized to the levels characteristic of lifelong non-smokers.
3. After cessation, heavy smokers may fully recover as well. However, according to SDNN, total power, Multiscale entropy low and Poincaré SD_2 the normalization might need up to 15 to 25 years.

Our findings suggested a full normalization of the Lyapunov Largest Exponent only after 25 years of cessation in former heavy smokers. This supports the hypothesis that nonlinear time series analysis techniques may be able to unveil subtle, but important changes in the regulation of the cardiovascular system more difficult to detect by traditional analysis methods. However, our sensitivity analysis suggested that this remaining significant change between 15 and 25 years of smoking cessation might be due to outliers in the data.

We also noticed that the specific settings used for the calculation of the nonlinear time series analysis parameters, e.g., the embedding dimension (**Online Supplement**), play a role as to whether certain effects are detected or not. This suggests that tobacco smoke exposure may trigger very specific alterations which might be better described using non-standard parameters calculated using certain settings. This property of the non-standard parameters makes them potentially suitable tools for exploring the mechanisms underlying the modifications in the regulation of the cardiovascular system triggered by tobacco smoke exposure.

Finally, we found a significant interaction between the total lifetime amount smoked and the time elapsed since cessation for standard and non-standard parameters. The fact that this interaction is positive means that the slope of the regression line relating HRV to cessation time is steeper in heavier than in lighter smokers. This suggests that the speed of recovery is

faster in former heavy smokers. Nevertheless, it takes longer for them to fully recover because they start out from a lower initial value.

3.5.2. Strength and weaknesses of the study (internal validity)

To the authors' knowledge, this is the first study examining the influence of long-term smoking cessation on parameters describing the HRV and heart rate dynamics. Additional strengths of the present study included a) the population-based design, involving a random sample of the Swiss population; b) the large number of participants; b) the detailed information available on participants, allowing for control of most potential confounders; c) the two assessment points, allowing a better understanding of the smoking history of the participants; d) the calculation, using long-duration ECG recordings, of a variety of HRV parameters, both in the time and frequency domain, and nonlinear time series analysis parameters as well; e) the control group of lifelong non-smokers, allowing to assess whether there was a full recovery in former smokers.

This study faced some limitations. First, smoking status was assessed using self-reported data rather than by means of measurements of biomarkers. However, CO measures were used to validate smoking status, which has been shown, in combination with self-reporting, to discriminate well between smokers and non-smokers (Stevens and Munoz 2004, Felber Dietrich, Schwartz et al. 2007), and to be highly concordant with biomarker measurements (Patrick, Cheadle et al. 1994). A second limitation was the absence of information on the use of nicotine substitution therapy after smoking cessation. Previous studies have shown that abstinence from smoking in combination with the use of nicotine transdermal patches results in increased HRV, which further increases after the cessation of substitution therapy (Stein, Rottman et al. 1996, Harte and Meston 2013). Therefore, nicotine substitutes might prevent former smokers from recovering as quickly as they normally would without the use of substitutes. As a consequence, we cannot rule out that some of the effect attributed to time since quitting may be confounded by an effect of using nicotine replacement and we might have underestimated the improvement in former smokers. Furthermore, no information about the type of cigarettes smoked was available within this cohort. Nevertheless, the subjects in this study did not use electronic cigarettes since they were introduced into the market in 2004. Finally, given the small number of participants who had only quit for a short time, we were not able to assess the short-term influence of smoking cessation.

3.5.3. Strengths and weaknesses of the study compared to other studies (external validity)

Our findings are in the line with the well-established view that smokers, compared to non-smokers, exhibit dysfunctional cardiac autonomic function, as evidenced by lower HRV indices, even at low levels of exposure (Raupach, Schafer et al. 2006, Dinas, Koutedakis et al. 2013, Harte and Meston 2013).

Moreover, using parameters describing HRV, or heart rate dynamics, we have given evidence that long-term smoking cessation allows for a full recovery within 15 years for former light smokers, and up to 15-25 years for former heavy smokers. We have only found one study assessing the long-term influence of smoking cessation on HRV (Gac and Sobieszczanska 2014). Using time-domain measures of HRV, measured in 145 patients with hypertension, Gac et al. have found a decreased HRV in current smokers and a partial recovery in former smokers. In former smokers, the mean number of cigarettes/24h smoked in the past was 16.2±6.5 and the mean time after cessation was 10.8±3.6 years. Therefore, heavy smokers who lacked sufficient time post cessation to fully recover may account for the remaining decrease in HRV. Furthermore, the long-term effect of smoking cessation has been extensively studied with respect to the risk of cardiovascular morbidity and mortality (Doll and Peto 1976, Novello 1990, Ockene, Kuller et al. 1990, Lightwood and Glantz 1997, Teo, Ounpuu et al. 2006). To the extent that we may translate the increase in HRV to a decrease in the risk of coronary heart disease, our findings are consistent with this large body of literature, which has demonstrated the substantial benefits of quitting cigarette use for the cardiovascular system, irrespective of the amount smoked (Novello 1990, Ockene, Kuller et al. 1990, Dobson, Alexander et al. 1991, Tverdal, Thelle et al. 1993, Negri, La Vecchia et al. 1994, Wannamethee, Shaper et al. 1995, Doll, Peto et al. 2004, Health 2004, Teo, Ounpuu et al. 2006, Honjo, Iso et al. 2010, Mannan, Stevenson et al. 2010, Shields, Garner et al. 2013, Shields and Wilkins 2013). Our findings are, in particular, highly consistent with the studies that stratified the analyses by light and heavy former smokers, since they have shown that (a) the risk in former light smokers was similar within 3 years after quitting to those who had never smoked (Rosenberg, Palmer et al. 1990, Negri, La Vecchia et al. 1994, Wannamethee, Shaper et al. 1995, Teo, Ounpuu et al. 2006); (b) there was a remaining risk in former heavy smokers after 10 years of smoking cessation (Rosenberg, Palmer et al. 1990, Negri, La Vecchia et al. 1994, Wannamethee, Shaper et al. 1995). Only Teo et al. has identified a still

increased risk of acute myocardial infarction in former heavy smokers after 20 years of smoking cessation (Teo, Ounpuu et al. 2006).

3.5.4. Relevance of the study results and implications for policymakers

First, this study provides evidence that smoking triggers adverse changes in the regulation of the cardiovascular system, even at low levels of exposure. This constitutes a strong argument for health policy makers advocating for more intensive prevention campaigns aimed at discouraging smoking.

Secondly, we show that long-term smoking cessation leads to a normalization of heart rate dynamics, therefore reducing the risk of developing cardiovascular disease later in life. This underpins the value of public healthcare programs supporting the benefits of smoking cessation.

However, while light smokers fully recovered within the 15 first years of cessation, heavy former smokers might need up to 15-25 years to fully recover. Thus, former heavy smokers remain exposed longer after cessation to a higher risk of cardiovascular morbidity and cardiovascular-related morbidity. In analogy to the recommendations of the American Cancer Society (2014) related to lung cancer screening, our data suggest that future studies need to demonstrate whether close monitoring of cardiovascular disease in heavy smokers, current and/or former, should be recommended as well.

3.6. Conclusion

In conclusion, findings of the present study indicate that smoking triggers changes in the cardiac autonomic function even at low levels of exposure. Moreover, there is evidence for a dose-response effect. Furthermore, our findings indicate that long-term smoking cessation allows for a full recovery within 15 years in former light smokers, and possibly within 15-25 years in former heavy smokers. Therefore, our study supports the substantial benefits of smoking cessation, but also warns of important alterations caused by heavy smoking.

3.7. Acknowledgements

Current SAPALDIA Team:

Study directorate: N.M. Probst-Hensch (PI; e/g); T. Rochat (p), C. Schindler (s), N. Künzli (e/exp), J.M. Gaspoz (c)

Scientific team: J.C. Barthélémy (c), W. Berger (g), R. Bettschart (p), A. Bircher (a), C. Brombach (n), P.O. Bridevaux (p), L. Burdet (p), D. Felber Dietrich (e), M. Frey (p), U. Frey (pd), M.W. Gerbase (p), D. Gold (e), E. de Groot (c), W. Karrer (p), F. Kronenberg (g), B. Martin (pa), A. Mehta (e), D. Miedinger (o), M. Pons (p), F. Roche (c), T. Rothe (p), P. Schmid-Grendelmeyer (a), D. Stolz (p), A. Schmidt-Trucksäss (pa), J. Schwartz (e), A. Turk (p), A. von Eckardstein (cc), E. Zemp Stutz (e).

Scientific team at coordinating centers: M. Adam (e), I. Aguilera (exp), S. Brunner (s), D. Carballo (c), S. Caviezel (pa), I. Curjuric (e), A. Di Pascale (s), J. Dratva (e), R. Ducret (s), E. Dupuis Lozeron (s), M. Eeftens (exp), I. Eze (e), E. Fischer (g), M. Foraster (e), M. Germond (s), L. Grize (s), S. Hansen (e), A. Hensel (s), M. Imboden (g), A. Ineichen (exp), A. Jeong (g), D. Keidel (s), A. Kumar (g), N. Maire (s), A. Mehta (e), R. Meier (exp), E. Schaffner (s), T. Schikowski (e), M. Tsai (exp)

(a) allergology, (c) cardiology, (cc) clinical chemistry, (e) epidemiology, (exp) exposure, (g) genetic and molecular biology, (m) meteorology, (n) nutrition, (o) occupational health, (p) pneumology, (pa) physical activity, (pd) pediatrics, (s) statistics

The study could not have been done without the help of the study participants, technical and administrative support and the medical teams and field workers at the local study sites. **Local fieldworkers:** Aarau: S. Brun, G. Giger, M. Sperisen, M. Stahel, Basel: C. Bürli, C. Dahler, N. Oertli, I. Harreh, F. Karrer, G. Novicic, N. Wyttenbacher, Davos: A. Saner, P. Senn, R. Winzeler, Geneva: F. Bonfils, B. Blicharz, C. Landolt, J. Rochat, Lugano: S. Boccia, E. Gehrig, M.T. Mandia, G. Solari, B. Viscardi, Montana: A.P. Bieri, C. Darioly, M. Maire, Payerne: F. Ding, P. Danieli A. Vonnez, Wald: D. Bodmer, E. Hochstrasser, R. Kunz, C. Meier, J. Rakic, U. Schafroth, A. Walder.

Administrative staff: N. Bauer Ott, C. Gabriel, R. Gutknecht.

Research funding:

The Swiss National Science Foundation (grants no 33CS30-148470/1, 33CSCO-134276/1, 33CSCO-108796, 3247BO-104283, 3247BO-104288, 3247BO-104284, 3247-065896, 3100-

059302, 3200-052720, 3200-042532, 4026-028099, PMPDP3_129021/1, PMPDP3_141671/1), the Federal Office of Public Health (Tobacco Control Fund: 07.00577 and REF-1014-50999), the Federal Office for the Environment, the Federal Office of Roads and Transport, the canton's government of Aargau, Basel-Stadt, Basel-Land, Geneva, Luzern, Ticino, Valais, and Zürich, the Swiss Lung League, the canton's Lung League of Basel Stadt/ Basel Landschaft, Geneva, Ticino, Valais, Graubünden and Zurich, Stiftung ehemals Bündner Heilstätten, SUVA, Freiwillige Akademische Gesellschaft, UBS Wealth Foundation, Talecris Biotherapeutics GmbH, Abbott Diagnostics, European Commission 018996 (GABRIEL), Wellcome Trust WT 084703MA.

We thank K. Hugentobler for language editing.

3.8. Online Supplement

Methods

Data collection

SAPALDIA questionnaire

In 1991, participants were first interviewed by a trained fieldworker using a standardised questionnaire. The questionnaire was developed along with the questionnaire of the European Community Respiratory Health Survey (Burney, Luczynska et al. 1994). The different sections of the questionnaire related to history of respiratory symptoms, allergic diseases, living and working environment, exposure to animals, smoking and general health. Additional questions concerning smoking habits and environmental tobacco smoke exposure were adopted from the MONICA questionnaires (Martin, Ackermann-Liebrich et al. 1997). In SAPALDIA 2, the follow-up study, the questionnaire was extended with additional questions about chronic diseases, including heart disease, physical activity (derived from the ECRHS II and the Questionnaire of the Swiss Health Survey), and present and past medication use was recorded in detail (Ackermann-Liebrich, Kuna-Dibbert et al. 2005).

Measurements

Details about environmental measurements (e.g., NO_2, PM_{10}) and biological measurements (e.g., blood pressure, weight, height) are reported elsewhere(Martin, Ackermann-Liebrich et al. 1997, Ackermann-Liebrich, Kuna-Dibbert et al. 2005). Recording of 24 hours ECG have been previously described (Felber Dietrich, Schindler et al. 2006). Interbeat interval time series were obtained from the raw ECG Holter recordings via QRS-complex recognition using the software Impresario, Version 3 (Del Mar Reynolds Medical, Inc. Irvine, CA, USA).

Potentially abnormal or ectopic beats, when recognized by the software, were marked as putative artifacts but not removed from the resulting time series. Our algorithms used for the calculation of time series analysis parameters as well as of measures of HRV (see Computational Methods below) make use of these marks in order to avoid the incorporation of possibly faulty values into the computations.

Computational methods

Traditional time and frequency domain measures were calculated in agreement with the standards of measurement proposed by the Task Force of the European Society of Cardiology

and the North American Society of Pacing and Electrophysiology (1996). The time domain measure used was the standard deviation of normal interbeat intervals (SDNN). For the frequency domain measures, Fast Fourier Transform procedures were used to derive the spectral distribution, which resulted in the calculation of total power, low frequency (LF) power (0.04–0.15 Hz), high frequency (HF) power (0.15–0.40 Hz), and the ratio between LF and HF (LF/HF).

The following time series analysis parameters were calculated using our own implementations in R (occasionally accessing C libraries to reduce run time) of well-known algorithms. Many of the implementations are based on the TISEAN package (Hegger, Kantz et al. 1999):

- The Largest Lyapunov exponent λ was calculated using an embedding dimension of two, and a time lag or delay time of one sample, considering at least 2000 reference points, and adjusting the neighbourhood size ε to obtain at least 10 neighbours per reference point, such that no neighbour was a direct chronological successor of the given reference point (i.e., the Theiler window was set to 2). The length of the embedding space trajectories compared for the estimation of λ was of 20. The algorithm used is described in (Hegger, Kantz et al. 1999, Kantz and schreiber 2004).
- The correlation dimension was calculated using an embedding dimension of two, and a time lag or delay time of one sample. The Theiler window was set to 2. The algorithm used is described in (Hegger, Kantz et al. 1999, Kantz and schreiber 2004).
- The scaling exponent α obtained via detrended fluctuation analysis (DFA) was calculated using a geometric window increase with exponent equal to 2 and no overlap of windows. Four different time scales were considered: 3-7 samples (α_1), 7-13 samples (α_2), 4-16 samples (α_3), and 16-64 samples (α_4). The method of DFA is described in (Kantz and schreiber 2004) and the citations therein. The data were detrended by means of a moving average method (Alvarez-Ramirez 2005).
- The sample entropy was calculated on the original time series of interbeat intervals as well as on coarse-grained time series constructed on the basis of collapsing the original values within a window of the size of the scale of interest to one value, namely the average of the measurements over the length of the window. The scales considered (sizes of windows) were 1-20. The parameters $SampEn_1$ to $SampEn_{10}$ correspond to the scales 1-10. In all cases, the sample entropy was calculated using a comparison length of m=2 points, and a tolerance of

r=0.2 *sdnn, where sdnn stands for the standard deviation of all normal interbeat intervals, according to the algorithm described in (Richman and Moorman 2000) and in the citations therein. The multiscale entropy was calculated according to (Costa, Goldberger et al. 2002) over all scales considered, that is 1-20. Local slopes of the plot of the sample entropy as a function of scale yielded the parameters MultiEnLow (slope within the scales 1-5), and MultiEnHigh (slope within the scales 6-20).

- A power-law relationship between the power spectral density (PSD) of the interbeat interval time series and frequency was determined by estimating the slope β of the linear best-fit of the PSD as a function of the frequency on a double logarithmic scale. Several parameters were obtained, depending on the range of frequencies used. The power spectral density (PSD) was estimated according to the method described in (Cusenza 2010). Regressing the power spectral for frequencies in the range 0.01 > freq > 0.0001 yielded the parameter PSD_1. Analogously, PSD_2 corresponds to 0.04 > freq > 0.02, PSD_3 to 0.45 > freq > 0.0001, PSD_4 to 0.5 > freq > 0.1, PSD_5 to 0.5 > freq > 0.2, and PSD_6 to 0.5 > freq > 0.3, respectively.

Statistical analysis

Hierarchical clustering analysis

To focus our attention on time series analysis parameters that provide orthogonal information about the participant's heart rate dynamics, we performed a clustering analysis. To this end, for a given time series analysis parameter, we grouped all the values obtained within the cohort to a row in a matrix. Then, we conducted hierarchical clustering on the rows of the matrix and identified clusters of time series analysis parameters that, in the context of the cohort analysed, seem to encode similar properties. A heat map representation of this procedure is depicted in **Figure E3.1**. With the exception of the standard measures of HRV which were all kept, we selected out of each of the identified clusters one parameter per computational method as a representative.

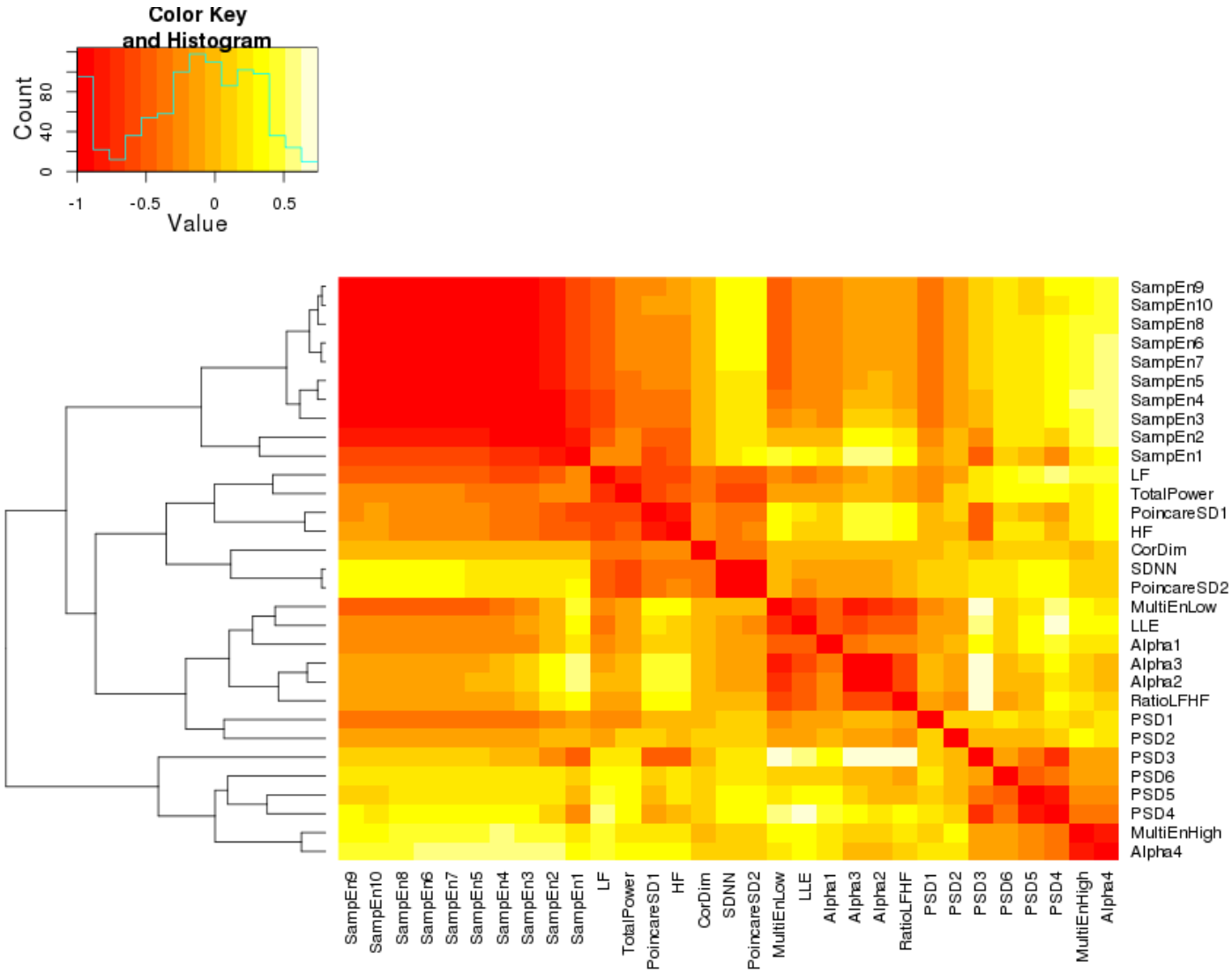

Figure E3.1. Hierarchical clustering analysis of parameters describing heart rate variability and heart rate dynamics, n=1481

Results

Hierarchical cluster analysis of the parameters of HRV and heart rate dynamics

By means of hierarchical clustering we identified four main clusters in the heart dynamics components (**Figure E3.1**). The first cluster included the exponent α_4, the Multiscale entropy high, and $PSD_{3\ to\ 6}$. In order to keep only one representative parameter per computational method in each cluster, we excluded $PSD_{3,4,6}$ for the following analyses. The second cluster included PSD_1 and PSD_2, the ratio LF/HF, $\alpha_{1\ to\ 3}$, the Largest Lyapunov Exponent and the Multiscale entropy low. We excluded PSD_1, α_2 and α_3. The third cluster included the Poincaré parameters SD_1 and SD_2, CD, and all the traditional parameters of HRV with the exception of the ratio LF/HF. Both Poincaré parameters SD_1 and SD_2 were kept for the following analyses since they are usually described together in the literature. The CD was excluded because of its specific distribution. Finally, the fourth cluster included all the parameters related to the sample entropy. Only one of them, the $SampEn_1$, was retained for the following analyses.

Exploration of the association between current smoking and the heart rate dynamics

Using non-standard time series analysis parameters

The parameters exponent α long-term time scale (α_4), PSD_2, multiscale entropy high and $SampEn_1$ did not detect any changes in the regulation of the cardiovascular system as a response to current tobacco smoke exposure (**Table E3.1**). Therefore, their association with the long-term smoking cessation was not assessed.

Exploration of the association between long-term smoking cessation and heart rate dynamics

A significant positive interaction between the total lifetime amount smoked (pack-years) and time elapsed since cessation (years) was found for SDNN, total power, LF, α short-term time scale, and multiscale entropy low, and Poincaré SD_2 (**Table E3.2**).

Table E3.1. Association between smoking status and non-standard parameters of HRV in current smokers (n=1420)

	α long-term time scale		**PSD_2**		**Multiscale entropy high**		**SampEn1**	
	coefficient±se	p-value	coefficient±se	p-value	coefficient±se	p-value	%GM, $_{95\%}$CI	p-value
Intercept	1.6±0.2		-2.1±0.4		-0.01±0.005			
Smoking status (ref.=Lifelong non-smokers)		0.72		0.74		0.77		0.25
Current light smoker	0.004±0.03	0.88	-0.002±0.07	0.98	0.0002±0.0007	0.80	-4.6% [-14.1;6.1]	0.39
Current heavy smoker	0.01±0.02	0.42	0.03±0.04	0.44	0.0003±0.0004	0.48	4.3% [-2.2;11.2]	0.20

All the models are adjusted for gender, age, ETS exposure, alcohol consumption, physical activity, diabetes, BMI, BMI squared, number of cardiovascular medication, average annual NO_2

Values shown are as percent changes in geometric means (GM) and 95% confidence interval ($_{95\%}$CI) or coefficient ± standard error (se)

Participants were classified as lifelong non-smokers if the total lifetime amount smoked was <0.1 pack-years. Smokers were defined as heavy smokers if the total lifetime amount smoked was ≥ 20 pack-years. Pack-years were calculated by multiplying the number of years smoked by the average number of packs smoked per day

Table E3.2. Tests of linear interactions between total lifetime amount smoked (pack-years) and time elapsed since cessation (years) (n=1420)

Outcome	Intercept	Pack-years		Time elapsed since cessation		Interaction	
	coefficient±se	coefficient±se or %GM, $_{95\%}$CI	p-value	coefficient±se or %GM, $_{95\%}$CI	p-value	coefficient±se or %GM, $_{95\%}$CI	p-value
SDNN		-0.2% [-0.3;-0.2]	<**0.001**	0.007% [-0.1;0.1]	0.91	0.009% [0.002;0.02]	**0.01**
Total power		-0.5% [-0.7;-0.3]	<**0.001**	0.1% [-0.1;0.4]	0.27	0.02% [0.005;0.04]	**0.01**
HF		-0.2% [-0.4;0.07]	0.18	0.09% [-0.3;0.4]	0.63	0.01% [-0.01;0.03]	0.32
LF		-0.5% [-0.7;-0.3]	<**0.001**	0.2% [-0.05;0.5]	0.11	0.02% [0.005;0.04]	**0.01**
Ratio LF/HF		-0.3% [-0.5;-0.2]	<**0.001**	0.1% [-0.1;0.4]	0.25	0.01% [-0.004;0.03]	0.16
PSD_5	-3.5±0.7	0.006±0.001	<**0.001**	-0.001±0.002	0.52	-0.0001±0.0001	0.20
α short-term time scale	2.9±0.3	-0.003±0.0005	<**0.001**	0.0009±0.0007	0.21	0.0001±0.00005	**0.02**
Multiscale entropy low	0.1±0.03	-0.0004±0.00005	<**0.001**	0.00007±0.002	0.37	0.00001±0.000005	**0.02**
Lyapunov Largest Exponent	0.4±0.06	-0.0007±0.0001	<**0.001**	-0.00004±0.0002	0.79	0.00001±0.00001	0.22
Poincaré SD_1		-0.06% [-0.2;0.06]	0.32	0.04% [-0.1;0.2]	0.63	0.006% [-0.005;0.02]	0.31
Poincaré SD_2	269.8±35.8	-0.43±0.07	<**0.001**	0.02±0.1	0.81	0.01±0.007	**0.04**

All the models are adjusted for gender, age, ETS exposure, alcohol consumption, physical activity, diabetes, BMI, BMI squared, number of cardiovascular medication, average annual NO_2

SDNN, standard deviation of all NN intervals; HF, power in the high frequency range; LF, power in the low frequency range; PSD, Power Spectral Density

Values shown are as percent changes in geometric means (GM) and 95% confidence interval ($_{95\%}$CI) or coefficient ± standard error (se)

Sensitivity analysis

Random effect of the study areas

Inclusion of a random effect for study area did not change the associations between heart rate dynamics and smoking status (data not shown).

Exclusion of outliers

For each of the standard and non-standard parameters, we excluded participants with a value lower than the 1th percentile or higher than the 99th percentile of the distribution of the parameter. Compared to the main analysis of the standard parameters, this analysis showed an additional significant decrease of the ratio LF/HF in former heavy smokers within the first 15 years of smoking cessation (-13.5% [-22.9%;-3.0%], p=0.01). The ratio LF/HF normalized to the level of lifelong non-smokers within 15-25 years of smoking cessation (-1.6% [-14.5%;13.3%], p=0.83). In regard with the non-standard parameters, a remarkable result was the normalization of the Largest Lyapunov Exponent in the former heavy smokers after 15-25 years of smoking cessation (-0.004±0.009, p=0.07).

Exclusion of the participants taking at least one cardiovascular medication or with missing information on the number of cardiovascular medications

Additional analysis of heart rate dynamics excluding the participants taking at least one cardiovascular medication or with missing information on the number of cardiovascular medications was performed. Regarding the standard parameters, we obtained the same results as in the main analysis, with the additional significant decrease of LF in former heavy smokers within the first 15 years of smoking cessation (-16.8% [-29.1%;-2.5%], p=0.02). The LF normalized to the level of lifelong non-smokers within 15-25 years of smoking cessation (5.5% [-15.4%;31.6%], p=0.63) (**Table E3.3**).

There were no noteworthy changes in the association between non-standard parameters and smoking status (data not shown).

Exclusion of both outliers and participants taking at least one cardiovascular medication or with missing information on the number of cardiovascular medications

Finally, after exclusion of both outliers and participants taking at least one cardiovascular medication, the Largest Lyapunov Exponent exhibited a full recovery within the first 15 years of smoking cessation (**Table E3.4**).

There were no noteworthy changes in the association between the standard parameters of HRV and smoking status (data not shown).

Interaction between smoking status and ACE inhibitors

Interaction between smoking status and ACE inhibitors was not significant for any outcome (data not shown).

Table E3.3. Association between smoking status and time-domain and frequency-domain measures of HRV in multivariable analysis, after exclusion of the participants taking at least one cardiovascular medication or with missing information on the number of cardiovascular medications, stratified by time elapsed since cessation (n=1020)

	SDNN %GM, $_{95\%}$CI	p-value	**Total power** %GM, $_{95\%}$CI	p-value	**HF** %GM, $_{95\%}$CI	p-value	**LF** %GM, $_{95\%}$CI	p-value	**Ratio LF/HF** %GM, $_{95\%}$CI	p-value
Time after cessation: 0 year										
Smoking status (ref.=Lifelong non-smoker)		*<0.001*		*<0.001*		*0.03*		*<0.001*		*0.04*
Current light smoker	-12.5% [-19.2;-5.2]	0.001	-27.5% [-39.4;-13.2]	<0.001	-20.0% [-37.9;3.0]	0.08	-25.0% [-38.3;-9.0]	0.004	-6.3% [-20.5;10.5]	0.44
Current heavy smoker	-13.8% [-17.8;-9.6]	<0.001	-29.1% [-36.3;-21.0]	<0.001	-15.7% [-27.6;-1.8]	0.03	-25.5% [-33.7;-16.3]	<0.001	-11.7% [-20.0;-2.5]	0.01
Time after cessation:]0-15[years										
Smoking status (ref.=Lifelong non-smoker)		*<0.001*		*0.002*		*0.43*		*0.05*		*0.09*
Former light smoker	-1.5% [-10.2;8.2]	0.76	-6.2% [-24.4;16.4]	0.56	9.0% [-19.7;47.9]	0.58	-11.8% [-30.2;11.5]	0.29	-19.0% [-34.1;-0.5]	0.04
Former heavy smoker	-12.5% [-17.9;-6.8]	<0.001	-23.2% [-33.7;-11.0]	<0.001	-11.0% [-27.7;9.6]	0.27	-16.8% [-29.1;-2.5]	0.02	-6.6% [-18.8;7.5]	0.34
Time after cessation: [15-25[years										
Smoking status (ref.=Lifelong non-smoker)		*0.79*		*0.72*		*0.46*		*0.50*		*0.74*
Former light smoker	-0.9% [-7.2;5.8]	0.78	4.5% [-10.2;21.6]	0.57	10.4% [-10.6;36.4]	0.36	9.5% [-7.1;29.1]	0.28	-0.8% [-13.7;14.0]	0.91
Former heavy smoker	-2.8% [-11.0;6.1]	0.52	6.3% [-13.3;30.3]	0.55	13.5% [-14.5;50.7]	0.38	5.5% [-15.4;31.6]	0.63	-7.0% [-22.8;12.0]	0.44
Time after cessation: ≥ 25 years										
Smoking status (ref.=Lifelong non-smoker)		*0.15*		*0.43*		*0.91*		*0.45*		*0.27*
Former light smoker	-4.4% [-8.7;0.2]	0.06	-6.2% [-15.7;4.3]	0.23	-2.8% [-16.3;12.8]	0.71	-3.5% [-14.0;8.4]	0.55	-0.7% [-10.1;9.7]	0.89
Former heavy smoker	1.3% [-9.5;13.4]	0.82	-7.8% [-28.8;19.4]	0.54	3.2% [-28.1;48.3]	0.86	-15.2% [-36.0;12.4]	0.25	-17.8% [-35.5;4.7]	0.11

All the models are adjusted for gender, age, ETS exposure, alcohol consumption, physical activity, diabetes, BMI, BMI squared, average annual NO_2
SDNN, standard deviation of all NN intervals; HF, power in the high frequency range; LF, power in the low frequency range
Values shown are as percent changes in geometric means (GM) and 95% confidence interval ($_{95\%}$CI)
Participants were classified as lifelong non-smokers if the total lifetime amount smoked was <0.1 pack-years. Smokers were defined as heavy smokers if the total lifetime amount smoked was ≥ 20 pack-years. Pack-years were calculated by multiplying the number of years smoked by the average number of packs smoked per day

Table E3.4. Association between smoking status and non-standard parameters in multivariable analysis, after exclusion of the outliers and participants taking at least one cardiovascular medication or with missing information on the number of cardiovascular medications, stratified by time elapsed since cessation (n=1000)

	Category 1								**Category 2**			
	PSD_5		**α short-term time scale**		**Multiscale entropy low**		**Largest Lyapunov exponent**		**Poincaré SD_1**		**Poincaré SD_2**	
	coefficient±se	p-value	coefficient±se	p-value	coefficient±se	p-value	coefficient±se	p-value	%GM, $_{95\%}$CI	p-value	coefficient± se	p-value
Time after cessation: 0 year												
Intercept	-2.3±0.8		3.3±0.3		0.2±0.03		0.4±0.07				200.4±44.5	
Smoking status (ref.=Lifelong non-smoker)		0.02		0.06		<0.001		<0.001		0.005		<0.001
Current light smoker	-0.03±0.1	0.82	-0.06±0.04	0.14	-0.004±0.005	0.40	-0.004±0.01	0.75	-12.9% [-22.3;-2.3]	0.02	-19.8±7.3	0.007
Current heavy smoker	0.2±0.08	0.006	-0.05±0.03	0.05	-0.01±0.003	<0.001	-0.03±0.007	<0.001	-8.4% [-14.4;-1.9]	0.01	-21.9±4.4	<0.001
Time after cessation:]0;15[years												
Intercept	-2.5±0.9		3.2±0.3		0.1±0.03		0.4±0.07				242.1±46.7	
Smoking status (ref.=Lifelong non-smoker)		0.16		0.30		0.04		0.34		0.54		<0.001
Former light smoker	-0.004±0.2	0.98	-0.02±0.05	0.70	-0.009±0.006	0.12	-0.008±0.01	0.53	5.4% [-8.2;21.1]	0.45	0.2±8.8	0.98
Former heavy smoker	0.2±0.1	0.06	-0.06±0.04	0.13	-0.009±0.004	0.03	-0.01±0.009	0.17	-3.5% [-12.3;6.2]	0.46	-23.8±6.1	<0.001
Time after cessation: [15;25[years												
Intercept	-3.4±0.8		3.2±0.3		0.1±0.03		0.4±0.07				237.6±46.6	
Smoking status (ref.=Lifelong non-smoker)		0.66		0.92		0.17		0.57		0.29		0.63
Former light smoker	0.06±0.1	0.60	0.005±0.04	0.89	0.0005±0.004	0.91	-0.005±0.009	0.61	5.8% [-4.2;16.8]	0.27	3.7±6.3	0.55
Former heavy smoker	0.1±0.2	0.44	-0.02±0.05	0.71	-0.01±0.006	0.06	-0.01±0.01	0.35	8.1% [-5.4;23.4]	0.25	-5.9±8.3	0.48
Time after cessation: ≥ 25 years												
Intercept	-2.8±0.8		3.4±0.3		0.1±0.03		0.4±0.07				226.8±45.0	
Smoking status (ref.=Lifelong non-smoker)		0.76		0.47		0.70		0.40		0.67		0.44
Former light smoker	-0.007±0.08	0.93	0.01±0.03	0.68	0.0009±0.003	0.77	-0.009±0.007	0.20	-2.6% [-9.1;4.4]	0.46	-5.1±4.5	0.25
Former heavy smoker	0.1±0.2	0.47	0.08±0.07	0.23	-0.005±0.007	0.45	-0.008±0.02	0.61	3.7% [-12.3;22.5]	0.68	4.5±10.9	0.68

All the models are adjusted for gender, age, ETS exposure, alcohol consumption, physical activity, diabetes, BMI, BMI squared, average annual NO_2

PSD, Power Spectral Density

Values shown are as percent changes in geometric means (GM) and 95% confidence interval ($_{95\%}$CI) or coefficient ± standard error (se)

Participants were classified as lifelong non-smokers if the total lifetime amount smoked was <0.1 pack-years. Smokers were defined as heavy smokers if the total lifetime amount smoked was ≥ 20 pack-years. Pack-years were calculated by multiplying the number of years smoked by the average number of packs smoked per day

4. Chapter 2: Evidence of adverse effects of long-term exposure to traffic-related PM_{10} on heart rate variability and heart rate dynamics in healthy subjects

D. Meier-Girard, E. Delgado, E. Schaffner, C. Schindler, N. Künzli, M. Adam, M. Imboden, U. Frey*, N. Probst-Hensch*

*equal senior author

Corresponding author: Urs Frey, Pediatric Pulmonology Research Group, University Children's Hospital, Spitalstrasse 33, 4056 Basel, Switzerland. Phone: +41 61 704 29 01, Fax: +41 61 704 12 13. E-mail address: urs.frey@ukbb.ch

This article will be submitted to

Environmental Health Perspectives

4.1. Abstract

Aim: To evaluate the influence of low-level, but long-term exposure (10 years), to traffic-related particulate matter (TPM_{10}) on the regulation of the autonomic cardiovascular system and heart rate dynamics in an aging general population, as well as the a priori selected effect modifiers sex, smoking status, obesity, and gene variation in selected glutathione S-transferases (GSTs).

Methods: We analyzed data from 1593 participants aged ≥ 50 years from the SAPALDIA cohort study. For each participant, heart rate variability and heart rate dynamics were characterized by means of various quantitative analyses of the inter-beat interval time series generated from 24-hour electrocardiogram recordings. Each parameter obtained was then used as the outcome variable in multivariable mixed linear regression models in order to evaluate the association with long-term exposure to traffic-related PM_{10}. The models were adjusted for known confounding factors. Interaction between long-term exposure to traffic-related PM_{10} and the a priori selected effect modifiers were tested.

Results: We did not observe an overall association between long-term exposure to TPM_{10} and heart rate variability/heart rate dynamics in the entire study population. However, significant changes in the heart rate dynamics were found in subjects without cardiovascular morbidity and significant changes both in the heart rate dynamics and in heart rate variability were found in non-obese subjects without cardiovascular morbidity. Furthermore, subjects with homozygous GSTM1 gene deletion appeared to be more susceptible to the effects of TPM_{10}.

Conclusion: This study provides evidences that long-term exposure to TPM_{10} triggers adverse changes in the regulation of the cardiovascular system. These adverse effects were more visible in the healthy subjects, in whom the overall relationship between TPM_{10} and heart rate variability/heart rate dynamics was not modified by an underlying health condition and the eventual countering effects of related drug treatments.

Keywords: heart rate variability; nonlinear dynamics; air pollution; particulate matter; vehicle emissions

4.2. Introduction

Short- and long-term exposure to particulate matter (PM) air pollution has been associated with increased cardiovascular morbidity and mortality, with greater risks in susceptible populations, such as the elderly, individuals with diabetes, patients with preexisting coronary heart disease, chronic lung disease, or heart failure, and individuals with low education or socioeconomic status (Pope, Burnett et al. 2004, Brook, Rajagopalan et al. 2010, Pieters, Plusquin et al. 2012). Current or previous smokers, obesity and sex could also be susceptibility factors (Brook, Rajagopalan et al. 2010).

Possible mechanisms for these associations include effects on the autonomic nervous system. Heart rate variability (HRV) is a useful non-invasive measure to assess the autonomic regulation of cardiac rhythm (1996). Lower HRV is associated with higher cardiovascular morbidity and mortality, and has proved itself as an important prognostic tool for several cardiovascular conditions (Kleiger, Miller et al. 1987, Bigger, Fleiss et al. 1992, 1996, Tsuji, Larson et al. 1996). There is strong overall epidemiological evidence that short-term PM exposure (days) is associated with reductions in most indices of HRV, and the association might be more pronounced among the elderly, patients with preexisting cardiovascular disease or diabetes, or people with reduced antioxidative defenses (Park, O'Neill et al. 2005, Brook, Rajagopalan et al. 2010, Pieters, Plusquin et al. 2012, Mordukhovich, Coull et al. 2015). In particular, recent observations have shown a strong effect modification of the HRV-PM relationship, as well as of the HRV-second-hand smoke and HRV-BMI relationships, by genes that modulate endogenous oxidative stress, such as glutathione S-transferase (GST) (Schwartz, Park et al. 2005, Park, O'Neill et al. 2006, Baccarelli, Cassano et al. 2008, Probst-Hensch, Imboden et al. 2008, Adam, Imboden et al. 2017), suggesting that air pollutants might impact in part through inflammatory and oxidative stress pathways.

Although long-term PM exposure is known to have a stronger effect on cardiovascular morbidity and mortality than acute exposure, there is limited or weak available epidemiological evidence that HRV is altered by low-level, but long-term exposure (years) (Brook, Rajagopalan et al. 2010). Indeed, studies on the chronic impact of PM air pollution on HRV are scarce (Adam, Felber Dietrich et al. 2012, Adam, Imboden et al. 2014, Mordukhovich, Coull et al. 2015) and the American Heart Association recently stated that studies on the long-term effects of air pollution on HRV and cardiovascular health are a major unresolved issue.

Finally, there is increasing evidence that the regulation of the cardiovascular system involves nonlinear control mechanisms (1996, Rajendra Acharya, Paul Joseph et al. 2006) which can best be characterized using nonlinear time series analysis techniques (Goldberger and West 1987, Pincus 1991, Pikkujamsa, Makikallio et al. 2001, Meyer and Stiedl 2003, Rajendra Acharya, Paul Joseph et al. 2006, Vandeput, Verheyden et al. 2012). The recent implementation of such methods to evaluate the influence of current smoking and smoking cessation on heart rate dynamics, in the large epidemiological dataset of SAPALDIA, which allowed for the control of the most potential confounders, enabled us to unveil long-term alterations in former heavy smokers who might need up to 15-25 years to fully recover. (Meier-Girard et al. 2016).

By applying the same kind of approach, the present study aimed first at evaluating the influence of low-level, but long-term (10 years), exposure to traffic-related particulate matter (TPM_{10}) on the regulation of the autonomic cardiovascular system and heart rate dynamics in an aging general population. Second, we specifically focused our investigation on the sub-populations with or without cardiovascular morbidity (i.e., cardiovascular disease and/or hypertension). Finally, we investigated the a priori selected effect modifiers - sex, smoking status, obesity, and gene variation in selected glutathione S-transferases (GSTs) - in the entire population, as well as in the sub-populations with or without cardiovascular morbidity.

4.3. Methods

4.3.1. Ethics statement

The study was approved by the Central Ethics Committee of the Swiss Academy of Medical Sciences and the Cantonal Ethics Committees for each of the study areas. Each subject was informed in detail about the health examinations and signed and written informed consent before any of the health examinations were conducted.

4.3.2. Study population

This study is part of the SAPALDIA (Swiss Cohort Study on Air Pollution and Lung and Heart Disease in Adults) study which was designed to assess the health effects of long-term exposure to air pollutants in the Swiss adult population. The study design has been described in detail elsewhere (Ackermann-Liebrich et al., 2005; Martin et al., 1997). In brief, the SAPALDIA cohort (n=9651) was enrolled in 1991, and consisted of a random sample of the

Swiss population aged 18 to 60 years, recruited from the local registries of inhabitants in eight areas featuring distinct geographical and environmental conditions.

In 2002, the follow-up study included 8047 (83.4%) participants. A random sample of 1846 out of 4417 participants, aged ≥ 50 years underwent a 24-hour electrocardiogram (ECG) Holter recording to assess HRV, as previously described in detail (Felber Dietrich et al., 2006). Exclusion criteria were general or spinal anaesthesia within 8 days before the ECG recording (n=5), a myocardial infarction within 3 months prior to the examination (n=2), taking digitalis (n=6), and an artificial internal pacemaker (n=0). Participants with recordings showing atrial fibrillation (n=12), ECG duration lower than 18 hours (n=73), or of insufficient quality (n=6), non-valid data on HRV (n=96) were also excluded (Felber Dietrich et al., 2006). This current analysis is restricted to 1593 participants with valid data on HRV, cardiovascular risk factors, and TPM_{10} exposure.

4.3.3. Questionnaires and measurements

Information about questionnaires and biological measurements (i.e., body mass index, blood pressure, heart rate, uric acid, high-sensitivity C-reactive protein) has been reported elsewhere (Martin, Ackermann-Liebrich et al. 1997, Ackermann-Liebrich, Kuna-Dibbert et al. 2005, Felber Dietrich, Schindler et al. 2006).

HRV measurements and measures of heart rate dynamics

Time series analysis parameters of heart rate variability were calculated for each individual time series of inter-beat intervals (RR series) generated from the 24-hour ECG recordings.

The traditional time domain measure used was the standard deviation of normal interbeat intervals (SDNN) (1996). Additionally, a power-law relationship between the power spectral density (PSD) of the interbeat interval time series and frequency was determined by estimating the slope β of the linear best-fit of the PSD as a function of the frequency on a double logarithmic scale. In our previous study, related to heart rate dynamics and smoking exposure, we found a positive association between slope β and smoking exposure (Girard, Delgado-Eckert et al. 2015).

We used nonlinear time series analysis methods to quantify and characterize the heart rate dynamics. The following heart rate dynamics (HRD) parameters were calculated:

- Exponent α: we used detrended fractal analysis (DFA) to measure the presence or absence of fractal correlation properties in signals (namely the "memory effect"). This method has been validated for interbeat intervals time series (Peng, Havlin et al. 1995). The fractal long-range correlations are characterized by a scaling exponent α. A fractal-like signal results in α=1. White Gaussian noise (totally random signal) results in a value of 0.5. In healthy young subjects, it is closer to 1, and this value falls within different ranges for various types of cardiac abnormalities. In our previous study, related to heart rate dynamics and smoking exposure, we found an inverse association between α and smoking exposure (Girard, Delgado-Eckert et al. 2015).
- Largest Lyapunov exponent: detection of chaos in a time series can be done by measuring the largest Lyapunov exponent in the appropriate phase space embedding (Rosenstein, Collins et al. 1993). It quantifies the exponential divergence of initially close state-space trajectories and estimates the amount of chaos in a system. The extent to which chaos relates to physiological or pathological dynamics is a subject of active investigation and some controversy (Goldberger, Amaral et al. 2000). In our previous study, related to heart rate dynamics and smoking exposure, we found an inverse association between the largest Lyapunov exponent and smoking exposure (Girard, Delgado-Eckert et al. 2015).

More details about the choice, implementation, and properties of the aforementioned time series analysis methods are described in the **Online Supplement**.

Air pollutant exposure estimation

TPM_{10} estimates were obtained over ten years (1990-2000) using a dispersion modeling approach (Liu, Curjuric et al. 2007). In accordance with previous investigations of traffic-related PM_{10} as part of the SAPALDIA cohort study (Adam, Felber Dietrich et al. 2012, Adam, Imboden et al. 2014), mean of the 10 indicators was used to obtain the average concentration of TPM_{10} over 10 years.

Genotyping

The genotyping has been described in detail elsewhere (Probst-Hensch, Imboden et al. 2008). In brief, all subjects were genotyped for GSTM1 (UniGene ID Hs.301961; UniGene 2008a) and GSTT1 (UniGene Hs.268573; UniGene 2008b) gene deletions.

4.3.4. Statistical analysis

All tests were two-sided with a significance level of 0.05. Statistical analysis was performed using R, Version 3.3.3 (2008).

Descriptive analysis

Results are expressed as numbers and percentages for categorical variables and as a mean ± standard deviation or median [25^{th}quartile; 75^{th}quartile] for continuous variables, according to their distribution.

Multivariable analysis

Each parameter describing the HRV, or heart rate dynamics, was used as the outcome variable in multivariable linear regression models in order to evaluate the association with long-term exposure to TPM_{10} (for an increase of 10 $\mu g/m^3$ of TPM_{10}). Initial inspection of the outcome variable showed a skewed distribution of the residuals for the traditional time and frequency domain measures and for some of the other time series analysis parameters. These variables were therefore log-transformed. Results of these analyses are therefore presented as geometric means and percent changes in geometric means. All the models were adjusted for known confounding factors (Felber Dietrich, Schindler et al. 2006, Adam, Felber Dietrich et al. 2012, Adam, Imboden et al. 2014). These factors were: sex (male as reference), age (for an increase of 1 year), age^2, body mass index (BMI, for an increase of 1 kg/m^2), BMI^2, alcohol consumption (<1 glass/day as reference, ≥ 1 glass/day), weekly physical activity – to the point of getting out of breath or sweating – (never as reference, between 0.5h and 2h, ≥ 2h/week), daily exposure to environmental tobacco smoke (for an increase of 1 hour/day), diabetes (no as reference, yes), smoking group (lifelong non-smoker as reference, former light smoker, former heavy smoker, current light smoker, current heavy smoker), uric acid concentration (µmol/l), high-sensitivity C-reactive protein (mg/l), street and railway noise exposure (mean dB(A) per night), seasonal effects (sine and cosine functions of the day of examination with a period of 1 year), education level (high as reference, middle, low), employment category (employed as reference, unemployed, house person, pensioner), occupational exposure (no as reference, yes if current exposure to dust, gas/smoke/aerosols/fumes/vapors at the working place), cardiovascular morbidity (no as reference, yes). Cardiovascular morbidity was defined as “no” if there was no evidence for cardiovascular disease or hypertension (i.e., the subject had no physician diagnosed heart disease, no major cardiovascular medication intake, and no

hypertension). Major cardiovascular medication consisted of beta-blockers, angiotensin-converting-enzyme (ACE) inhibitors, angiotensin II receptor antagonists, calcium channel blockers, diuretic medications, antiarrhythmic drugs class I + III, sympathomimetic medications. Absence of hypertension was defined as absence of a physician diagnosis of hypertension, blood pressure in the hypertensive range, and antihypertensive medication.

Investigation of susceptible groups

Modifying effect of cardiovascular morbidity, sex, as well as of inflammation and oxidative stress related parameters such as smoking status (defined as ever smoker or lifelong non-smoker), obesity (defined as BMI $\geq$ 30 kg/m^2), and known polymorphisms in the GSTM1, GSTT1, and GSTP1 genes previously found to modify the smoking-HRV association in the SAPALDIA cohort study (Probst-Hensch, Imboden et al. 2008) were assessed for each outcome.

Sensitive analysis

Given previous evidence of higher air pollution susceptibility for patients with diabetes, participants with an according physician diagnosis of diabetes were excluded in a sensitivity analysis.

According to previous work on the SAPALDIA cohort, which provided evidence that participants under ACE inhibitor therapy represented a specific subgroup susceptible to the adverse effects of TPM_{10} on the traditional parameters of HRV (Adam, Felber Dietrich et al. 2012), a sensitive analysis was conducted to assess the TPM_{10}-HRV/HRD relationship by excluding patients under ACE inhibitors therapy from the entire study population.

Investigation of the PM_{10}-HRV/HRD relationship

In addition, we investigated the PM_{10}-HRV/HRD relationship. Methods and results are provided in the **Online Supplement**.

4.4. Results

4.4.1. Study population

The study population consisted of 1593 subjects. The mean age of the subjects was 60.5±6.2 years. Demographic characteristics, lifestyle factors, cardiovascular health and diabetes, long-term exposure to air pollution, and GST genotypes are summarized in **Table**

4.1. A more detailed description has been reported elsewhere (Adam, Felber Dietrich et al. 2012, Girard, Delgado-Eckert et al. 2015).

Table 4.1. Characteristics of the study population and subpopulations investigated

Characteristic	**Entire study population (n=1593)**	**MD**	**Subpopulation without cardiovascular morbidity (n=510)**	**MD**
Demographic characteristics				
Age, years	60.5±6.2	-	59.6±6.01	-
Sex, Men	773 (48.5)	-	200 (39.2)	-
BMI, kg/m^2	26.7±4.34	3	25.1±3.78	-
Education		-		-
low	144 (9.0)		48 (9.4)	
middle	1048 (65.8)		325 (63.7)	
high	401 (25.2)		137 (26.9)	
Employment		10		6
employed	862 (54.5)		291 (57.7)	
house person	352 (22.2)		121 (24)	
unemployed	81 (5.1)		21 (4.2)	
pensioner	288 (18.2)		71 (14.1)	
Lifestyle factors				
Smoking status		60		12
lifelong non-smoker	692 (45.1)		230 (46.2)	
current light smoker	65 (4.2)		28 (5.6)	
current heavy smoker	222 (14.5)		71 (14.3)	
former light smoker	314 (20.5)		112 (22.5)	
former heavy smoker	240 (15.7)		57 (11.4)	
Time elapsed since cessation, years		7		3
<15	154 (28.2)		46 (27.7)	
15-25	150 (27.4)		44 (26.5)	
≥ 25	243 (44.4)		76 (45.8)	
Daily ETS exposure, hours		2		2
none	1253 (78.8)		407 (80.1)	
<3	216 (13.6)		63 (12.4)	
≥ 3	122 (7.7)		38 (7.5)	
Alcohol, ≥ 1 glass/day	731 (45.9)	2	211 (41.5)	2
Weekly physical activity		14		5
none	666 (42.2)		203 (40.2)	
30min-1h	516 (32.7)		183 (36.2)	
2h or more	397 (25.1)		119 (23.6)	
Noise exposure, dB(A)	56.6±7.31	7	57±7.11	3
Cardiovascular health and diabetes				
Diabetes	80 (5.0)	-	9 (1.8)	-
Heart disease diagnosed by a doctor	126 (7.9)	-	0 (0)	-

Table 4.1. Characteristics of the study population and subpopulations investigated (continued)

Characteristic	**Entire study population (n=1593)**	**MD**	**Subpopulation without cardiovascular morbidity (n=510)**	**MD**
Hypertension	861 (54.0)	-	0 (0)	-
Major cardiovascular medication (≥1)	403 (31.7)	321	0 (0)	-
ACE inhibitor therapy	102 (8.0)	321	0 (0)	-
Uric acid, µmol/l	326±85.93	56	300±77.01	15
hs-CRP, mg/l	1.2 [0.6;2.6]	56	1 [0.5;2]	15
Heart rate (bpm)	74.2±9.1	1	74.6±7.85	1
SDNN (msec)	136.5±35.22	-	140.5±34.49	-
Air pollutants exposure				
Occupational exposure	400 (25.2)	3	131 (25.8)	2
PM_{10}, µg/m^3	20.9 [17.8;25.1]	9	20.2 [17.1;24.7]	5
Traffic-related PM_{10}, µg /m^3	1.9 [1.2;3.1]	9	1.5 [1.1;2.9]	5
GST genotypes				
GSTM1 deletion	781 (52.3)	100	240 (50.3)	33
GSTT1 deletion	261 (17.5)	100	79 (16.6)	33
GSTM1T1 deletion	145 (9.7)	100	41 (8.6)	33

ACE inhibitor, angiotensin-converting-enzyme inhibitor; BMI, body mass index; ETS, Environmental Tobacco Smoke; hs-CRP, high-sensitivity C-reactive protein; GST, glutathione S-transferase; MD, missing data; PM, particulate matter; SDNN, standard deviation of all NN intervals

Values shown are mean ± standard deviation, median [25thquartile; 75thquartile] and numbers (percentages)

4.4.2. Relationship between long-term exposure to TPM_{10} and heart rate variability/heart rate dynamics

There was no significant association between long-term exposure to TPM_{10} and HRV/HRD parameters (**Table 4.2**), as well as when stratifying by sex, smoking status, obesity, and GST genotypes (**Online Supplement**).

However, stratification by cardiovascular morbidity revealed significant associations between TPM_{10} and the HRD parameters slope β (0.8±0.3, p=0.01, interaction TPM_{10}*cardiovascular morbidity: p=0.04) and largest Lyapunov exponent (-0.06±0.03, p=0.03, interaction TPM_{10}*cardiovascular morbidity: p=0.08) in subjects without cardiovascular morbidity (**Table 4.3, Figure 4.1**).

Table 4.2. Association between HRV and heart rate dynamics parameters (outcome variable) and traffic-related PM_{10} in linear mixed effects regression models (random intercepts for study area)

	Entire study population (n=1237*)		
	intercept	%GM, 95%CI or coefficient±SE	p-value
SDNN		-6.5 [-17.2;5.7]	0.27
α		-3.5 [-12.0;5.3]	0.42
Slope β	-4.0±2.8	-0.03±0.2	0.89
Largest Lyapunov exponent	0.2±0.2	-0.01±0.02	0.63

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity

SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

Table 4.3. Association between HRV and heart rate dynamics parameters (outcome variable) and traffic-related PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by cardiovascular morbidity

	Cardiovascular morbidity (n=775*)			**No cardiovascular morbidity (n=462*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-2.2 [-16.2;12.7]	0.76		-14.3 [-29.0;6.3]	0.15	0.36
α		-0.3 [-11.8;12.7]	0.96		-8.8 [-20.1;2.0]	0.11	0.24
Slope β	-3.4±3.7	-0.3±0.3	0.36	-3.8±4.4	0.8±0.3	0.01	0.04
Largest Lyapunov exponent	0.3±0.3	0.02±0.02	0.47	0.4±0.4	-0.06±0.03	0.03	0.08

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, and diabetes

SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

¥ Interaction between TPM_{10} and cardiovascular morbidity

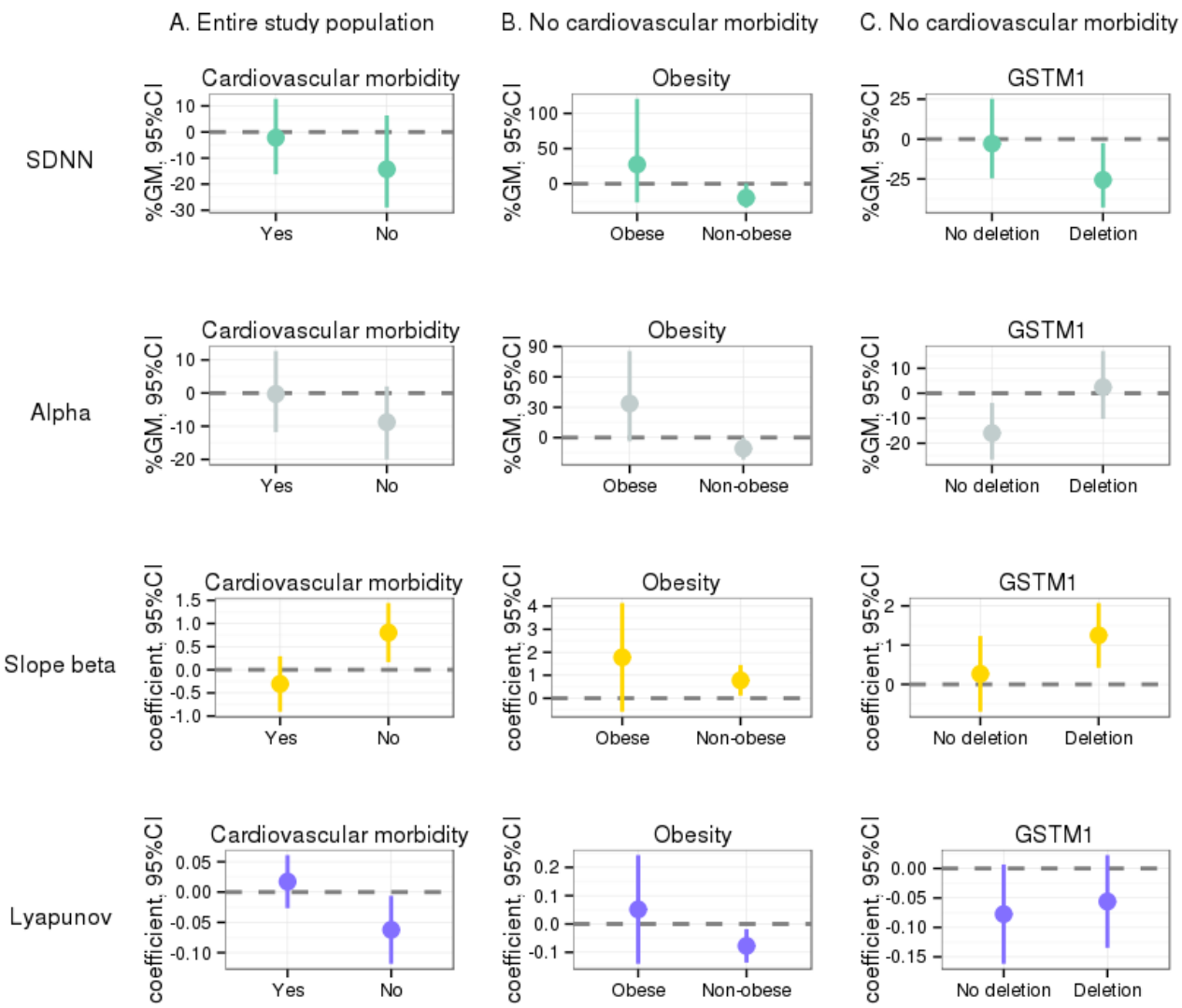

Figure 4.1. Percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient and 95%CI of SDNN, α, slope β, and largest Lyapunov exponent, for an increase of 10 $\mu g/m^3$ of traffic-related PM_{10}, in models stratified by cardiovascular morbidity in the entire study population (A), and by obesity (B) and GSTM1 (C) in the subpopulation without cardiovascular morbidity

GST, glutathione S-transferase; SDNN, standard deviation of all NN intervals

4.4.3. Investigation of the subgroup without cardiovascular morbidity

TPM_{10} effects became particularly visible in non-obese subjects both in the HRV parameter (SDNN: -20.0% [-33.7%;0.2%], p=0.05, interaction TPM_{10}*obesity: p=0.01) and in the HRD parameters (α: -10.8% [-21.7%;0.06%], p=0.05; slope β: 0.8±0.3, p=0.02; largest Lyapunov exponent: -0.08±0.03, p=0.01) (**Table 4.4, Figure 4.1**). This finding suggests that TPM_{10} effects might be more visible in subjects without any comorbidity ("healthy subjects").

There was no effect modification by sex (**Online Supplement Table E4.1**) and no clear effect modification by smoking status (**Online Supplement Table E4.2**).

We found strong significant associations between TPM_{10} and HRV/HRD parameters (SDNN: -25.6% [-42.9%;-2.5%], p =0.03; slope β: 1.2±0.4, p<0.001) in subjects with homozygous GSTM1 gene deletion (**Table 4.5, Figure 4.1**). These findings are consistent with the hypothesis that air pollutants might impact in part through oxidative stress pathways. Conversely, the HRD parameter α was significantly decreased (-16.0% [-26.6%;-3.9%], p=0.01) in subjects without GSTM1 deficiency. That might be explained by the fact that those subjects are likely to be more healthy (not likely to have systemic inflammation and oxidative stress) and thus, similarly to our findings in non-obese subjects, TPM_{10} effects might be visible in such subjects as well.

When stratifying by GSTT1 genotype, we observed significant associations in subjects without GSTT1 deficiency (**Online Supplement Table E4.3**). However, coefficients in subjects with homozygous GSTT1 gene deletion were similar to that in subjects without GSTT1 deficiency, and confidence intervals were very broad. The small sample size of the subgroup of subjects with homozygous GSTM1 gene deletion (n=69) might have limited the statistical power of our analyses. Consequently, there is no evidence of effect modification by GSTT1 genotype.

Table 4.4. Association between HRV and heart rate dynamics parameters (outcome variable) and traffic-related PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by obesity

Entire study population	**Non-obese (n=974*)**			**Obese (n=263*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-6.7 [-18.2;8.5]	0.33		-6.5 [-27.5;20.7]	0.61	0.72
α		-2.7 [-10.9;6.1]	0.53		-2.9 [-19.5;17.2]	0.76	0.88
Slope β	-4.4±3.5	-0.05±0.2	0.82	-12.4±7.2	0.3±0.4	0.54	0.34
Largest Lyapunov exponent	-0.3±0.3	-0.005±0.02	0.98	1.0±0.7	-0.03±0.04	0.41	0.21
Subpopulation without cardiovascular morbidity	**Non-obese (n=415*)**			**Obese (n=47*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-20.0 [-33.7;0.2]	0.05		27.4 [-26.5;120.6]	0.38	0.01
α		-10.8 [-21.7;0.06]	0.05		33.6 [-3.8;85.6]	0.08	0.13
Power spectral density	-0.9±4.9	0.8±0.3	0.02	-21.9±26.3	1.8±1.2	0.14	0.56
Largest Lyapunov exponent	-0.07±0.4	-0.08±0.03	0.01	2.8±2.1	0.05±0.1	0.60	0.11

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity (except for the subpopulation without cardiovascular morbidity)

SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

¥ Interaction between TPM_{10} and obesity

Table 4.5. Association between HRV and heart rate dynamics parameters (outcome variable) and traffic-related PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by GSTM1 genotype

Entire study population	**Deletion in GSTM1 (n=620*)**			**No deletion in GSTM1 (n=572*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-10.4 [-24.0;5.6]	0.19		-3.0 [-16.2;12.3]	0.69	0.38
α		2.5 [-9.0;16.2]	0.68		-8.1 [-18.6;2.2]	0.12	0.23
Slope β	-5.0±4.0	0.2±0.3	0.48	0.6±4.1	-0.07±0.3	0.82	0.65
Largest Lyapunov exponent	0.2±0.4	-0.009±0.03	0.76	0.1±0.3	-0.01±0.02	0.66	0.79
Subpopulation without cardiovascular morbidity	**Deletion in GSTM1 (n=227*)**			**No deletion in GSTM1 (n=217*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-25.6 [-42.9;-2.5]	0.03		-2.8 [-24.6;25.2]	0.82	0.11
α		2.4 [-10.3;16.9]	0.72		-16.0 [-26.6;-3.9]	0.01	0.21
Slope β	1.3±5.8	1.2±0.4	<0.001	0.2±6.5	0.3±0.5	0.59	0.29
Largest Lyapunov exponent	1.2±0.6	-0.06±.04	0.16	-0.2±0.6	-0.08±0.04	0.08	0.57

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity (except for the subpopulation without cardiovascular morbidity)

GST, glutathione S-transferase; SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

¥ Interaction between TPM_{10} and GSTM1 genotype

4.4.4. Sensitive analyses

As previously shown in the SAPALDIA cohort study, there was a strong decrease in SDNN (-49.9% [-74.4%;-6.8%], p=0.03) in subjects under ACE inhibitor therapy (**Online Supplement Table E4.4**). In contrast, we did not find any particular changes in the HRD parameters. The relationship between long-term exposure to TPM_{10} and HRV/HRD parameters was not modified by exclusion of patients under ACE inhibitor therapy from the entire study population (data not shown).

Exclusion of subjects with diabetes made associations even stronger in non-obese subjects for SDNN (-20.9 [-34.6;-1.5], p=0.04) and α (-11.3 [-22.3;-0.5], p=0.04) (**Online Supplement Table E4.5**), as well as in subjects without GSTT1 deletion for α (-11.3 [-22.3;-0.1], p=0.05) and slope β (0.8±0.4, p=0.05) (**Online Supplement Table E4.6**).

4.4.5. Relationship between long-term exposure to PM_{10} and heart rate variability/heart rate dynamics

Findings with PM_{10} were very similar to that found by investigating the relationship between long-term exposure to TPM_{10} and HRV/HRD, though the effect size of PM_{10} was smaller than that of $TPM_{10,}$ and resulted in the same conclusions (**Online Supplement**).

4.5. Discussion

4.5.1. Main findings

This study evaluates the influence of long-term exposure to TPM_{10} on HRV and heart rate dynamics. While we did not find any overall association in the entire study population, we observed strong significant associations of long-term exposure to TPM_{10} with the HRD parameters slope β and largest Lyapunov exponent in subjects without cardiovascular morbidity. These findings might be explained by the fact that the relative contribution of both the underlying health condition and the countering effects of drug treatments on the TPM_{10}–HRV/HRD relationship might render this relationship so variable that the overall TPM_{10}–HRV/HRD relationship in such subjects might be null. In contrast, the TPM_{10}–HRV/HRD relationship might become more visible in subjects without cardiovascular disease and related drug treatments.

This hypothesis is supported by the fact that TPM_{10} effects became even more visible in the subgroup of non-obese subjects without cardiovascular morbidity, as shown by both HRV and HRD parameters. Again, in these subjects, the underlying health condition and the

countering effects of drug treatments (e.g., statins) might have rendered the overall TPM_{10}–HRV/HRD relationship null.

Additionally, our findings support the hypothesis that TPM_{10} might impact in part through oxidative stress pathways. We found significant associations between TPM_{10} and HRV/HRD parameters in subjects with homozygous GSTM1 gene deletion (as shown by SDNN and slope β).

Finally, the fact that adverse effects of TPM_{10} were revealed in subjects without cardiovascular morbidity only by HRD parameters supports the hypothesis that measuring changes in complexity in heart rate dynamics in response to exposure to environmental elements, might unveil subtle but important changes in the regulatory mechanisms of the cardiovascular system not detectable by traditional analysis methods.

4.5.2. Strengths and weaknesses of the study (internal validity)

To the best of our knowledge, this is the first study examining the influence of low-level, but long-term, particulate matter air pollution exposure on parameters describing the HRV and heart rate dynamics (using nonlinear time series analysis methods). Additional strengths of the present study included the population-based design, involving a random sample of the Swiss population; the large number of participants; and the detailed information available on participants, allowing for the control of most potential confounders.

A limitation of this study was the absence of a physiological interpretation of the parameters calculated with methods from nonlinear dynamics. Physiological interpretation of such metrics constitutes a major limitation for their use (1996, Goldberger, Amaral et al. 2000, Francesco, Maria Grazia et al. 2012, Manor and Lipsitz 2013). Though it is reasonable to assume that these concepts from mathematics could help gain insight into mechanisms underlying systems fluctuation behavior (e.g., modulations of heart period), efforts are needed to improve our understanding of their physiological correlates. In the present study, this uncertain knowledge limited the interpretation of associations between parameters, and their translation into risk of cardiac events. Another limitation was the small sample size of some subgroups, as well as the low prevalence of some genotypes, which limited statistical power of the explanatory analyses.

4.5.3. Strengths and weaknesses of the study compared to other studies (external validity)

To the best of our knowledge, the association between long-term traffic-related particulate matter exposure and HRV has only been examined in the SAPALDIA cohort study (TPM_{10} levels averaged over a 10 year period) (Adam, Felber Dietrich et al. 2012, Adam, Imboden et al. 2014), and by Mordukhovich et al., who evaluated sub-chronic (3-84 days) and longer-term (1 year) $PM_{2.5}$ or black carbon (a marker of traffic pollution) exposure in relation to HRV (Mordukhovich, Coull et al. 2015). These studies did not observe any consistent overall association.

Interestingly, in the present study, by examining the TPM_{10}-HRV/HRD relationship in the subgroup of subjects without cardiovascular morbidity (i.e., no hypertension or heart disease), we observed significant changes in the heart rate dynamics, whereas we found no significant association in the subgroup with cardiovascular morbidity. These findings corroborate those from Barclay et al. who did not observe any hematological or electrocardiogram response to ambient air pollution in patients with cardiac failure, thought to be a susceptible group (Barclay, Miller et al. 2009), in contrast to their earlier findings in healthy elderly people (Seaton, Soutar et al. 1999). They concluded that modern cardiac therapy was likely to give a measure of protection against the adverse cardiac effects of pollution.

Several studies have provided evidence that the relation between HRV and cardiovascular drug therapies varies and depends on the type of therapy. Adam et al. observed that the adjusted HRV of subjects treated with ACE inhibitors or beta blockers was generally increased, while the HRV of subjects treated with angiotensin receptor blockers, calcium channel blockers, or diuretics, was decreased when compared with the average HRV levels of participants without any heart medication intake (Adam, Felber Dietrich et al. 2012). Furthermore, they provided suggestive evidence that participants under ACE inhibitor treatment may represent a specific subgroup susceptible to the adverse effects of TPM_{10} on HRV. In some other studies, beta blockers (Gold, Litonjua et al. 2000, Park, O'Neill et al. 2005), calcium channel blockers (Park, O'Neill et al. 2005) and statins (Schwartz, Park et al. 2005) have been shown to attenuate the effects of air pollutants; while another study found no evidence of effect modification by beta blockers (Schwartz, Litonjua et al. 2005). These findings suggest that response to long-term TPM_{10} exposure might result from the relative contribution of both the underlying cardiovascular condition and the countering effects of

drug treatments, and might therefore explain the heterogeneous effects of short- and long-term PM air pollution found in subjects with a cardiovascular morbidity (Holguin, Tellez-Rojo et al. 2003, Chuang, Chan et al. 2005, Park, O'Neill et al. 2005, Schwartz, Litonjua et al. 2005, Pieters, Plusquin et al. 2012, Buteau and Goldberg 2016).

The TPM_{10}-HRV/HRD relationship became even more visible while we investigated healthier subjects in the subgroup of subjects without cardiovascular morbidity (i.e., non-obese subjects, and subjects without diabetes). These findings are consistent with those from Yingying et al., who found greater reductions in HRV in relation to PM_{10} exposure in subjects with low Framingham risk score (i.e., low global cardiac risks) (Feng, Huang et al. 2015).

Finally, we found strong and significant associations in subjects with homozygous GSTM1 gene deletion, which is in the line with previous studies that provided evidences that air pollutants might impact in part through oxidative stress pathways (Schwartz, Park et al. 2005, Probst-Hensch, Imboden et al. 2008, Pieters, Plusquin et al. 2012).

4.5.4. Relevance of the study results and implications for policymakers

First, this study provides evidence of the adverse effects of long-term exposure to TPM_{10} on HRV and heart rate dynamics in healthy subjects, believed to be less susceptible than specific subpopulations with morbidities (e.g., the elderly, patients with preexisting cardiovascular disease or diabetes, obese subjects) though. This constitutes a strong argument for health policy makers advocating for more intensive prevention campaigns aimed at reducing traffic-related pollution. However, further studies are needed to see whether these alterations in HRV/HRD in healthy people lead to increased mortality and morbidity later in life.

Second, this study provides evidence that the TPM_{10}-HRV/HRD relationship in subjects with cardiovascular morbidity might be modified by both the underlying cardiovascular condition and the related treatments. Thus, some cardiac therapies, for a given underlying cardiovascular condition, might be protective against the adverse cardiac effects of pollution, whereas some other cardiac therapies/conditions might render subjects particularly susceptible to those effects. Further studies investigating the TPM_{10}-HRV/HRD relationship in subjects with cardiovascular morbidity are necessary.

4.6. Conclusion

In conclusion, findings from the present study indicate that long-term exposure to TPM_{10}, even at low level, triggers adverse changes in the regulation of the cardiovascular system and in the heart rate dynamics. These adverse effects were more visible in healthy subjects, in whom the overall TPM_{10}-HRV/HRD relationship was not modified by an underlying health condition and the eventual countering effects of related drug treatments. Therefore, our findings constitute a strong argument for health policy makers advocating for more intensive prevention campaigns aimed at reducing traffic-related pollution. Finally, we provide some evidence that subjects with homozygous GSTM1 gene deletion might be more susceptible to the effects of TPM_{10}.

4.7. Acknowledgments

Current SAPALDIA Team:

Study directorate: N.M. Probst-Hensch (PI; e/g); T. Rochat (p), C. Schindler (s), N. Künzli (e/exp), J.M. Gaspoz (c)

Scientific team: J.C. Barthélémy (c), W. Berger (g), R. Bettschart (p), A. Bircher (a), C. Brombach (n), P.O. Bridevaux (p), L. Burdet (p), D. Felber Dietrich (e), M. Frey (p), U. Frey (pd), M.W. Gerbase (p), D. Gold (e), E. de Groot (c), W. Karrer (p), F. Kronenberg (g), B. Martin (pa), A. Mehta (e), D. Miedinger (o), M. Pons (p), F. Roche (c), T. Rothe (p), P. Schmid-Grendelmeyer (a), D. Stolz (p), A. Schmidt-Trucksäss (pa), J. Schwartz (e), A. Turk (p), A. von Eckardstein (cc), E. Zemp Stutz (e).

Scientific team at coordinating centers: M. Adam (e), I. Aguilera (exp), S. Brunner (s), D. Carballo (c), S. Caviezel (pa), I. Curjuric (e), A. Di Pascale (s), J. Dratva (e), R. Ducret (s), E. Dupuis Lozeron (s), M. Eeftens (exp), I. Eze (e), E. Fischer (g), M. Foraster (e), M. Germond (s), L. Grize (s), S. Hansen (e), A. Hensel (s), M. Imboden (g), A. Ineichen (exp), A. Jeong (g), D. Keidel (s), A. Kumar (g), N. Maire (s), A. Mehta (e), R. Meier (exp), E. Schaffner (s), T. Schikowski (e), M. Tsai (exp)

(a) allergology, (c) cardiology, (cc) clinical chemistry, (e) epidemiology, (exp) exposure, (g) genetic and molecular biology, (m) meteorology, (n) nutrition, (o) occupational health, (p) pneumology, (pa) physical activity, (pd) pediatrics, (s) statistics

The study could not have been done without the help of the study participants, technical and administrative support and the medical teams and field workers at the local study sites. **Local fieldworkers:** Aarau: S. Brun, G. Giger, M. Sperisen, M. Stahel, Basel: C. Bürli, C. Dahler, N. Oertli, I. Harreh, F. Karrer, G. Novicic, N. Wyttenbacher, Davos: A. Saner, P. Senn, R. Winzeler, Geneva: F. Bonfils, B. Blicharz, C. Landolt, J. Rochat, Lugano: S. Boccia, E. Gehrig, M.T. Mandia, G. Solari, B. Viscardi, Montana: A.P. Bieri, C. Darioly, M. Maire, Payerne: F. Ding, P. Danieli A. Vonnez, Wald: D. Bodmer, E. Hochstrasser, R. Kunz, C. Meier, J. Rakic, U. Schafroth, A. Walder.

Administrative staff: N. Bauer Ott, C. Gabriel, R. Gutknecht.

Research funding:

The Swiss National Science Foundation (grants no 33CS30-148470/1, 33CSCO-134276/1, 33CSCO-108796, 3247BO-104283, 3247BO-104288, 3247BO-104284, 3247-065896, 3100-

059302, 3200-052720, 3200-042532, 4026-028099, PMPDP3_129021/1, PMPDP3_141671/1), the Federal Office of Public Health (Tobacco Control Fund: 07.00577 and REF-1014-50999), the Federal Office for the Environment, the Federal Office of Roads and Transport, the canton's government of Aargau, Basel-Stadt, Basel-Land, Geneva, Luzern, Ticino, Valais, and Zürich, the Swiss Lung League, the canton's Lung League of Basel Stadt/ Basel Landschaft, Geneva, Ticino, Valais, Graubünden and Zurich, Stiftung ehemals Bündner Heilstätten, SUVA, Freiwillige Akademische Gesellschaft, UBS Wealth Foundation, Talecris Biotherapeutics GmbH, Abbott Diagnostics, European Commission 018996 (GABRIEL), Wellcome Trust WT 084703MA.

We thank K. Hugentobler for language editing.

4.8. Online Supplement

Methods

HRV measurements and measures of heart rate dynamics

The following time series analysis parameters were calculated using our own implementations in R (occasionally accessing C libraries to reduce run time) of well-known algorithms. Many of the implementations are based on the TISEAN package (Hegger, Kantz et al. 1999):

- A power-law relationship between the power spectral density (PSD) of the interbeat interval time series and frequency was determined by estimating the slope β of the linear best-fit of the PSD as a function of the frequency on a double logarithmic scale. Several parameters were obtained, depending on the range of frequencies used. The power spectral density (PSD) was estimated according to the method described in (Cusenza 2010). Regressing the power spectral for frequencies in the range 0.01 > freq > 0.0001 yielded the parameter PSD_1. Analogously, PSD_2 corresponds to 0.04 > freq > 0.02, PSD_3 to 0.45 > freq > 0.0001, PSD_4 to 0.5 > freq > 0.1, PSD_5 to 0.5 > freq > 0.2, and PSD_6 to 0.5 > freq > 0.3, respectively. According to our previous work on heart rate dynamics and smoking exposure, we considered PSD_5 in the present study (Girard, Delgado-Eckert et al. 2015).
- The Largest Lyapunov exponent λ was calculated using an embedding dimension of two, and a time lag or delay time of one sample, considering at least 2000 reference points, and adjusting the neighbourhood size ε to obtain at least 10 neighbours per reference point, such that no neighbour was a direct chronological successor of the given reference point (i.e., the Theiler window was set to 2). The length of the embedding space trajectories compared for the estimation of λ was of 20. The algorithm used is described in (Hegger, Kantz et al. 1999, Kantz and schreiber 2004).
- The scaling exponent α obtained via detrended fluctuation analysis (DFA) was calculated using a geometric window increase with exponent equal to 2 and no overlap of windows. Four different time scales were considered: 3-7 samples (α_1), 7-13 samples (α_2), 4-16 samples (α_3), and 16-64 samples (α_4). According to our previous work on heart rate dynamics and smoking exposure, we considered α_3 in the present study (Girard, Delgado-Eckert et al. 2015). The method of DFA is described in (Kantz and schreiber 2004) and the citations therein. The data were detrended by means of a moving average method (Alvarez-Ramirez 2005).

Results

Table E4.1. Association between HRV and heart rate dynamics parameters (outcome variable) and traffic-related PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by sex

Entire study population	**Male (n=590*)**			**Female (n=647*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-5.2 [-21.2;16.7]	0.58		-6.8 [-19.6;7.9]	0.34	0.59
α		-2.4 [-11.8;8.0]	0.64		-6.1 [-17.0;4.7]	0.26	0.89
Slope β	-1.0±4.5	-0.1±0.3	0.68	-4.2±3.5	0.07±0.3	0.79	0.54
Largest Lyapunov exponent	0.7±0.4	-0.01±0.02	0.56	-0.3±0.3	-0.02±0.02	0.53	0.50
Subpopulation without cardiovascular morbidity	**Male (n=180*)**			**Female (n=282*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-30.2 [-50.7;5.1]	0.08		-4.2 [-23.1;19.5]	0.70	0.12
α		-8.1 [-21.9;8.3]	0.31		-6.4 [-16.7;5.2]	0.27	0.72
Slope β	-4.0±7.5	0.1±0.6	0.84	-3.4±5.4	1.2±0.4	<0.001	0.28
Largest Lyapunov exponent	1.5±0.6	-0.09±0.05	0.06	-0.5±0.5	-0.07±0.04	0.06	0.28

All the models are adjusted for age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity (except for the subpopulation without cardiovascular morbidity)

SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

¥ Interaction between TPM_{10} and sex

Table E4.2. Association between HRV and heart rate dynamics parameters (outcome variable) and traffic-related PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by smoking status

Entire study population	**Ever smoker (n=675*)**			**Lifelong non-smoker (n=562*)**			**Interaction[¥]**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-5.8 [-18.8;9.4]	0.43		-6.9 [-22.6;10.8]	0.42	0.51
α		-2.8 [-11.4;6.7]	0.55		-7.0 [-21.8;7.0]	0.33	0.96
Slope β	-1.3±4.0	0.1±0.3	0.60	-6.9±3.8	-0.3±0.3	0.36	0.73
Largest Lyapunov exponent	0.3±0.3	-0.009±0.03	0.75	-0.04±0.3	0.0009±0.03	0.97	0.86
Subpopulation without cardiovascular morbidity	**Ever smoker (n=248*)**			**Lifelong non-smoker (n=214*)**			**Interaction[¥]**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-11.2 [-31.5;15.0]	0.37		-22.9 [-40.3;-0.4]	0.05	0.51
α		1.2 [-11.2;15.3]	0.86		-15.5 [-24.4 ;-1.1]	0.04	0.06
Slope β	-3.9±6.6	1.0±0.5	0.03	-4.6±5.8	0.6±0.5	0.23	0.70
Largest Lyapunov exponent	0.9±0.6	-0.05±0.04	0.20	-0.08±0.5	-0.07±0.04	0.08	0.52

All the models are adjusted for gender, age, BMI, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity (except for the subpopulation without cardiovascular morbidity)

SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

[¥] Interaction between TPM_{10} and smoking status

Table E4.3. Association between HRV and heart rate dynamics parameters (outcome variable) and traffic-related PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by GSTT1 genotype

Entire study population	**Deletion in GSTT1 (n=199*)**			**No deletion in GSTT1 (n=993*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-6.0 [-30.7;26.4]	0.68		-7.1 [-19.0;6.8]	0.29	0.33
α		10.4 [-17.2;39.5]	0.46		-6.0 [-13.7;3.0]	0.17	0.02
Slope β	-0.03±6.6	-0.3±0.4	0.55	-4.5±3.2	0.08±0.3	0.75	0.81
Largest Lyapunov exponent	0.5±0.7	0.08±0.05	0.15	0.2±0.3	-0.02±0.02	0.28	0.08
Subpopulation without cardiovascular morbidity	**Deletion in GSTT1 (n=69*)**			**No deletion in GSTT1 (n=375*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-35.0 [-59.9;21.5]	0.18		-17.5 [-33.3;4.2]	0.10	0.05
α		-8.1 [-23.6;44.7]	0.79		-11.0 [-22.0;0.2]	0.05	0.06
Slope β	-16.9±11.4	0.2±0.9	0.84	-3.2±5.0	0.7±0.4	0.06	0.59
Largest Lyapunov exponent	-0.5±1.0	-0.06±0.09	0.50	0.2±0.4	-0.07±0.03	0.03	0.59

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity (except for the subpopulation without cardiovascular morbidity)

GST, glutathione S-transferase; SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

¥ Interaction between TPM_{10} and GSTT1 genotype

Sensitive analyses

Table E4.4. Association between HRV and heart rate dynamics parameters (outcome variable) and traffic-related PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by ACE inhibitor intake in the entire study population

	ACE inhibitor (n=90*)			**No ACE inhibitor (n=1038*)**			**Interaction[¥]**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-49.9 [-74.4;-6.8]	0.03		-2.6 [-13.6;9.8]	0.66	0.16
α		3.9 [-25.2;69.9]	0.82		-3.9 [-12.4;4.7]	0.36	0.64
Slope β	-5.5±12.8	-0.1±0.8	0.89	-2.8±3.0	-0.001±0.2	0.99	0.68
Largest Lyapunov exponent	1.0±1.4	0.02±0.07	0.83	0.2±0.3	-0.02±0.02	0.33	0.80

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity

ACE inhibitor, angiotensin-converting-enzyme inhibitor; SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

[¥] Interaction between TPM_{10} and ACE inhibitor

Table E4.5. Association between HRV and heart rate dynamics parameters (outcome variable) and traffic-related PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by obesity in the subpopulation without cardiovascular morbidity and without diabetes

	Non-obese (n=408*)			**Obese (n=46*)**			**Interaction[¥]**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-20.9 [-34.6;-1.5]	0.04		27.4 [-26.9;122.0]	0.39	0.01
α		-11.3 [-22.3;-0.5]	0.04		33.6 [-4.1;86.3]	0.09	0.09
Slope β	-0.4±5.0	0.8±0.3	0.02	-21.9±26.5	1.8±1.2	0.15	0.56
Largest Lyapunov exponent	-0.02±0.4	-0.08±0.03	0.01	2.8±2.2	0.05±0.1	0.61	0.11

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, and occupational exposure

SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

[¥] Interaction between TPM_{10} and obesity

Table E4.6. Association between HRV and heart rate dynamics parameters (outcome variable) and traffic-related PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by GSTT1 genotype in the subpopulation without cardiovascular morbidity and without diabetes

	Deletion in GSTT1 (n=69*)			**No deletion in GSTT1 (n=367*)**			**Interaction[¥]**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-35.0 [-59.9;21.5]	0.18		-19.0 [-34.8;2.2]	0.07	0.05
α		-8.1 [-23.6;44.7]	0.79		-11.3 [-22.3;-0.1]	0.05	0.05
Slope β	-16.9±11.4	0.2±0.9	0.84	-2.8±5.1	0.8±0.4	0.05	0.64
Largest Lyapunov exponent	-0.5±1.0	-0.06±0.09	0.50	0.2±0.4	-0.08±0.03	0.03	0.59

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, and occupational exposure

GST, glutathione S-transferase; SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

[¥] Interaction between TPM_{10} and GSTT1 genotype

Investigation of the PM_{10}-HRV/HRD relationship

Air pollutant exposure estimation

PM_{10} estimates were obtained over ten years (1990-2000) using a dispersion modeling approach (Liu, Curjuric et al. 2007). The same as for investigations of traffic-related $PM_{10,}$ mean of the 10 indicators was used to obtain the average concentration of PM_{10} over 10 years.

Results

Table E4.7. Association between HRV and heart rate dynamics parameters (outcome variable) and PM_{10} in linear mixed effects regression models (random intercepts for study area)

	Entire study population (n=1237*)		
	intercept	%GM, 95%CI or coefficient±SE	p-value
SDNN		-1.1 [-3.6;1.3]	0.35
α		-0.7 [-3.4;1.1]	0.47
Power spectral density	-4.2±2.8	-0.04±0.04	0.25
Largest Lyapunov exponent	0.2±0.2	-0.0008±0.004	0.84

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity

SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

Table E4.8. Association between HRV and heart rate dynamics parameters (outcome variable) and PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by the presence of cardiovascular morbidity

	Cardiovascular morbidity (n=775*)			**No cardiovascular morbidity (n=462*)**			**Interaction[¥]**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		0.4 [-2.8;3.2]	0.77		-2.6 [-6.0;1.0]	0.14	0.21
α		0.4 [-2.5;3.0]	0.79		-2.0 [-5.7;0.04]	0.06	0.11
Power spectral density	-3.5±3.7	-0.1±0.005	0.02	-3.8±4.4	0.1±0.6	0.02	0.01
Largest Lyapunov exponent	0.3±0.3	0.006±0.004	0.19	0.4±0.4	-0.009±0.005	0.06	0.05

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, and diabetes

SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

[¥] Interaction between TPM_{10} and cardiovascular morbidity

Table E4.9. Association between HRV and heart rate dynamics parameters (outcome variable) and PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by sex

Entire study population	**Male (n=590*)**			**Female (n=647*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-1.3 [-5.3;2.7]	0.48		-1.3 [-4.0;1.4]	0.33	0.75
α		-0.6 [-2.4;1.3]	0.55		-0.6 [-3.9;1.6]	0.64	0.23
Power spectral density	-1.2±4.5	-0.1±0.05	0.08	-4.1±3.5	0.02±0.05	0.69	0.11
Largest Lyapunov exponent	0.7±0.4	-0.002±0.004	0.67	-0.3±0.3	-0.0009±0.005	0.84	0.38
Subpopulation without cardiovascular morbidity	**Male (n=180*)**			**Female (n=282*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-6.2 [-11.3;0.3]	0.06		-1.5 [-5.2;2.3]	0.44	0.15
α		-1.8 [-4.9;0.8]	0.17		-1.5 [-3.5;0.5]	0.14	0.66
Power spectral density	-4.0±7.5	-0.01±0.09	0.88	-3.4±5.4	0.2±0.07	<0.001	0.05
Largest Lyapunov exponent	1.5±0.6	-0.02±0.008	0.04	-0.4±0.5	-0.008±0.006	0.23	0.16

All the models are adjusted for age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity (except for the subpopulation without cardiovascular morbidity)

SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

¥ Interaction between TPM_{10} and sex

Table E4.10. Association between HRV and heart rate dynamics parameters (outcome variable) and PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by smoking status

Entire study population	**Ever smoker (n=675*)**			**Lifelong non-smoker (n=562*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-1.0 [-3.6;1.8]	0.49		-0.9 [-4.9;2.4]	0.59	0.61
α		-0.6 [-2.3;1.1]	0.48		-0.8 [-6.5;2.1]	0.64	0.57
Power spectral density	-1.7±4.0	-0.04±0.05	0.38	-6.7±3.8	-0.04±0.06	0.42	0.65
Largest Lyapunov exponent	0.3±0.3	-0.004±0.006	0.48	-0.03±0.3	0.005±0.005	0.33	0.32
Subpopulation without cardiovascular morbidity	**Ever smoker (n=248*)**			**Lifelong non-smoker (n=214*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-1.7 [-6.0;2.8]	0.45		-4.8 [-8.9;-0.6]	0.03	0.35
α		-1.0 [-3.3 ;1.3]	0.37		-1.5 [-3.8;0.8]	0.20	0.35
Power spectral density	-4.7±6.7	0.1±0.08	0.13	-4.3±5.8	0.1±0.08	0.07	0.72
Largest Lyapunov exponent	0.9±0.6	-0.01±0.007	0.15	-0.1±0.5	-0.01±0.007	0.16	0.89

All the models are adjusted for gender, age, BMI, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity (except for the subpopulation without cardiovascular morbidity)

SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

¥ Interaction between TPM_{10} and smoking status

Table E4.11. Association between HRV and heart rate dynamics parameters (outcome variable) and PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by obesity

Entire study population	**Non-obese (n=974*)**			**Obese (n=263*)**			**Interaction[¥]**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-1.2 [-3.9;1.7]	0.38		-2.1 [-6.9;3.0]	0.41	0.75
α		-0.8 [-3.0;1.0]	0.38		0.4 [-3.3;4.1]	0.85	0.88
Power spectral density	-4.5±3.5	-0.05±0.04	0.25	-12.4±7.2	0.03±0.09	0.70	0.20
Largest Lyapunov exponent	-0.3±0.3	-0.0002±0.004	0.96	1.1±0.7	-0.001±0.008	0.87	0.48
Subpopulation without cardiovascular morbidity	**Non-obese (n=415*)**			**Obese (n=47*)**			**Interaction[¥]**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-3.8 [-7.1;-0.04]	0.05		7.5 [-2.7;18.7]	0.15	0.03
α		-2.5 [-6.2;-0.4]	0.02		5.1 [-1.1;11.7]	0.10	0.05
Power spectral density	-0.6±4.9	0.2±0.06	0.01	-25.0±28.2	0.1±0.2	0.54	0.35
Largest Lyapunov exponent	-0.09±0.4	-0.01±0.005	0.01	3.8±2.2	0.03±0.02	0.15	0.03

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity (except for the subpopulation without cardiovascular morbidity)

SDNN, standard deviation of all NN intervals

Values shown are as percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

[¥] Interaction between TPM_{10} and obesity

Table E4.12. Association between HRV and heart rate dynamics parameters (outcome variable) and PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by GSTM1 genotype

Entire study population	**Deletion in $GSTM_1$ (n=620*)**			**No deletion in $GSTM_1$ (n=572*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-0.9 [-4.1;2.2]	0.54		-1.4 [-4.2;1.4]	0.32	0.91
α		0.01 [-2.5;2.4]	0.99		-1.2 [-4.6;1.1]	0.31	0.77
Power spectral density	-5.4±4.0	-0.03±0.06	0.60	0.6±4.1	-0.01±0.06	0.79	0.61
Largest Lyapunov exponent	0.2±0.3	-0.003±0.006	0.61	0.1±0.3	-0.0008±0.005	0.87	0.52
Subpopulation without cardiovascular morbidity	**Deletion in $GSTM_1$ (n=227*)**			**No deletion in $GSTM_1$ (n=217*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-4.4 [-8.9;0.7]	0.08		-2.3 [-6.5;2.0]	0.29	0.29
α		-1.1 [-4.0;1.3]	0.36		-3.0 [-6.6;-0.5]	0.02	0.80
Power spectral density	1.2±5.8	0.2±0.08	0.03	0.4±6.4	0.1±0.08	0.14	0.79
Largest Lyapunov exponent	1.2±0.6	-0.009±0.007	0.22	-0.2±0.6	-0.01±0.007	0.08	0.98

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity (except for the subpopulation without cardiovascular morbidity)

GST, glutathione S-transferase; SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

¥ Interaction between TPM_{10} and GSTM1

Table E4.13. Association between HRV and heart rate dynamics parameters (outcome variable) and PM_{10} in linear mixed effects regression models (random intercepts for study area) stratified by GSTT1 genotype

Entire study population	**Deletion in $GSTT_1$ (n=199*)**			**No deletion in $GSTT_1$ (n=993*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-3.2 [-8.8;2.7]	0.28		-0.8 [-3.6;2.0]	0.57	0.65
α		6.3 [1.4;10.6]	0.02		-1.5 [-3.8;0.2]	0.08	<0.001
Power spectral density	-0.07±6.6	0.01±0.09	0.87	-4.8±3.2	-0.04±0.04	0.38	0.63
Largest Lyapunov exponent	-0.5±0.6	0.02±0.009	0.02	0.2±0.3	-0.003±0.004	0.34	0.01
Subpopulation without cardiovascular morbidity	**Deletion in $GSTT_1$ (n=69*)**			**No deletion in $GSTT_1$ (n=375*)**			**Interaction¥**
	intercept	%GM, 95%CI or coefficient±SE	p-value	intercept	%GM, 95%CI or coefficient±SE	p-value	p-value
SDNN		-6.8 [-16.6;3.0]	0.16		-2.9 [-6.7;1.2]	0.15	0.53
α		3.7 [-2.2;9.1]	0.19		-2.9 [-6.8;-0.8]	0.01	0.02
Power spectral density	-14.5±10.6	0.4±0.2	0.01	-3.4±5.0	0.1±0.06	0.10	0.23
Largest Lyapunov exponent	-0.3±0.9	-0.006±0.02	0.71	0.2±0.4	-0.01±0.006	0.06	0.52

All the models are adjusted for gender, age, BMI, smoking status, environmental tobacco smoke exposure, alcohol consumption, physical activity, uric acid, high-sensitivity C-reactive protein, noise exposure, seasonal effect, education level, employment category, occupational exposure, diabetes, and cardiovascular morbidity (except for the subpopulation without cardiovascular morbidity)

GST, glutathione S-transferase; SDNN, standard deviation of all NN intervals

Values shown are percent change in geometric mean (GM) and 95% confidence interval (95%CI) or coefficient ± standard error (SE)

*Number of subjects included in the regression models (i.e., with no missing values on variables included in the regression models)

¥ Interaction between TPM_{10} and GSTT1

5. Chapter 3: Novel phenotyping based on lung function fluctuation clusters in asthma and COPD (BIOAIR study)

Delphine Meier-Girard[1], Edgar Delgado-Eckert[1], Maciej Kupczyk[2,3], Anna James[2], Roelinde Middelveld[2], Barbro Dahlén[2], Mina Gaga[4], Nikos M. Siafakas[5], Alberto Papi[6], Leonardo M. Fabbri[7], Guy Joos[8], Klaus F. Rabe[9], Peter Sterk[10], Elisabeth H. Bel[10], Sebastian L. Johnston[11], Pascal Chanez[12], Mark Gjormarkaj[13], Peter H. Howarth[14], Ewa Niżankowska-Mogilnicka[15], Sven-Erik Dahlén[2], Urs Frey[1]

On behalf of the BIOAIR (Longitudinal Assessment of Clinical Course and BIOmarkers in Severe Chronic AIRway Disease) and ChAMP consortium

[1]University of Basel, University Children's Hospital (UKBB), Basel, Switzerland, [2]Center for Allergy Research, Karolinska Institutet, Stockholm, Sweden; [3]Department of Internal Medicine, Asthma and Allergy, Medical University of Lodz, Poland; [4]University of Athens, Greece; [5]University of Crete, Greece; [6]University of Ferrara, Italy; [7]University of Modena, Italy; [8]University of Gent, Belgium; [9]Christian Alberts University of Kiel, Germany; [10]Amsterdam Medical Centre, The Netherlands; [11]The Imperial College of Science and Technology, London, UK; [12]University of Marseilles, France; [13]Italian Research Council, Palermo, Italy; [14]University of Southampton, UK; [15]The Jagellonian University, Krakow, Poland

Corresponding author: Urs Frey, Pediatric Pulmonology Research Group, University Children's Hospital, Spitalstrasse 33, 4056 Basel, Switzerland. Phone: +41 61 704 29 01, Fax: +41 61 704 12 13. E-mail address: urs.frey@ukbb.ch

This article will be submitted to

the American Journal of Critical Care and Respiratory Medicine

5.1. At a glance commentary

Scientific Knowledge on the Subject: Phenotyping appears especially relevant in severe asthma, COPD and the transition forms between these entities, in which the heterogeneity of response to drug therapy and the unpredictable nature of exacerbations are a major clinical challenge. For clinicians, identification of phenotypes related to specific treatable traits is of primary concern. Airway function dynamics are at the intersection between pathophysiological mechanisms and the expression of particular clinical patterns or treatment responses. Consequently, investigation of lung function fluctuation might give new insight into the relationship between specific pathological features and clinically meaningful outcomes.

What This Study Adds to the Field: The present study uses a novel clustering approach, based on the fluctuations of a single lung function parameter, namely, the twice-daily FEV_1 recorded over one year. We identify five phenotypes, of those three distinct phenotypes of severe asthma, in which the progressive functional alteration of the lung corresponds to a gradually increasing clinical severity and translates into specific risks of exacerbation and treatment response features. Such phenotypes might help identify patients who may benefit from different treatment strategies, further clinical investigations in a referral center, and/or closer monitoring, for example in a telemonitoring setting.

5.2. Abstract

Rationale: Identification of phenotypes related to specific treatable traits is of primary concern in asthma and COPD. Airway function dynamics are at the intersection between pathophysiological mechanisms and the expression of particular clinical patterns or treatment responses. Consequently, investigation of lung function fluctuation might give new insight into the relationship between specific pathological features and clinically meaningful outcomes.

Objective: To evaluate whether the subgrouping of patients with obstructive airway diseases, including mild-to-moderate asthma, severe asthma, and COPD, according to their pattern of lung function fluctuation, allows for the identification of phenotypes with specific treatable traits.

Methods: We conducted a time series clustering analysis based on the fluctuation of twice-daily FEV_1 measurements recorded over a one year period in a mixed group of 134 adults with mild-to-moderate asthma, severe asthma, or COPD from the longitudinal Pan-European BIOAIR study.

Measurements and Main Results: We identified a group of mild-to-moderate asthmatics (M), three distinct groups of severe asthmatics (S_1, S_2, S_3) and a group with COPD patients (C). These 5 groups presented a gradually increasing clinical severity and functional alteration of the lung (from M to C) and identified phenotypes of patients with high exacerbation risks (S_2, S_3), of patients high likelihood to respond to steroids (S_2, S_3), but also of patients with severe functional alterations (C) disallowing a clinical response to steroids despite an appropriate cellular anti-inflammatory response.

Conclusions: Lung function fluctuation based clustering identifies phenotypes of severe asthmatics in which the functional alteration of the lung translates into specific risks of exacerbation and treatment response features.

5.3. Introduction

Asthma and COPD are increasingly recognized as entities in a continuum of heterogeneous obstructive airway disease (Zeki, Schivo et al. 2011, Carolan and Sutherland 2013, 2015) with distinct phenotypes (Siroux and Garcia-Aymerich 2011, Wenzel 2012). For clinicians, there is a great need to identify phenotypes with direct relevance to choice of treatment and risk of worsening. Especially in severe asthma (Heaney and Robinson 2005, Wenzel 2012), COPD, and the transition forms between these entities (Chung 2013), in which the unpredictable nature of exacerbations and the heterogeneity of response to drug therapy are a major clinical challenge (Moore and Peters 2006, Donaldson, Seemungal et al. 2012, Kupczyk, Haque et al. 2013, Phipatanakul, Mauger et al. 2016). Clinical treatment success is determined both by anti-inflammatory and lung functional response to bronchodilators. It is still poorly understood why some patients benefit more from step-up treatment with long-acting bronchodilators than from anti-inflammatory treatment; and why certain patients have a poor overall response.

The information content in airway function dynamics is high and largely underestimated. It is recognized that airway function dynamics are influenced by airway inflammation, but also independently by mechanical factors in the lung and airways (Tschumperlin and Drazen 2006). For instance, rapid bronchial obstruction, due to exaggerated bronchial responsiveness, contributes to a specific dynamic behavior, and such patients might be clinically characterized by a high exacerbation risk. On the other hand, irreversible obstruction, due to mechanical impairments, contributes to another specific dynamic behavior, and such patients might be clinically characterized by a poor response to bronchodilator due to the mechanical impairment (Stern, de Jongste et al. 2011, Thamrin, Frey et al. 2016). That suggests that functional alterations are reflected in lung function fluctuation and might translate into particular clinical patterns or treatment responses. Indeed, lung function fluctuation has been found to be associated with disease progression and control (Frey and Suki 2008, Thamrin, Nydegger et al. 2011), risk of exacerbations (Frey, Brodbeck et al. 2005, Thamrin, Zindel et al. 2011, Donaldson, Seemungal et al. 2012), and treatment response (Kaminsky, Wang et al. 2016). Therefore, its characterization might give new insights into the relationship between specific pathological features and clinically meaningful outcomes.

Consequently, we hypothesized that a lung function fluctuation based clustering (FBC) approach (Delgado, Kumar et al. 2015) might help identify subgroups of patients with distinct lung functional abnormalities, which may be related to specific treatable traits (i.e., specific

treatable mechanisms related to specific symptom features). The FBC method identifies clusters of patients with similar patterns of lung function fluctuation over a predetermined window of observation. We conducted the FBC analysis in a mixed group of 134 adults with mild-to-moderate asthma, severe asthma, or COPD, with a unique one-year collection of twice-daily lung function data, from the longitudinal European BIOAIR (Longitudinal Assessment of Clinical Course and BIOmarkers in Severe Chronic AIRway Disease) multicenter study.

The aim of this study was to assess whether the subgrouping of patients with obstructive airway diseases, including mild-to-moderate asthma, severe asthma, and COPD, according to their pattern of lung function fluctuation, allows for the identification of phenotypes with specific treatable traits.

5.4. Methods

5.4.1. Study design

This is a post-hoc analysis of the Pan-European BIOAIR study which was designed to characterize the course of severe chronic airway diseases over time. The design has been described in detail elsewhere (Kupczyk, Haque et al. 2013). Briefly, 169 adults with asthma, and 64 with COPD were included. Patients with asthma were screened at visit 1 (**Figure 5.1**) and allocated to mild-to-moderate asthma (n=76) and severe asthma (n=93) groups according to established criteria aligned with current guidelines (Kupczyk, Haque et al. 2013, Chung, Wenzel et al. 2014). Patients underwent a treatment optimization period of 4 weeks (from visit 1 to visit 2), which was followed by a 2-week double blind placebo-controlled oral prednisone intervention (0.5 mg/kg body weight, from visit 2 to visit 3) permitting assessment of lung function and biomarker responses to oral corticosteroid (OCS) intervention. Finally, patients were followed up for 12 months (from visit 3 to visit 6).

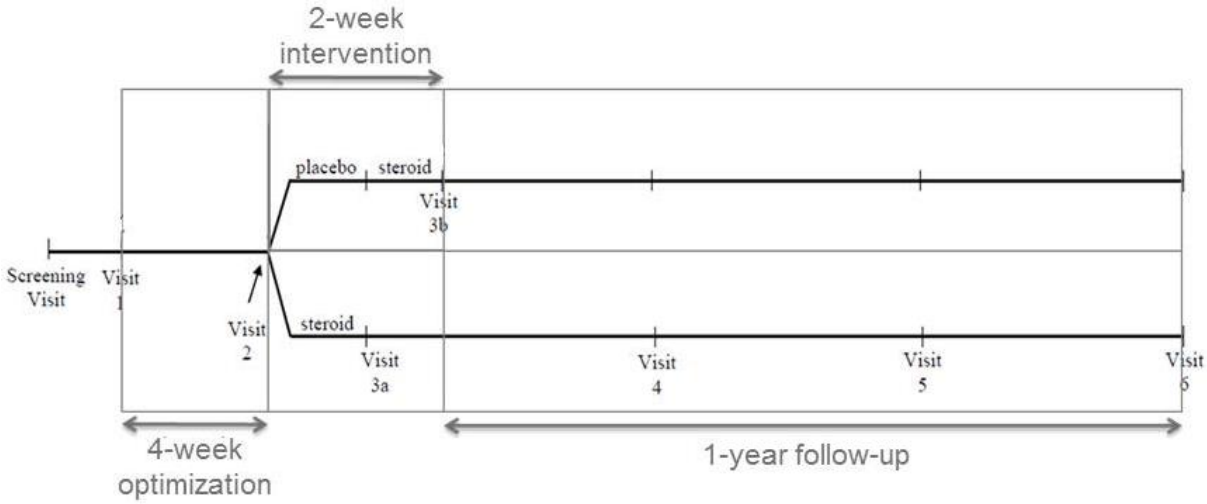

Figure 5.1. Design of the BIOAIR study

5.4.2. Data collection and measurements

Lung function measurements, including FEV_1 and PEF, clinical symptoms and medication use were recorded daily throughout the entire study using an electronic diary (Vitalograph Electronic PEF/FEV1 Diary, XM version, Vitalograph Ltd, Buckingham, UK) (Kupczyk, Haque et al. 2013). Detailed information about lung function measurements performed during visits and questionnaires completed by patients can be found in the **Online Supplement**.

5.4.3. Lung function fluctuation based clustering of patients

Patients with a similar fluctuation behavior in the twice-daily measurements of FEV_1 from the one-year follow-up (visit 3 to visit 6) were grouped into clusters, using the FBC approach (Delgado-Eckert, Fuchs et al. 2017) (**Online Supplement**). The measurements performed during the interventional phases of the study (i.e., treatment optimization and oral steroid intervention) where not included in the clustering analysis. FEV_1 was expressed as the age, sex, height and ethnicity adjusted z-score (denoted $zFEV_1$) (Quanjer, Stanojevic et al. 2012).

5.4.4. Statistical analysis

Results are expressed as numbers and percentages for categorical variables, and as mean (± standard deviation) or median [25^{th} percentile; 75^{th} percentile] for continuous variables, according to their distribution.

In each of the clusters identified, an enrichment analysis was performed using the hypergeometric test (Agresti 1992) in order to assess whether there was a significant enrichment (i.e., over-representation) of mild-to-moderate asthmatics, severe asthmatics, or patients with COPD.

Comparisons between defined groups were provided using the one-way ANOVA or the Kruskal-Wallis test, as appropriate, for continuous variables, and the Chi^2 or the Fisher's exact test, as appropriate, for categorical variables. Post-hoc tests for pair-wise multiple comparisons were performed using the Tukey's test or the Nemenyi test for continuous variables, as appropriate. For categorical variables, the multiple testing issue was addressed using an enrichment analysis (hypergeometric test) combined to a resampling method, setting the family-wise error rate at the 5% level (**Online Supplement**). We were then able to assess whether there was a significant over-representation of a given parameter in a given group as compared to the entire analysis population.

Response to OCS intervention (visit 2 to visit 3) was defined as ≥ 10 percent improvement of predicted FEV_1 (Phipatanakul, Mauger et al. 2016).

All tests were two-sided with a significance level of 0.05. Statistical analysis was performed using R, Version 3.2.1 (2008).

5.5. Results

5.5.1. Description of the analysis population

Among the 233 patients included in the BIOAIR study, 6 were excluded at screening, 12 were lost to follow-up before beginning the 1-year follow-up, and 29 did not perform any self-measurements of FEV_1. Furthermore, 52 patients were excluded from the analysis population because they did not have the minimum number of FEV_1 measurements required for the cluster analysis (**Online Supplement**). Patients excluded (n=99) did not significantly differ from the 134 patients analyzed (**Online Supplement**).

Among the 134 patients analyzed, there were 53 (39.6%) mild-to-moderate asthmatics, 54 (40.3%) severe asthmatics, and 27 (20.1%) patients with COPD. The mean age of the subjects was 51.7±13.6 years. Characteristics at study inclusion, according to the airway disease (i.e., mild-to-moderate asthma, severe asthma, COPD), are summarized in **Table 5.1**. The mean number of FEV_1 measurements per patient during follow-up was 428±170.

Table 5.1. Characteristics of patients at inclusion according to the airway disease (n=134)

	Mild-to-moderate asthma (N=53)	**MD**	**Severe asthma (N=54)**	**MD**	**COPD (N=27)**	**MD**	**p-value***
Clinical characteristics							
Age, years	42.6±12.6	-	50.6±10.5	-	64.8±7.9	-	**<0.001**
Gender, male	20 (37.7%)	-	21 (38.9%)	-	19 (70.4%)	-	**0.01**
BMI, kg/m^2	25.1±4.0	-	28.5±5.1	-	27.0±4.7	-	**<0.001**
Age of disease onset, years	18.0 [5.5;33.0]	2	33.0 [20.3;43.3]	4	60.0 [51.0;66.0]	2	**<0.001**
ACQ, Juniper	0.9 [0.4;1.3]	2	2.0 [1.2;2.7]	4	NA	NA	**<0.001**
QoL, SGRQ	15.4 [10.5;29.6]	11	41.6 [31.9;57.1]	5	39.6 [32.5;50.6]	4	**<0.001**
Atopy	24 (47.1%)	2	20 (40.0%)	4	0 (0%)†	1	0.47‡
Lung function							
Reversibility, percentage change	10.5±6.0	1	8.5±6.0	1	3.0±3.8	-	**<0.001**
FeNO, ppb	32.9 [20.6;51.5]	23	33.9 [13.5;71.0]	28	10.8 [8.05;13.3]	16	**0.004**
FEV_1, z-score	-1.4±1.3	2	-2.0±1.3	1	-3.3±0.7	1	**<0.001**
FEV_1, % predicted	82.2±16.9	2	71.5±19.5	1	45.9±10.8	1	**<0.001**
FVC, z-score	-0.2±0.9	2	-1.1±1.2	1	-1.4±0.8	1	**<0.001**
FVC, % predicted	97.0±12.4	2	85.2±16.6	1	78.2±11.8	1	**<0.001**
FEV_1/FVC	-1.7±1.2	2	-1.7±1.5	1	-3.5±1.1	1	**<0.001**
DLCO, %predicted corrected	94.5±14.5	5	86.0±16.6	8	59.4±20.0	1	**<0.001**
FRC, %predicted corrected	96.1 [82.4;119.7]	6	92.6 [82.2;113.5]	9	126.2 [104.0;147.1]	3	**<0.001**
IVC, %predicted corrected	102.3±14.1	6	96.8±19.1	1	89.4±11.2	3	**0.007**
TLC, %predicted corrected	104.3±12.5	2	103.1±15.4	1	109.2±18.0	-	0.22
RV, %predicted corrected	104.6 [92.5;126.1]	3	118.0 [97.8;139.4]	1	150.7 [111.6;174.1]	-	**0.001**
RV/TLC	1.0 [0.9;1.2]	3	1.2 [1.0;1.4]	1	1.4 [1.2;1.6]	-	**<0.001**
Inflammatory response							
hs-CRP, mg/l	2.0 [1.0;3.6]	5	3.4 [1.6;9.2]	6	6.0 [2.5;8.2]	-	**0.03**
Sputum cells, $\times 10^6$	0.6 [0.2;2.0]	14	0.7 [0.4;2.9]	14	1.1 [0.7;2.4]	12	0.14
Sputum eosinophils, %	1.2 [0.1;7.2]	17	4.8 [0.7;26.6]	16	0.6 [0.03;1.6]	13	**0.002**
Sputum neutrophils, %	40.1 [15.3;58.6]	17	26.1 [15.3;48.1]	16	54.9 [34.8;73.1]	13	0.10

Table 5.1. Characteristics of patients at inclusion according to the airway disease (n=134) (continued)

	Mild-to-moderate asthma (N=53)	**MD**	**Severe asthma (N=54)**	**MD**	**COPD (N=27)**	**MD**	**p-value***
White blood cells, $\times 10^9$/l	6.5±1.6	2	8.2±2.5	1	8.1±2.0	-	**<0.001**
Blood eosinophils, $\times 10^9$/l	0.3 [0.2;0.4]	2	0.3 [0.1;0.5]	2	0.2 [0.1;0.3]	-	0.21
Blood neutrophils, $\times 10^9$/l	3.5 [2.6;4.4]	5	4.8 [3.6;6.7]	3	4.9 [3.8;6.4]	1	**<0.001**

Values shown are mean ± standard deviation, median [25th percentile; 75th percentile], and numbers (percentages)

ACQ, Asthma Control Questionnaire; BMI, body mass index; hs-CRP, high-sensitivity C-reactive protein; D_{LCO}, diffusing capacity of the lung for carbon monoxide; FeNO, fraction of exhaled nitric oxide; FEV_1, forced expiratory volume in one second; FRC, forced residual volume; FVC, forced vital capacity; IVC, inspiratory vital capacity; MD, missing data; NA, not applicable; QoL, quality of life; RV, residual volume; SGRQ, St George's Respiratory Questionnaire; TLC, total lung capacity

*Comparison between groups using the one-way ANOVA or the Kruskal-Wallis test, as appropriate, for continuous variables, and using the Chi^2 or the Fisher's exact test, as appropriate, for categorical variables; †Inclusion criteria; ‡Comparison between mild-to-moderate asthmatics and severe asthmatics

5.5.2. The fluctuation based clustering analysis identifies four clusters

The FBC analysis identified four clusters, which consisted of 12 (9.0%), 49 (36.7%), 31 (23.1%), and 42 (31.3%) subjects.

Cluster 1 consisted of 10 (83.3%) mild-to-moderate asthmatics and 2 (16.7%) severe asthmatics. There was a significant over-representation (i.e., significant enrichment) of mild-to-moderate asthmatics ($p<0.001$); the cluster was therefore labeled "*mild-type lung function fluctuation*" (**Table 5.2**).

Cluster 2 consisted of 29 (59.2%) mild-to-moderate asthmatics, 18 (36.7%) severe asthmatics, and 2 (4.1%) patients with COPD. There was a significant over-representation of mild-to-moderate asthmatics ($p<0.001$), but given the mixture with severe asthmatics and patients with COPD, the cluster was labeled "*moderate-type lung function fluctuation*".

Cluster 3 consisted of 10 (32.3%) mild-to-moderate asthmatics, 16 (51.6%) severe asthmatics, and 5 (16.1%) patients with COPD. There was a significant over-representation of severe asthmatics ($p=0.048$); the cluster was therefore labeled "*severe-type lung function fluctuation*".

Cluster 4 consisted of 4 (9.5%) mild-to-moderate asthmatics, 18 (42.9%) severe asthmatics, and 20 (47.6%) patients with COPD. There was a significant over-representation of patients with COPD ($p<0.001$); the cluster was therefore labeled "*COPD-type lung function fluctuation*".

Table 5.2. Over-representation of airway diseases in each of the clusters identified (n=134)

	Cluster 1 (N=12)	p-value	**Cluster 2** (N=49)	p-value	**Cluster 3** (N=31)	p-value	**Cluster 4** (N=42)	p-value
Airway disease								
Mild-to-moderate asthma	**10 (83.3%)**	**<0.001**	**29 (59.2%)**	**<0.001**	10 (32.3%)	0.77	4 (9.5%)	0.99
Severe asthma	2 (16.7%)	0.93	18 (36.7%)	0.67	**16 (51.6%)**	**0.048**	18 (42.9%)	0.27
COPD	0 (0.0%)	0.94	2 (4.1%)	0.99	5 (16.1%)	0.64	**20 (47.6%)**	**<0.001**

p-value from the hypergeometric test

5.5.3. Phenotyping based on the combination of clinical classification (i.e., mild-to-moderate asthma, severe asthma, and COPD) and lung function fluctuation based clusters unveils three subgroups of severe asthmatics

Most of the mild-to-moderate asthmatics (n=39/53, 73.6%) were assigned to the clusters "*mild-type lung function fluctuation*" and "*moderate-type lung function fluctuation*"; they were defined as group M (**Table 5.3, Figure 5.2**). Mild-to-moderate asthmatics who were assigned to the clusters "*severe-type lung function fluctuation*" (n=10, labelled group MS) and "*COPD-type lung function fluctuation*" (n=4) were not further investigated, due to the small sample size of these groups. Description of group MS is provided in the **Online Supplement**.

Most patients with COPD (n=20/27, 74.0%) were assigned to the cluster "*COPD-type lung function fluctuation*"; they were defined as group C. COPD patients who were assigned to the clusters "*moderate-type lung function fluctuation*" (n=2) and "*severe-type lung function fluctuation*" (n=5) were not further investigated, due to the small sample size of these groups.

Severe asthmatics were spread across the clusters "*moderate-type lung function fluctuation*" (n=18/54, 33.3%), "*severe-type lung function fluctuation*" (n=16/54, 29.6%), and "*COPD-type lung function fluctuation*" (n=18/54, 33.3%). These 3 groups were labeled S_1, S_2, and S_3, respectively. Severe asthmatics who were assigned to the cluster "*mild-type lung function fluctuation*" (n=2) were not further investigated, due to the small sample size of this group.

Description of groups M, S_1, S_2, S_3, and C is provided in **Table 5.4** and **Figures 5.3-5.5**.

Table 5.3. Distribution of mild-to-moderate asthmatics, severe asthmatics, and patients with COPD according to clusters (n=134)

	Cluster 1 (N=12) called «mild-type lung function fluctuation»	**Cluster 2** (N=49) called «moderate-type lung function fluctuation»	**Cluster 3** (N=31) called «severe-type lung function fluctuation»	**Cluster 4** (N=42) called «COPD-type lung function fluctuation»
Mild-to-moderate asthmatics (N=53)	10 (18.9%) **Group M**	29 (54.7%) **Group M**	10 (18.9%)	4 (7.5%)
Severe asthmatics (N=54)	2 (3.7%)	18 (33.3%) **Group S_1**	16 (29.6%) **Group S_2**	18 (33.3%) **Group S_3**
Patients with COPD (N=27)	0	2 (7.4%)	5 (18.5%)	20 (74.1%) **Group C**

Values shown are and numbers (percentages)

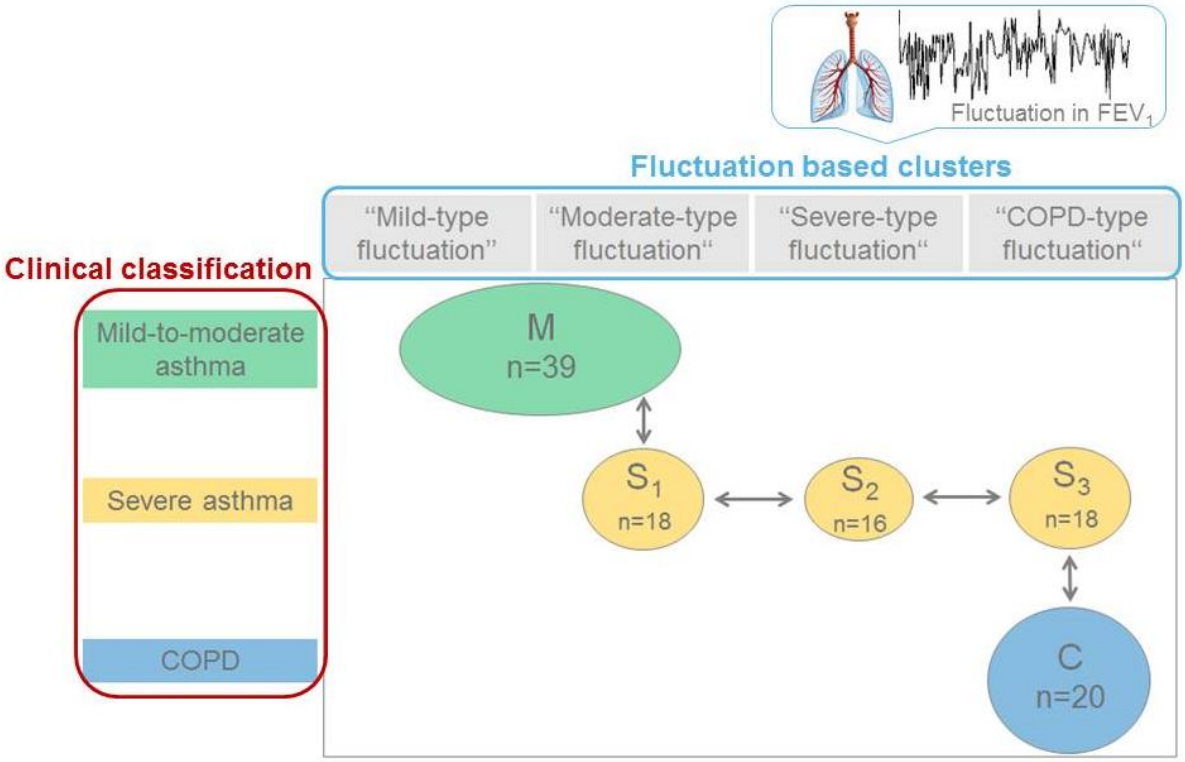

Figure 5.2. Subgrouping of mild-to-moderate asthmatics, severe asthmatics and patients with COPD (i.e., clinical classification) according to their pattern of lung function fluctuation (i.e., data-driven classification)

Table 5.4. Characteristics of patients according to subgroups M, S_1, S_2, S_3, and C (n=111)

	Group M (N=39)	MD	**Group S_1 (N=18)**	MD	**Group S_2 (N=16)**	MD	**Group S_3 (N=18)**	MD	**Group C (N=20)**	MD	**p-value** *
Airway disease	**Mild-to-moderate asthma**		**Severe asthma**		**Severe asthma**		**Severe asthma**		**COPD**		
Clinical characteristics											
Age, years	43.2±13.7 ‡	-	51.2±11.5 ‡	-	53.6±10.3 † ‡	-	47.8±9.3 ‡	-	65.1±7.5 †	-	**<0.001**
Gender, male	14 (35.9%)	-	8 (44.4%)	-	6 (37.5%)	-	7 (38.9%)	-	13 (65.0%)	-	0.28
BMI, kg/m^2	24.9±3.3	-	27.3±4.5	-	28.5±4.9	-	29.9±6.0 †	-	26.9±5.3	-	**0.003**
Age of disease onset, years	18.5 [5.3;32.5] ‡	1	34.0 [24.3;43.8] ‡	-	24.5 [13.3;37.5] ‡	2	35.0 [19.0;45.0] ‡	2	60.5 [52.5;65.5] †	2	**<0.001**
Atopy	18 (47.4%)	1	5 (33.3%)	3	8 (50.0%)	-	7 (38.9%)	-	0 (0%) §	-	0.73 ll
QoL, SGRQ	11.4 [6.8;20.7] ‡	8	39.3 [26.0;42.6] †	3	43.7 [37.7;48.8] †	4	51.1 [33.1;59.9] †	1	40.3 [34.2;56.2] †	7	**<0.001**
ACQ, Juniper	0.3 [0.1;0.6]	-	1.4 [0.9;2.0] †	-	2.0 [1.3;2.6] †	1	2.4 [2.0;3.0] †	2	NA	-	**<0.001**
Number of exacerbation during follow-up	0 [0;1]	-	1 [0;2]	-	1 [1;2]	-	2 [1;2] †	-	0 [0;1]	-	**<0.001**
At least one exacerbation during follow-up	13 (33.3%)	-	10 (55.6%)	-	13 (81.2%)	-	13 (72.2%)	-	8 (40.0%)	-	**0.005**
Lung function											
Reversibility, percentage change	10.2 [7.7;14.1] ‡	1	9.1 [5.6;12.9] ‡	-	5.5 [0.9;9.3]	1	7.6 [4.6;14.9] ‡	-	3.1 [0.8;4.4] †	-	**<0.001**
FeNO, ppb	21.3 [14.3;32.2]	12	21.4 [14.5;42.3]	10	40.5 [23.7;53.8]	7	29.7 [10.1;59.9]	8	11.2 [7.0;13.8]	11	**0.001**
FEV1, z-score	-0.7±1.1 ‡	-	-1.4±0.9 †‡	-	-2.1±1.0 †‡	-	-2.6±1.2 †‡	-	-3.5±0.6 †	-	**<0.001**
FEV_1, % predicted	96.5±15.8 ‡	-	83.4±13.6 †‡	-	73.7±16.9 †‡	-	66.7±19.8 †‡	-	44.6±10.4 †	-	**<0.001**
FVC, z-score	0.04±0.9 ‡	-	-1.0±1.3 †‡	-	-1.3±1.1 †	-	-1.3±1.0 †	-	-2.0±0.7 †	-	**<0.001**
FVC, % predicted	100.5±11.6 ‡	-	86.6±16.2 †‡	-	82.2±14.4 †‡	-	82.3±13.8 †‡	-	70.1±11.1 †	-	**<0.001**
FEV_1/FVC	-1.0±1.2 ‡	-	-0.7±1.4 ‡	-	-1.5±1.4 ‡	-	-2.4±1.4 †‡	-	-3.5±0.9 †	-	**<0.001**
DLCO, %predicted corrected	94.8±13.2 ‡	2	86.5±10.8 ‡	6	91.8±17.7 ‡	-	80.4±18.3 † ‡	1	56.3±21.3 † ‡	-	**<0.001**
FRC, %predicted corrected	101.3±22.3 ‡	2	91.9±24.8 ‡	6	98.2±24.8 ‡	-	99.9±26.5 ‡	2	128.3±30.0 †	2	**<0.001**
IVC, %predicted corrected	107.2±10.2 ‡	4	98.0±13.4	-	95.5±16.8	1	92.3±22.5 †	-	89.8±11.4 †	2	**<0.001**
TLC, %predicted corrected	105.5±10.9	1	97.5±14.4 ‡	1	105.8±12.5	-	103.3±16.7	-	111.3±17.9	-	0.07
RV, %predicted corrected	104.5±29.2 ‡	2	105.0±34.8 ‡	1	130.6±29.3	-	129.5±44.2	-	158.8±46.3	-	**<0.001**
RV/TLC	0.98±0.2 ‡	2	1.1±0.2 ‡	1	1.2±0.2 †	-	1.2±0.3 †	-	1.4±0.2 †	2	**<0.001**
Inflammatory biomarkers											
Sputum cells, ×10^6	0.6 [0.2;1.3]	8	2.2 [0.6;4.0]	6	1.0 [0.4;2.0]	5	1.1 [0.6;3.8]	4	1.1 [0.4;1.5]	7	0.14
Sputum eosinophils, %	1.3 [0.1;4.7]	9	1.5 [0.7;17.8]	6	4.9 [1.2;18.1]	6	4.5 [1.0;10.4]	8	2.1 [0.2;5.8]	9	0.26
Sputum eosinophilia ≥ 2%	12 (40.0%)	9	5 (41.7%)	6	6 (60.0%)	6	6 (60.0%)	8	6 (54.5%)	9	0.68

Table 5.4. Characteristics of patients according to subgroups M, S_1, S_2, S_3, and C (n=111) (continued)

	Group M (N=39)	**MD**	**Group S1 (N=18)**	**MD**	**Group S2 (N=16)**	**MD**	**Group S3 (N=18)**	**MD**	**Group C (N=20)**	**MD**	**p-value ***
Airway disease	**Mild-to-moderate asthma**		**Severe asthma**		**Severe asthma**		**Severe asthma**		**COPD**		
Sputum neutrophils, %	46.0 [17.6;67.0]	9	51.3 [26.7;69.4]	6	25.1 [23.4;58.6]	6	56.5 [45.0;63.6]	8	68.3 [54.2;74.2]	9	0.14
Sputum neutrophilia ≥ 40%	17 (56.7%)	9	7 (58.3%)	6	3 (30.0%)	6	8 (80.0%)	8	10 (90.9%)	9	**0.04**
Mixed granulocytic inflammation, %	6 (20.0%)	9	3 (25.0%)	6	2 (20.0%)	6	5 (50.0%)	8	5 (45.5%)	9	0.27
hs-CRP, mg/l	0.8 [0.4;2.2] ‡	7	2.6 [1.6;2.9]	3	2.2 [0.8;5.1]	1	4.2 [2.3;5.6] †	1	4.0 [2.4;5.9] †	3	**<0.001**
White blood cells, $\times 10^9$/l	6.3±1.5	1	7.0±1.9	-	8.1±2.4 †	-	8.8±2.6 †	-	7.7±2.3	-	**<0.001**
Blood eosinophils, $\times 10^9$/l	0.3 [0.2;0.3]	1	0.3 [0.1;0.4]	-	0.4 [0.1;0.5]	1	0.2 [0.1;0.3]	-	0.2 [0.2;0.3]	-	0.82
Blood neutrophils, $\times 10^9$/l	3.4 [3.0;4.2]	1	3.7 [2.9;5.3]	-	5.9 [3.5;6.3]	-	5.4 [4.3;6.9] †	-	4.4 [3.7;5.5]	-	**<0.001**
Response to treatment											
Response to oral corticosteroids	3 (7.7%)	-	3 (16.7%)	-	4 (25.0%)	-	8 (44.4%)	-	3 (15.0%)	-	**0.02**
Biomarkers											
sRAGE, pg/ml	1602 [1197;1897]	7	1694 [1310;2209]	3	1287 [952.6;2054]	1	1319 [1114;1508]	1	1145 [854;1386]	4	**0.03**
MMP-3 pg/mL	11910 [8062;16200]	7	15840 [5561;22400]	3	19960 [13410;25500]	1	23170 [9236;66400]	1	18680 [14710;26380]	4	**0.04**
DPPIV pg/mL	113200±35220	7	108500±37970	3	87530±34646	1	95320±52699	1	82250±31184	4	**0.05**
YKL-40 (or Chitinase 3-Like 1), pg/ml	15470 [12080;18410] ‡	7	15120 [13100;33000]	3	27720 [19500;41550]	1	15400 [10710;25020] ‡	1	37380 [21420;54580] †	4	**<0.001**

Values shown are mean ± standard deviation, median [25th percentile;75th percentile], and numbers (percentages)

ACQ, Asthma Control Questionnaire; BMI, body mass index; CD40 L, CD 40 ligand; hs-CRP, high-sensitivity C-reactive protein; D_{LCO}, diffusing capacity of the lung for carbon monoxide; FeNO, fraction of exhaled nitric oxide; FEV_1, forced expiratory volume in one second; FRC, forced residual volume; FVC, forced vital capacity; IVC, inspiratory vital capacity; MD, missing data; NA, not applicable; OCS, oral corticosteroid; QoL, quality of life; RV, residual volume; SGRQ, St George's Respiratory Questionnaire; sRAGE, soluble receptor for advanced glycation end products; TLC, total lung capacity

*Comparison between groups using the one-way ANOVA or the Kruskal-Wallis test, as appropriate, for continuous variables, and using the Chi^2 or the Fisher's exact test, as appropriate, for categorical variables; †As compared to group M; ‡As compared to group C; §Inclusion criteria; ll Comparaison between groups M, S_1, S_2, and S_3

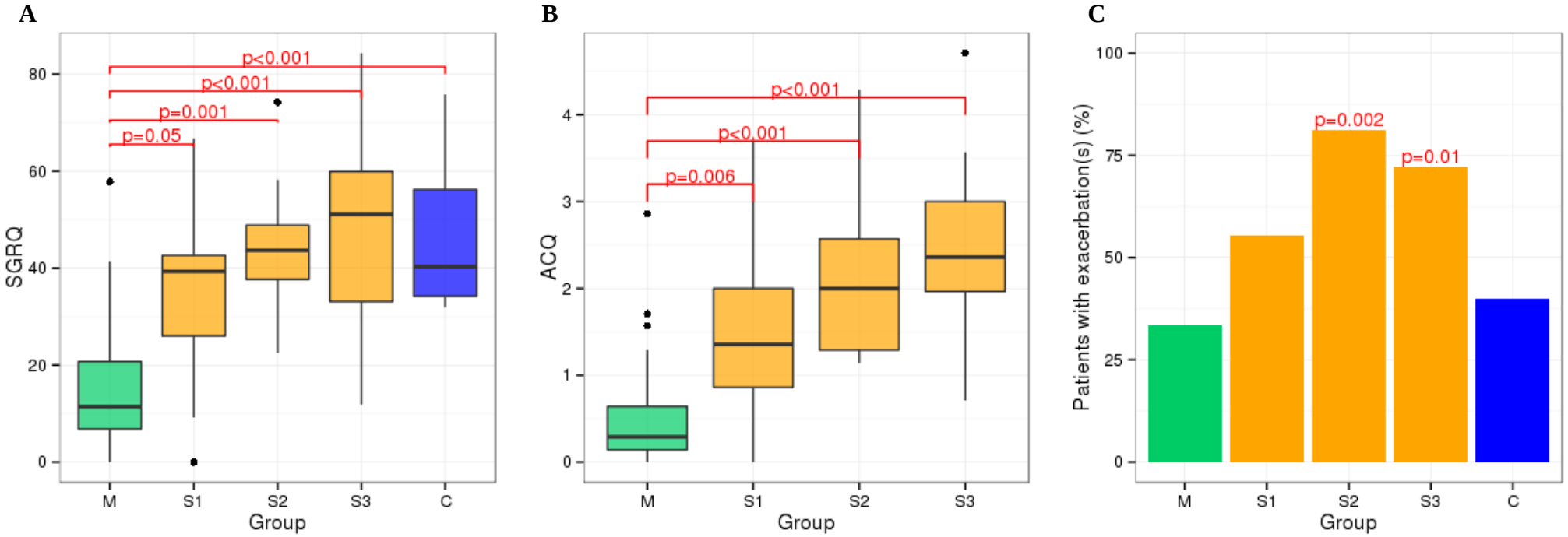

Figure 5.3. Distribution of clinical characteristics according to groups M, S_1, S_2, S_3, and C

(A) Score of St George's Respiratory Questionnaire, (B) Score of Asthma control Questionnaire, (C) Frequency of patients who experienced at least one exacerbation during follow-up

Group M includes mild-to-moderates asthmatics, groups S_1, S_2, and S_3 include severe asthmatics, group C includes COPD patients

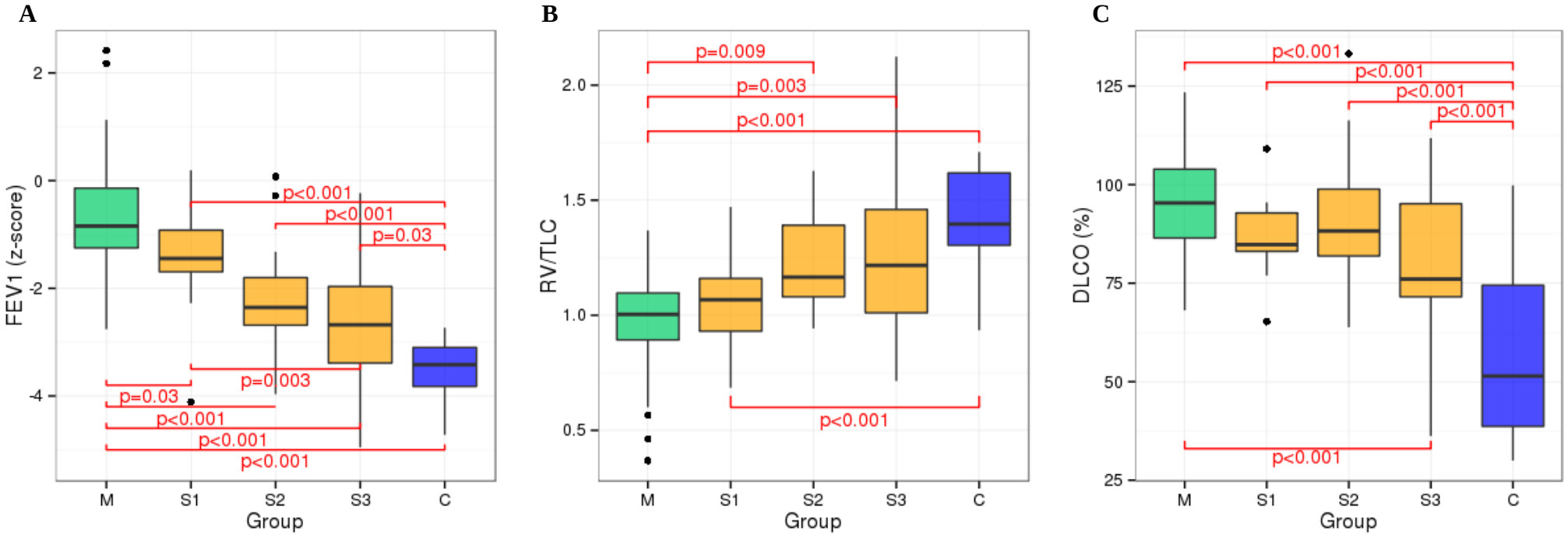

Figure 5.4. Distribution of lung function features according to groups M, S_1, S_2, S_3, and C

(A) FEV_1, (B) RV/TLC, (C) D_{LCO}

Group M includes mild-to-moderates asthmatics, groups S_1, S_2, and S_3 include severe asthmatics, group C includes COPD patients

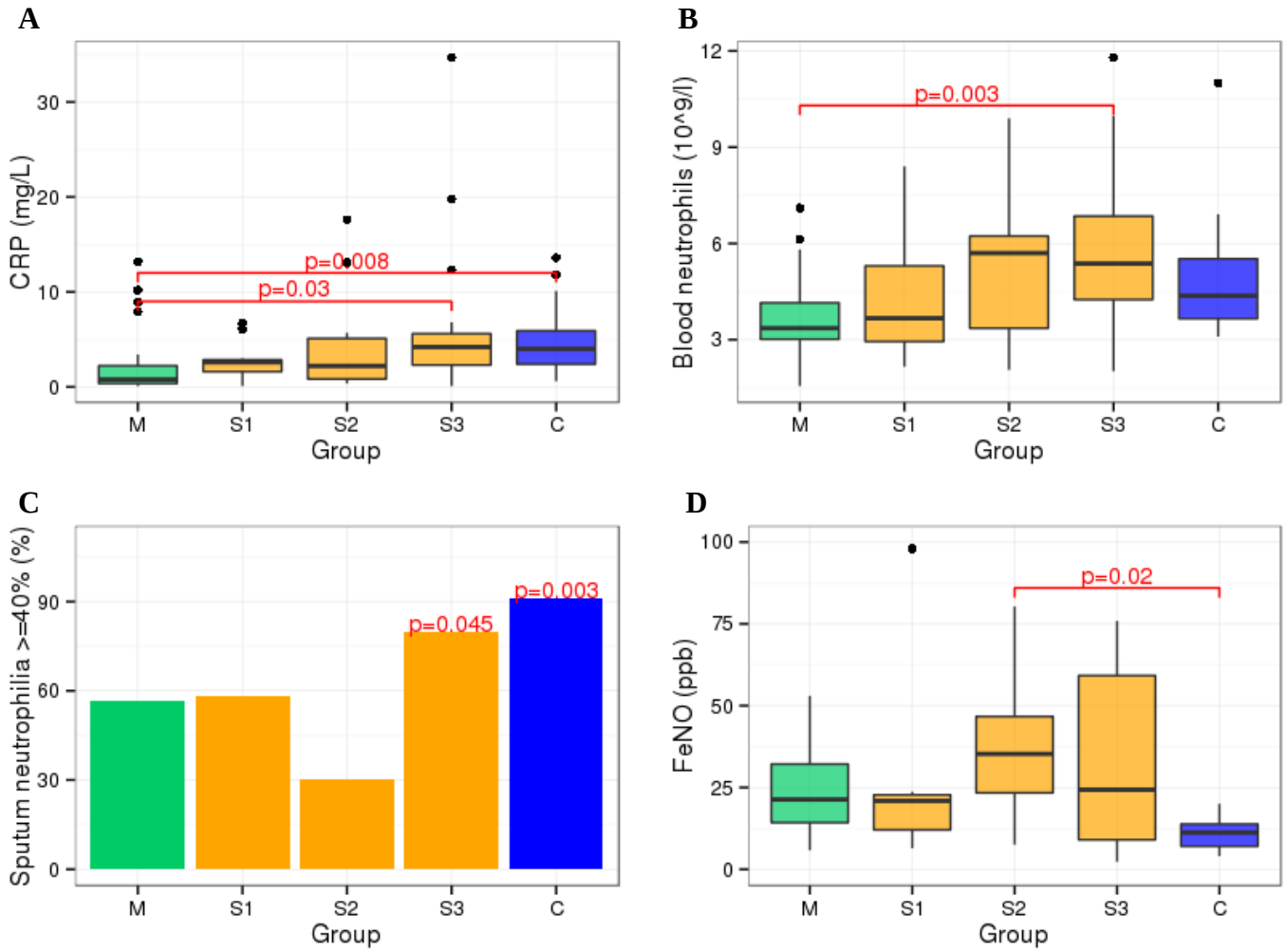

Figure 5.5. Distribution of inflammatory biomarkers according to groups M, S_1, S_2, S_3, and C

(A) High-sensitivity C-reactive protein, (B) Blood neutrophils, (C) Frequency of patients with sputum neutrophilia ≥ 40%, (D) FeNO

Group M includes mild-to-moderates asthmatics, groups S_1, S_2, and S_3 include severe asthmatics, group C includes COPD patients

Group M

Group M (n=39) was characterized by a significant over-representation of patients with early-onset disease (i.e., <18 years old) (19 (50.0%) in M as compared to 30 (28.8%) in the entire analysis population, adjusted p-value<0.001) and by a significant over-representation of patients with atopy (i.e., skin prick tests >9mm^2) (18 (47.4%) in M as compared to 38 (43.4%) in the entire analysis population, adjusted p-value=0.02). Independent of group, we found that atopic condition was associated with an early-onset disease in patients with asthma (median age of onset was 16.0 [5.0;32.0] years in subjects with atopy *vs.* 33.0 [20.5;46.5] years in subject without atopy, adjusted p-value<0.001). Score of Asthma Control Questionnaire (ACQ) (0.3 [0.1;0.6]) was significantly lower as compared to S_1 (1.4 [0.9;2.0], adjusted p-value=0.006), S_2 (2.0 [1.3;2.6], adjusted p-value<0.001), and S_3 (2.4 [2.0;3.0], adjusted p-value<0.001). Quality of life, i.e., score of St George's Respiratory Questionnaire, (11.4 [6.8;20.7]) was significantly better compared to S_1 (39.3 [26.0;42.6], adjusted p-value=0.049), S_2 (43.7 [37.7;48.8], adjusted p-value=0.001), S_3 (51.1 [33.1;59.9], adjusted p-value<0.001), and C (40.3 [34.2;56.2], adjusted p-value<0.001). Response to OCS was poor with a significant over-representation of non-responders (36 (92.3%) in M compared to 90 (81.0%) in the entire analysis population, adjusted p-value=0.004). This is probably due to the fact that the optimal lung function was gained already with the use of inhaled corticosteroids during the treatment optimization period.

Group S_1

Group S_1 (n=18) was characterized by a significant over-representation of patients with late-onset disease (16 (88.9%) in S_1 compared to 74 (71.1%) in the entire analysis population, adjusted p-value<0.001). With regard to demographic characteristics (i.e., age, gender, and BMI) and inflammatory biomarkers, patients were similar to those in group M. However, they exhibited increased airway obstruction with $zFEV_1$ (-2.1±1.0) and zFVC (-1.3±1.1) significantly lower compared to M (-0.7±1.1, adjusted p-value<0.001, and 0.04±0.9, adjusted p-value<0.001, respectively). Three patients (16.7%) responded to OCS.

Group S_2

Patients in group S_2 (n=16) were significantly older than in M (53.6±10.3 years *vs.* 43.2±13.7 years, adjusted p-value=0.02). There was a significant over-representation of patients who had at least one exacerbation during follow-up (13 (81.2%) in S_2 compared to 57

(51.4%) in the entire analysis population, adjusted p-value=0.002). Patients exhibited a decreased pulmonary function with $zFEV_1$ (-2.1±1.0) and zFVC (-1.3±1.1) significantly lower compared to group M (-0.7±1.1, adjusted p-value<0.001, and 0.04±0.9, adjusted p-value<0.001, respectively), and a hyperinflation characterized by a ratio RV/TLC significantly higher compared to group M (1.2±0.2 *vs.* 0.98±0.2, adjusted p-value=0.009). There were signs of airway inflammation with higher levels of sputum eosinophils (4.9 [1.2;18.1] %) compared to M (1.3 [0.1;4.7] %), though not significant (adjusted p-value=0.39). Fraction of exhaled nitric oxide (FeNO) (29.7 [10.1;59.9] ppb) was significantly higher compared to group C (11.2 [7.0;13.8] ppb, adjusted p-value=0.002). In response to OCS, these parameters decreased to the levels of that in M (**Online Supplement**). Four (25.0%) patients responded to OCS.

Group S_3

Group S_3 (n=18) was characterized by a significantly higher BMI compared to M (29.9±6.0 kg/m^2 *vs.* 24.9±3.3 kg/m^2, adjusted p-value=0.002), with a significant over-representation of obese subjects (i.e., BMI ≥ 30 kg/m^2) (9 (50.0%) in S_3 compared to 28 (25.0%) in the entire analysis population, adjusted p-value=0.003). There was a significant over-representation of patients who had a least one exacerbation during follow-up (13 (72.2%) in S_3 compared to 57 (51.4%) in the entire analysis population, adjusted p-value=0.01), as well as a significantly higher number of exacerbations compared to M (2 [1;2] *vs.* 0 [0;1], adjusted p-value=0.046). Patients exhibited decreased pulmonary function with $zFEV_1$ (-2.6±1.2), zFVC (-1.3±1.0), and IVC (92.3%±22.5%) significantly lower compared to M (-0.7±1.1, adjusted p-value<0.001, 0.04±0.9, adjusted p-value<0.001, and 107.2%±10.2%, adjusted p-value=0.006, respectively), a hyperinflation characterized by a ratio RV/TLC (1.2±0.3) significantly higher compared to M (0.98±0.2, adjusted p-value=0.003), as well as a loss of diffusion capacity, which was significantly lower compared to M (D_{LCO}: 80.4%±18.3% *vs.* 94.8%±13.2%, adjusted p-value=0.02). There were signs of airway inflammation with significantly higher levels of blood neutrophils (5.4 [4.3;6.9] $10^9/l$) and CRP (4.2 [2.3;5.6] mg/l) compared to M (3.4 [3.0;4.2], adjusted p-value=0.003, and 0.8 [0.4;2.2] mg/l, adjusted p-value=0.03, respectively), higher levels of sputum eosinophils (4.5 [1.0;10.4] %) compared to M (1.3 [0.1;4.7] %), though not significant, a significant over-representation of patients with sputum neutrophils ≥ 40% (8 (80.0%) in S_3 compared to 45 (61.6%) in the entire analysis population, adjusted p-value=0.045), and a mixed granulocytic

inflammation (i.e., combined increase in sputum eosinophils ≥ 2% and sputum neutrophils ≥ 40% (Moore, Meyers et al. 2010)) (5 (50.0%) in S_3 compared to 21 (28.8%) in the entire analysis population, adjusted p-value=0.03). In response to OCS, these parameters decreased to the levels of that in M (**Online Supplement**). There was significant over-representation of OCS responders (8 (44.4%) in S_3 compared to 21 (18.9%) in the entire analysis population, adjusted p-value<0.001).

Group C

Group C (n=20) was markedly different from the other groups and mainly consisted of older patients (65.1±7.5 years), predominantly males (13 (65.0%) in C compared to 48 (43.2%) in the entire analysis population, adjusted p-value=0.008), with late-onset disease (age of onset: 60.5 [52.5;65.5] years). Patients exhibited severe reductions in pulmonary function. Lung function was markedly decreased with $zFEV_1$ (-3.5±0.6) and $zFEV_1$/zFVC (-3.5±0.9) which were significantly lower compared to each of the 4 groups of asthmatics, and zFVC (-2.0±0.7) significantly lower compared to M (0.04±0.9, adjusted p-value<0.001) and S_1 (-1.0±1.3, adjusted p-value<0.02). Patients presented with a fixed obstruction characterized by reversibility (3.1% [0.8%;4.4%]) significantly lower compared to M (10.2% [7.7%;14.1%], adjusted p-value<0.001), S_1 (9.1% [5.6%;12.9%], adjusted p-value=0.005) and S_3 (7.6% [4.6%;14.9%], adjusted p-value=0.04). Lung function mechanics revealed a hyperinflation characterized by FRC (128.3%±30.0%) significantly higher compared to each of the 4 groups of asthmatics, and RV/TLC (1.4±0.2) significantly higher compared to M (0.98±0.2, adjusted p-value<0.001) and S_1 (1.1±0.2, adjusted p-value<0.001). Diffusion capacity (D_{LCO}: 56.3%±21.3%) was significantly lower compared to each of the 4 groups of asthmatics. Similar to S_3, there were signs of airway inflammation with CRP levels being significantly higher (4.0 [2.4;5.9] mg/l) compared to M (0.8 [0.4;2.2] mg/l, adjusted p-value=0.008), a significant over-representation of patients with sputum neutrophils ≥ 40% (10 (90.9%) in C compared to 45 (61.6%) in the entire analysis population, p=0.003), and a mixed granulocytic inflammation (5 (45.5%) in C compared to 21 (28.8%) in the entire analysis population, adjusted p-value=0.049). In response to OCS, these parameters decreased to the levels of that in M (**Online Supplement**). Three (15.0%) patients responded to OCS. There were also meaningful and distinct differences with respect to certain biomarkers between groups S_3 and C, as displayed in **Table 4,** and discussed in the **Online Supplement**.

5.6. Discussion

5.6.1. Main results

Our approach to phenotyping based on lung function fluctuations allowed for the identification of 5 groups (M, S_1, S_2, S_3, and C) corresponding to relevant phenotypes with a gradual severity regarding clinical, inflammatory, and pulmonary features, distinct exacerbation rate, and traits suggesting specific treatment response features (**Figure 5.6**).

In particular, we found a gradually increasing clinical severity from M to C, characterized by a gradual decrease in quality of life and asthma control, and a higher number of exacerbations in S_2 and S_3.

Related to the increasing clinical severity from M to C, we found increasing changes in lung function (i.e., increased airway obstruction, increased hyperinflation, and a loss of diffusion capacity) which might be due to progressive structural changes. Airways gradually reached a more rigid, narrow state, and became less reversible to β_2-mimetics. Particularly, in S_3, seriously impaired lung function combined with the severe clinical phenotype and signs of inflammation might be related to obesity since this group mainly consisted of obese subjects.

Response to OCS differed according to groups and seemed to result from the relative pathophysiological contributions of airway obstruction, inflammation, and irreversible mechanical impairment. Response in M was weak, probably due to a ceiling effect. Indeed, since this group of patients had minimal airway obstruction, they were probably controlled with inhaled corticosteroids, reducing any benefit of adding OCS. From S_1 to S_3, where degree of obstruction and signs of inflammation gradually increased, we found a gradually increasing response to OCS, with a particularly good response in S_3. In C, irreversible mechanical impairment of the lung might have rendered patients clinically unresponsive to OCS despite a satisfactory anti-inflammatory response. Thus, in these patients the mechanical impairment dominated the clinical picture.

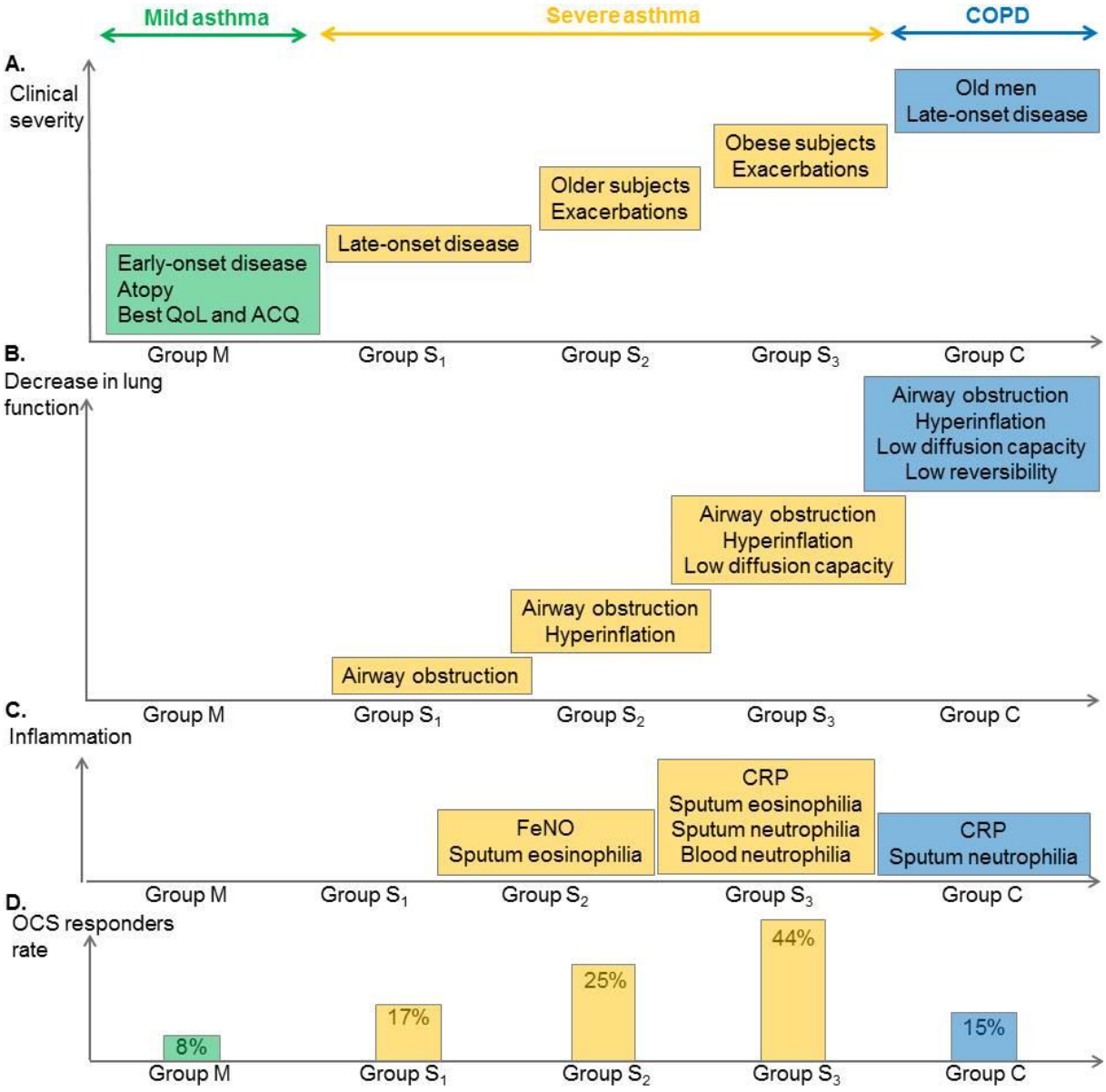

Figure 5.6. Predominant (A) clinical, (B) pulmonary and (C) inflammatory features, and (D) response to oral corticosteroid according to groups M, S_1, S_2, S_3, and C

CRP, high-sensitivity C-reactive protein; FeNO, fraction of exhaled nitric oxide; OCS, oral corticosteroid

Group M includes mild-to-moderates asthmatics, groups S_1, S_2, and S_3 include severe asthmatics, group C includes COPD patients

5.6.2. Strengths and weaknesses of the study (internal validity)

The BIOAIR cohort, due to its unique design, enabled, for the first time, comparison of patients with mild-to-moderate asthma, severe asthma, and COPD in regards to their lung function fluctuation over a one year period. Additional strengths of the study are a) the use of a single variable to perform the clustering analysis, thereby circumventing the issue of variable selection, which renders many clustering approaches subjective; b) the use of a time-related variable to perform the clustering analysis, instead of parameters measured at a single point in time; c) the detailed information available about the patients, allowing for an extensive description of the groups identified; d) the 2-week double blind placebo-controlled OCS intervention, allowing for the assessment of response to treatment. However, despite the high number of patients included in the study, the power of the analysis might have been limited sometime by the multiple testing correction.

5.6.3. Strengths and weaknesses of the study compared to other studies (external validity)

Our results are in line with the view that pulmonary function is an important determinant of disease severity in asthma and COPD (Sorkness, Bleecker et al. 2008, Moore, Meyers et al. 2010). Indeed, solely on the basis of lung function fluctuations, we have identified a range of groups with gradual phenotype severity. Similarly, in the SARP study, lung function best differentiated the mildest from the most severe groups (Moore, Meyers et al. 2010).

Furthermore, alterations in pulmonary function may reflect specific underlying pathophysiological mechanisms. It has been shown that persistent airflow obstruction is increasingly common as asthma severity increases (Pascual and Peters 2009, Konstantellou, Papaioannou et al. 2015), and has been described as the manifestation of progressive structural changes in the airway walls (Thamrin, Nydegger et al. 2011). Sorkness et al. (Sorkness, Bleecker et al. 2008) found that severe asthmatics have a greater component of air trapping, relative to the airflow limitation component, contributing to airway obstruction. Therefore, the greater airway obstruction combined with the hyperinflation found in groups S_2 and S_3 are in accordance with features of persistent airflow obstruction, air trapping, and airway remodeling described in more severe asthma phenotypes. Interestingly, Choi et al. identified four clusters very similar to M, S_1, S_2, and S_3, using an imaging-based clustering approach (Choi, Hoffman et al. 2017). In particular, our group S_2 was similar to their luminal narrowing-dominant cluster, and S_3 was similar to their wall thickening-dominant cluster.

Finally, we identified a group of severe asthmatics (S3) characterized by an over-representation of obese subjects; meaning that obesity might be associated with a specific pattern of lung function fluctuation. There is strong evidence that obesity increases the prevalence and incidence of asthma, and reduces asthma control (Hakala, Stenius-Aarniala et al. 2000, Stenius-Aarniala, Poussa et al. 2000, Saint-Pierre, Bourdin et al. 2006, Beuther and Sutherland 2007, Mosen, Schatz et al. 2008). A more limited body of evidence suggests that obesity may also increase the severity of asthma (Peters-Golden, Swern et al. 2006, Saint-Pierre, Bourdin et al. 2006, Taylor, Mannino et al. 2008). Our findings are in accordance with these observations, since S3 was characterized by a particularly low quality of life, poor asthma control, exacerbations, and notably altered pulmonary function. Alteration of pulmonary function included airway obstruction, hyperinflation, and low diffusion capacity, which concord with reduced lung volumes and changes in airway resistance described in the literature (Zerah, Harf et al. 1993, Collins, Hoberty et al. 1995). Moreover, we found signs of inflammation which are consistent with the low-grade systemic inflammation described in obese subjects (Shore 2008), as well as a mixed granulocytic inflammation which has been previously described in very severe asthma (Moore, Meyers et al. 2010). These findings support the hypothesis that obesity might be associated with a specific type of asthma, related to specific alterations in lung function, and clinically presenting a greater and/or more difficult to control disease state (Mosen, Schatz et al. 2008, Umetsu 2016), which may better profit from a multidimensional treatment approach (Sodlerlund, Fischer et al. 2009, Frey, Latzin et al. 2015).

5.6.4. Relevance of the study results for disease management and monitoring strategies

Our findings might help identify patients who could benefit from different treatment strategies, further clinical investigations in a referral center, and/or closer monitoring. In particular, the lung function measurements might be implemented in a telemonitoring setting for diagnostic purpose (e.g., graduation of asthma severity, or asthma-COPD-overlap-syndrome), or for monitoring purpose (e.g., in patients with severe phenotypes, especially if there is a high risk exacerbation, periods of closer monitoring could be recommended, for instance after implementing a new treatment strategy). While conventional disease phenotyping usually relies on many characterizing parameters, which tend to be expensive and limited to in-hospital assessment (Delgado, Kumar et al. 2015), fluctuation in FEV_1 can

be implemented in a simple and cost-effective way in a telemonitoring setting with an appropriate adherence measure (Kupczyk, Haque et al. 2013).

Furthermore, asthma and COPD patients phenotyping based on airway dynamics might have in the near future relevant research applications. First, the characterization of structural alterations of the lung according to such phenotypes, using imaging techniques, might improve the understanding of disease pathogenesis (Choi, Hoffman et al. 2017). Second, evaluation of new treatment strategies according to such phenotypes in future controlled treatment trials might be of great value.

5.7. Conclusion

The present study uses a novel clustering approach, solely based on the lung function fluctuation recordings over one year. This approach identifies phenotypes, in which the progressive functional alteration of the lung corresponds to a gradually increasing clinical severity, and which may translates into specific treatable traits.

5.8. Acknowledgements

We thank K. Hugentobler for language editing, and the staff at all clinical sites for careful supervision of patients and collection of data during the trial.

Source of support: Funding for this project was provided by the Swiss National Science Foundation (3200-B0-112099). The BIOAIR study was supported by The Fifth and Sixth Framework Programmes of the European Union, contract numbers: QLG1-CT-2000-01185 (BIOAIR) and FOOD-CT-2004-506378 (GA2LEN), several national funding bodies (Sweden: Heart-Lung Foundation, Asthma and Allergy Foundation and the Stockholm County Council), and received unconditional support from Vitalograph. The ChAMP consortium is supported by the Swedish Strategic Research Foundation and an unconditional and competitive grant donated by AstraZeneca. MK and AJ were supported by the Bernard Osher Initivative for Severe Asthma Research.

5.9. Online Supplement

Methods

Data collection and measurements

Data were collected using an electronic Case Report Form (eCRF) developed specifically for the BIOAIR study. A reversibility test, as well as skin prick tests to common aeroallergens, were performed at the screening visit. At inclusion (visit 1), an extensive spirometry (inspiratory vital capacity (IVC), total lung capacity (TLC), and residual volume (RV)) was performed according to published guidelines (Quanjer, Tammeling et al. 1993), as well as diffusion capacity measurements (forced residual volume (FRC), and diffusing capacity of the lung for carbon monoxide (D_{LCO})) according to the technique described by Kerstjens et al. (Kerstjens, Brand et al. 1992). Induced sputum was obtained at inclusion and at the end of the optimization period (visit 2) using inclusion and exclusion criteria according to published recommendations (Paggiaro, Chanez et al. 2002). At each visit (from visit 1 to visit 6) the fraction of exhaled nitric oxide (FeNO) using a NIOX analyser (Aerocrine AB, Solna, Sweden) was measured according to guidelines (1999), as well as ordinary spirometry indices (forced expiratory volume in one second (FEV_1), forced vital capacity (FVC) and FEV_1/FVC) with calibrated spirometers using pneumotachometry according to standardised guidelines (Quanjer, Tammeling et al. 1993). Serum CRP levels were measured using a standardised high sensitivity assay with a clinical Cobas c502 (8000) instrument (Roche Diagnostics). Multiplex assays for (Lumican, metalloproteinase-3 (MMP-3), Dipeptidyl peptidase-4 (DPPIV), soluble receptor for advanced glycation end products (sRAGE), YKL-40 (or Chitinase 3-Like 1) were performed using human Luminex® screening assay reagents from R&D Systems (Bio-Techne, Abingdon, UK). Patients completed at each visit the St. George's respiratory questionnaire (SGRQ) (Jones, Quirk et al. 1992), and the Juniper's Asthma Control Questionnaire (ACQ) (Juniper, O'Byrne et al. 1999).

Lung function fluctuation based clustering of patients

The FBC method consists of identifying clusters of patients with similar patterns of lung function fluctuation by comparing each patient's empirical distribution of daily lung function measurements recorded over a predetermined window of observation. The FBC method has been described in detail elsewhere (Delgado-Eckert, Fuchs et al. 2017). In brief, the FBC method consists of the following steps:

1. Quantification of similarity in lung function fluctuation between individuals

2. Grouping of individuals into clusters such that similarity between members of the same clusters is strong and between different clusters is weak.

Furthermore, the FBC method includes a data-driven process for determining the tolerable amount of missing measurements. This data-driven process has been described in detail elsewhere (Delgado-Eckert, Fuchs et al. 2017). In brief, a highly compliant subset of patients (i.e., with a high number of FEV_1 measurements), the so-called "gold standard", was selected. Patients were defined as having a high number of FEV_1 measurements if their individual set of FEV_1 measurements contained at least as many measurements as the 60^{th} percentile of the overall distribution of the number of FEV_1 measurements from the entire analysis population. Then, within the gold standard, in order to quantify similarities in lung function fluctuation between individuals, the distribution of z-score FEV_1 values of a given patient was compared with the distributions of all other patients in the gold standard. This pair-wise comparison was done using the Earth mover's distance (EMD). A low value of EMD indicates high similarity in lung function fluctuation between two individuals. Patients were grouped into clusters such that the similarity between members of the same clusters was strong, and between different clusters was weak using the Ward's minimum-variance hierarchical clustering method. Afterwards, a cluster stability analysis was performed in order to assess whether further patients who had fewer measurements could be included in the analysis without disturbing the clusters identified. The outcome of this stability analysis enabled us to establish the minimum number of FEV_1 measurements required to ensure the stability of the clusters. Finally, patients who performed the minimum number of FEV_1 measurements required were added to the gold standard, and the cluster analysis was repeated with this larger subset to obtain the final clusters.

Statistical analysis

For categorical variables, we chose a resampling method to address the multiple testing issues instead of the commonly used Bonferroni correction, which is known to be very conservative. The resampling method consisted in randomly selecting artificial groups from the analysis population, with the same sample sizes as the groups M, S_1, S_2, S_3, and C, in which the tests were performed. We then applied the enrichment test to each of the 5 artificial groups (i.e., groups M', S_1', S_2', S_3', and C'). The resampling was done 10 000 times (i.e., 5 artificial groups 10 000 times randomly selected out of the analysis population), and the enrichment test was performed in the 5 artificial groups in each iteration.

We then adjusted the significance level of the individual tests such that no more than 5% of the 10 000 tests conducted on artificial groups were significant. In other words, we set the family-wise error rate at the 5% level. The adjusted significance levels of the individual tests were then used when analyzing the actual groups M, S_1, S_2, S_3, and C.

Phenotypes identified (based on patients' lung function fluctuation over the one-year follow-up period) were described at baseline (i.e., at the end of the treatment optimization period (visit 2)), and at study end (visit 6) in order to explore the characteristics of the phenotypes found as a function of time.

Results

Selection of the analysis population according to the determined tolerable amount of missing measurements

Mean number of FEV_1 measurements in the 186 patients initially included in the clustering analysis was 310±238. The 60^{th} percentile of the distribution of number of FEV_1 measurements was equal to 481 measurements. Thus, the highly compliant subset of patients consisted of patients who performed at least 481 FEV_1 measurements during follow-up, namely 59 patients. As a result of the cluster stability analysis, we were able to decrease this threshold by 90%, namely 49 measurements were required, instead of 481 measurements. Thus, after the cluster stability analysis, all those patients who performed at least 49 measurements during the follow-up could be considered. The analysis population consisted, therefore, of 134 patients.

Characteristics of patients excluded from the analysis population (n=99)

Regarding patients with severe asthma, $zFEV_1$ and zFVC were significantly lower in patients excluded than in patients from the analysis population ($zFEV_1$: -2.7±1.3 *vs.* -2.0±1.3, p=0.01; zFVC: -1.8±1.3 *vs.* -1.1±1.2, p=0.02) (**Table E5.1**). Regarding patients with COPD, zFVC, FRC and IVC were significantly lower in patients excluded than in patients from the analysis population (zFVC: -2.3±1.2 vs. -1.4±0.8, p=0.003; FRC: 94.9% [80.3%;125.4%] vs. 126.2% [104.0%;147.1%, p=0.01; IVC: 72.7% [64.0%;90.7%] vs. 90.2% [81.7%;96.7%], p=0.008).

Table E5.1. Characteristics of patients excluded from the analysis population (n=99)

	Mild-to-moderate asthma (N=23)	**MD**	**p-value***	**Severe asthma (N=39)**	**MD**	**p-value†**	**COPD (N=37)**	**MD**	**p-value‡**
Clinical characteristics									
Age, years	44.0 [30.5;51.5]	-	0.82	53.0 [35.5;64.0]	-	0.98	67.0 [57.0;70.0]	-	0.73
Gender, male	10 (43.5%)	-	0.64	18 (46.2%)	-	0.48	29 (78.4%)	-	0.46
BMI, kg/m^2	25.0 [23.0;27.5]	-	0.59	27.0 [24.5;32.0]	-	0.92	26.0 [23.0;31.0]	-	0.76
Age of disease onset, years	26.2±19.3	11	0.53	28.3±18.1	9	0.56	54.1±10.9	9	0.16
ACQ, Juniper	1.3±0.7	8	0.07	2.2±1.2	9	0.47	NA	NA	NA
QoL, SGRQ	22.1 [16.7;31.6]	7	0.17	48.8 [32.5;62.8]	5	0.48	43.1 [33.3;58.9]	8	0.45
Atopy	5 (21.7%)	-	0.04	10 (25.6%)	-	0.16	0 (0.0%) §	-	NA
Lung function									
Reversibility, % of change	11.0±4.3	2	0.74	10.6±9.6	1	0.24	4.2±3.5	2	0.22
FeNO, ppb	44.0 [34.8;53.6]	17	0.32	33.8 [23.2;41.5]	21	0.93	15.0 [9.3;26.8]	24	0.25
FEV_1, z-score	-1.4±1.3	7	0.80	-2.7±1.3	3	**0.01**	-3.5±1.0	3	0.37
FVC, z-score	-0.5±1.2	7	0.40	-1.8±1.3	3	**0.02**	-2.3±1.2	3	**0.003**
FEV_1/FVC, z-score	-1.7±1.0	7	0.84	-2.0±1.4	3	0.48	-3.1±1.4	3	0.13
DLCO, %predicted corrected	90.3 [88.3;94.2]	16	0.80	83.1 [71.3;99.3]	15	0.74	57.7 [43.0;67.6]	17	0.71
FRC, %predicted corrected	87.7 [72.3;101.2]	15	0.24	92.4 [78.2;108.1]	15	0.59	94.9 [80.3;125.4]	8	**0.01**
IVC, %predicted corrected	101.4 [93.0;106.5]	13	0.46	87.2 [75.5;101.9]	11	0.10	72.7 [64.0;90.7]	11	**0.008**
TLC, %predicted corrected	109.2 [103.8;115.4]	9	0.20	97.0 [90.4;109.2]	10	0.27	105.1 [90.6;116.0]	8	0.32
RV, %predicted corrected	116.7 [108.9;140.4]	10	0.19	121.6 [95.6;143.3]	10	0.72	148.7 [116.0;171.0]	8	0.99
RV/TLC	1.1 [1.0;1.2]	10	0.24	1.3 [1.0;1.4]	10	0.20	1.4 [1.2;1.6]	8	0.24
Inflammatory response									
hs-CRP, mg/litre	1.3 [0.08;3.0]	15	0.28	2.8 [0.3;7.6]	16	0.19	5.0 [1.4;9.5]	18	0.90
Sputum cells, ×10^6	1.3 [0.3;4.0]	7	0.36	0.9 [0.4;2.2]	17	0.76	1.8 [0.3;4.2]	13	0.97
Sputum eosinophils, %	1.6 [0.3;8.1]	7	0.91	1.5 [0.0,8.3]	16	0.14	0.2 [0.0;2.3]	13	0.68
Sputum neutrophils, %	30.8 [17.8;51.0]	7	0.66	36.5 [9.1;64.5]	16	0.50	63.6 [41.9;83.4]	13	0.52
White blood cells, ×10^9/l	7.3±1.5	5	0.07	8.4±3.2	4	0.80	7.8±1.6	8	0.86

Table E5.1. Characteristics of patients excluded from the analysis population (n=99) (continued)

	Mild-to-moderate asthma (N=23)	**MD**	**p-value***	**Severe asthma (N=39)**	**MD**	**p-value†**	**COPD (N=37)**	**MD**	**p-value‡**
Blood eosinophils, $\times 10^9$/l	0.3 [0.2;0.3]	7	0.99	0.2 [0.1;0.6]	4	0.54	0.2 [0.1;0.3]	9	0.82
Blood neutrophils, $\times 10^9$/l	4.0 [3.6;4.2]	7	0.26	4.5 [3.7;6.7]	4	0.99	4.9 [3.6;5.6]	10	0.37

Values shown are mean ± standard deviation, median [25^{th} percentile; 75^{th} percentile] and numbers (percentages)

ACQ, Asthma Control Questionnaire; BMI, body mass index; hs-CRP, high-sensitivity C-reactive protein; D_{LCO}, diffusing capacity of the lung for carbon monoxide; FeNO, fraction of exhaled nitric oxide; FEV_1, forced expiratory volume in one second; FRC, forced residual volume; FVC, forced vital capacity; IVC, inspiratory vital capacity; MD, missing data; NA, not applicable; QoL, quality of life; RV, residual volume; SGRQ, St George's Respiratory Questionnaire; TLC, total lung capacity

*Compared with mild-to-moderate asthmatics from analysis population; †Compared with severe asthmatics from analysis population; ‡Compared with COPD patients from analysis population; §Inclusion criteria

Description of group MS (n=10)

Characteristics of patients according to groups M, MS, S_1, and S_2 are provided in **Table E5.2**. Group MS was similar to group M regarding age, BMI and number of exacerbations. There was a gradual increase in the scores of QoL and ACQ (meaning a gradual decrease in QoL and asthma control) from group M - MS - S1 - S2. Finally, group MS exhibited a poor lung function, similar to that found in patients from group S_2.

Table E5.2. Characteristics of patients according to subgroups M, S_1, S_2, and C (n=111)

	Group M (N=39)	MD	**Group MS (N=10)**	MD	**Group S_1 (N=18)**	MD	**Group S_2 (N=16)**	MD	**p-value**
Airway disease	**Mild-to-moderate asthma**		**Mild-to-moderate asthma**		**Severe asthma**		**Severe asthma**		
Clinical characteristics									
Age, years	41.0 [32.0;52.5] ‡	-	42.0 [36.0;44.0]	-	52.5 [44.3;60.3]	-	55.5 [44.3;60.5] †	-	**0.01**
Gender, male	14 (35.9%)	-	4 (40.0%)	-	8 (44.4%)	-	6 (37.5%)	-	0.93
BMI, kg/m^2	24.0 [23.0;26.0] ‡	-	24.0 [20.3;28.5]	-	26.5 [25.3;28.0]	-	28.0 [26.3;30.0] †	-	**0.01**
Age of disease onset, years	18.5 [5.3;32.5]	1	32.0 [7.0;38.0]	1	34.0 [24.3;43.8]	-	24.5 [13.3;37.5]	2	0.10
Atopy	18 (47.4%)	1	3 (33.3%)	1	5 (33.3%)	3	8 (50.0%)	-	0.73
QoL, SGRQ	11.4 [6.8;20.7] ‡	8	25.4 [16.4;35.7]	2	39.3 [26.0;42.6]	3†	43.7 [37.7;48.8] †	4	**<0.001**
ACQ, Juniper	0.3 [0.1;0.6] ‡	-	1.0 [0.9;1.3] †	1	1.4 [0.9;2.0] †	-	2.0 [1.3;2.6] †	1	**<0.001**
Number of exacerbation during follow-up	0 [0;1] ‡	-	0 [0;0] ‡	-	1 [0;2]	-	1 [1;2] †	-	**<0.001**
At least one exacerbation during follow-up	13 (33.3%)	-	1 (10.0%)	-	10 (55.6%)	-	13 (81.2%)	-	**<0.001**
Lung function									
Reversibility, percentage change	10.2 [7.7;14.1]	1	7.1 [6.0;10.6]	-	9.1 [5.6;12.9]	-	5.5 [0.9;9.3]	1	0.06
FeNO, ppb	21.3 [14.3;32.2]	12	28.4 [14.4;41.2]	6	21.4 [14.5;42.3]	10	40.5 [23.7;53.8]	7	0.15
FEV_1, z-score	-0.8 [-1.3;-0.1] ‡	-	-2.0 [-2.8;-1.7] †	1	-1.4 [-1.7;-0.9]	-	-2.4 [-2.7;-1.8] †	-	**<0.001**
FEV_1, % predicted	95.4 [89.5;104.2] ‡	-	73.5 [67.2;79.6] †	-	83.9 [75.4;92.4]	-	70.3 [60.7;80.6] †	-	**<0.001**
FVC, z-score	0.1 [-0.5;0.4] ‡	-	-1.4 [-1.5;-0.3]	1	-1.1 [-1.5;-0.3] †	-	-1.4 [-2.0;-0.5] †	-	**<0.001**
FVC, % predicted	101.5 [92.9;105.3] ‡	-	82.6 [80.6;96.6]	1	85.2 [78.9;96.3] †	-	81.4 [72.9;92.34] †	-	**<0.001**
FEV_1/FVC	-1.1 [-1.9;-0.2]	-	-1.7 [-2.6;-1.3]	1	-0.8 [-1.7;0.03]	-	-1.8 [-2.4;-1.2]	-	0.08
DLCO, %predicted corrected	95.4 [86.4;103.9]	2	92.0 [82.2;107.0]	2	84.8 [83.1;92.8]	6	88.2 [81.9;98.8]	-	0.22
FRC, %predicted corrected	96.1 [82.7;115.4]	2	107.9 [83.4;147.9]	2	93.2 [48.0;110.9]	6	89.9 [81.7;112.0]	-	0.61
IVC, %predicted corrected	108.9 [101.6;112.7] ‡	4	90.6 [87.3;96.7] †	-	98.0 [92.2;100.7] †	-	93.7 [85.3;107.0] †	1	**<0.001**
TLC, %predicted corrected	107.0 [95.5;112.3]	1	101.1 [92.8;113.6]	-	94.3 [86.9;106.5] †	1	106.1 [97.4;110.8]	-	0.18
RV, %predicted corrected	103.8 [92.1;124.7]	2	123.2 [94.0;166.3]	-	95.1 [86.9;121.7]	1	119.7 [112.3;150.2]	-	0.07
RV/TLC	1.0 [0.9;1.1] ‡	2	1.2 [1.0;1.5]	-	1.1 [0.9;1.2]	1	1.2 [1.1;1.4] †	-	**0.005**
Inflammatory biomarkers									
Sputum cells, $\times 10^6$	0.6 [0.2;1.3]	8	0.6 [0.3;3.0]	3	2.2 [0.6;4.0]	6	1.0 [0.4;2.0]	5	0.19
Sputum eosinophils, %	1.3 [0.1;4.7]	9	1.4 [0.5;3.5]	3	1.5 [0.7;17.8]	6	4.9 [1.2;18.1]	6	0.21
Sputum eosinophilia ≥ 2%	12 (40.0%)	9	3 (42.9%)	3	5 (41.7%)	6	6 (60.0%)	6	0.73

Table E5.2. Characteristics of patients according to subgroups M, S_1, S_2, and C (n=111) (continued)

Airway disease	**Group M (N=39) Mild-to-moderate asthma**	**MD**	**Group MS (N=10) Mild-to-moderate asthma**	**MD**	**Group S1 (N=18) Severe asthma**	**MD**	**Group S2 (N=16) Severe asthma**	**MD**	**p-value**
Sputum neutrophils, %	46.0 [17.6;67.0]	9	44.5 [35.5;73.9]	3	51.3 [26.7;69.4]	6	25.1 [23.4;58.6]	6	0.62
Sputum neutrophilia ≥ 40%	17 (56.7%)	9	4 (57.1%)	3	7 (58.3%)	6	3 (30.0%)	6	0.49
Mixed granulocytic inflammation, %	6 (20.0%)	9	1 (14.3%)	3	3 (25.0%)	6	2 (20.0%)	6	0.76
hs-CRP, mg/l	0.8 [0.4;2.2]	7	1.8 [1.4;5.0]	1	2.6 [1.6;2.9]	3	2.2 [0.8;5.1]	1	0.06
White blood cells, $\times 10^9$/l	5.9 [5.4;6.8]	1	6.5 [6.1;8.0]	-	6.6 [5.7;8.0]	-	8.6 [6.0;9.8]	-	**0.05**
Blood eosinophils, $\times 10^9$/l	0.3 [0.2;0.3]	1	0.2 [0.1;0.5]	-	0.3 [0.1;0.4]	-	0.4 [0.1;0.5]	1	0.74
Blood neutrophils, $\times 10^9$/l	3.4 [3.0;4.2]	1	4.0 [3.3;4.7]	-	3.7 [2.9;5.3]	-	5.9 [3.5;6.3]	-	0.09
Response to treatment									
Response to oral corticosteroids	3 (7.7%)	-	2 (22.2%)	-	3 (16.7%)	-	4 (25.0%)	-	0.26

Values shown are mean ± standard deviation, median [25th percentile; 75th percentile] and numbers (percentages)

ACQ, Asthma Control Questionnaire; BMI, body mass index; hs-CRP, high-sensitivity C-reactive protein; D_{LCO}, diffusing capacity of the lung for carbon monoxide; FeNO, fraction of exhaled nitric oxide; FEV_1, forced expiratory volume in one second; FRC, forced residual volume; FVC, forced vital capacity; IVC, inspiratory vital capacity; MD, missing data; QoL, quality of life; RV, residual volume; SGRQ, St George's Respiratory Questionnaire; TLC, total lung capacity

‡ significant difference as compared to S_2; † significant difference as compared to M

Inflammatory response to oral corticosteroids

After the OCS intervention, FeNO decreased (from 40.5 [23.7;53.8] ppb to 27.5 [29.3;44.6] ppb) in group S_2 so that there was no significant difference between groups anymore (p=0.40) (**Figure E5.1**). CRP levels decreased in groups S_3 (from 4.2 [2.3;5.6] mg/l to 1.1 [0.6;5.0] mg/l) and C (from 4.0 [2.4;5.9] to 1.2 [0.5;4.1] mg/l) so that there was no significant difference between groups anymore (p=0.12). Finally, blood neutrophils levels increased in all groups so that there was no significant difference between groups anymore (p=0.29).

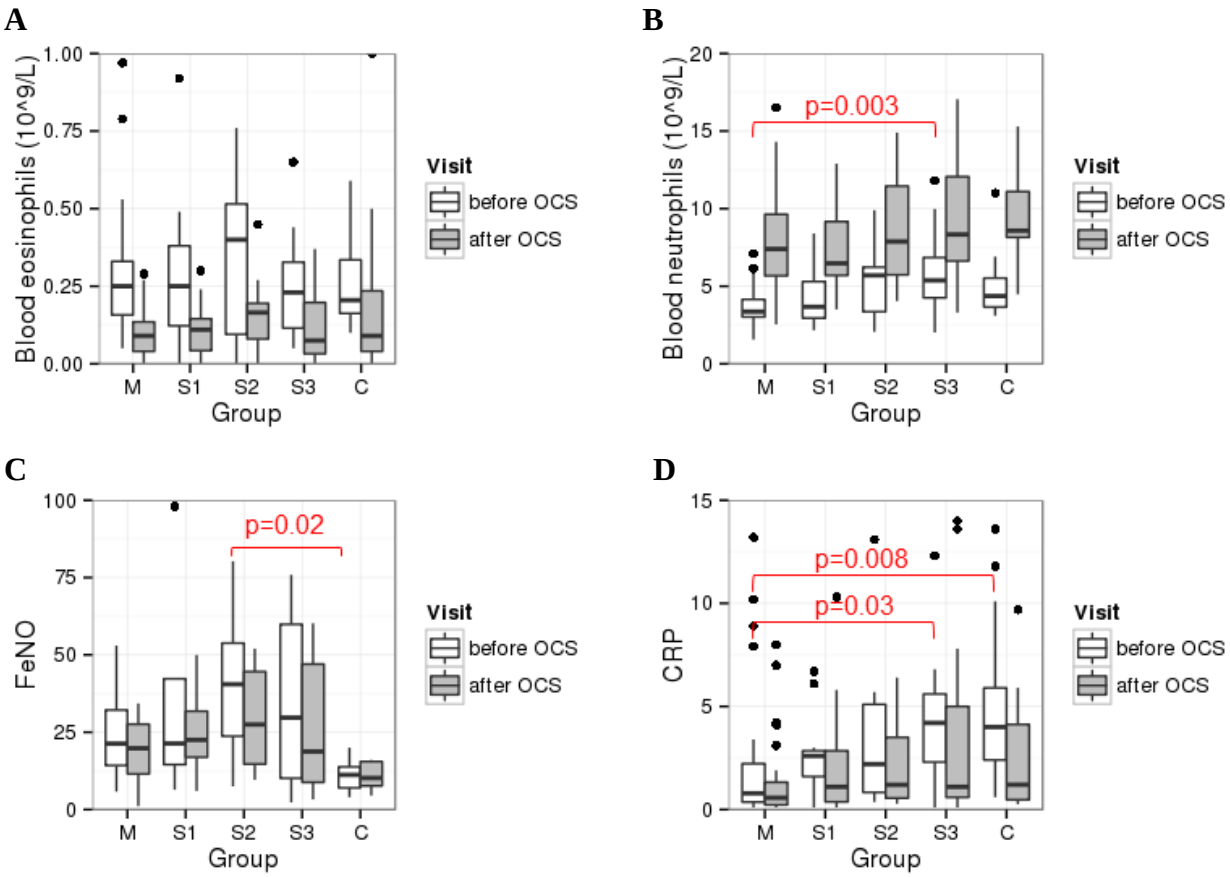

Figure E5.1. Distribution of (A) blood eosinophils, (B) blood neutrophils, (C) FeNO, and (D) CRP, before and after a 2-week double blind placebo-controlled oral prednisone intervention, according to groups M, S_1, S_2, S_3, and C

hs-CRP, high-sensitivity C-reactive protein; FeNO, fraction of exhaled nitric oxide; OCS, oral corticosteroid

Group M includes mild-to-moderates asthmatics, groups S_1, S_2, and S_3 include severe asthmatics, group C includes COPD patients

Stability of characteristics of phenotypes (comparison of characteristics at baseline and study end)

Characteristics of phenotypes were stable from baseline to study end according to clinical characteristics (i.e., quality of life and asthma control) and lung function (i.e., FEV_1, hyperinflation, loss of diffusion capacity) (**Table E5.3, Figures E5.2-E5.4**). However, the patients showed more intragroup variability in their inflammatory biomarker characteristics. Consequently, phenotypes obtained using the FBC method seem to be predominantly determined by lung mechanics, namely, lung mechanics are the phenotype determining characteristic.

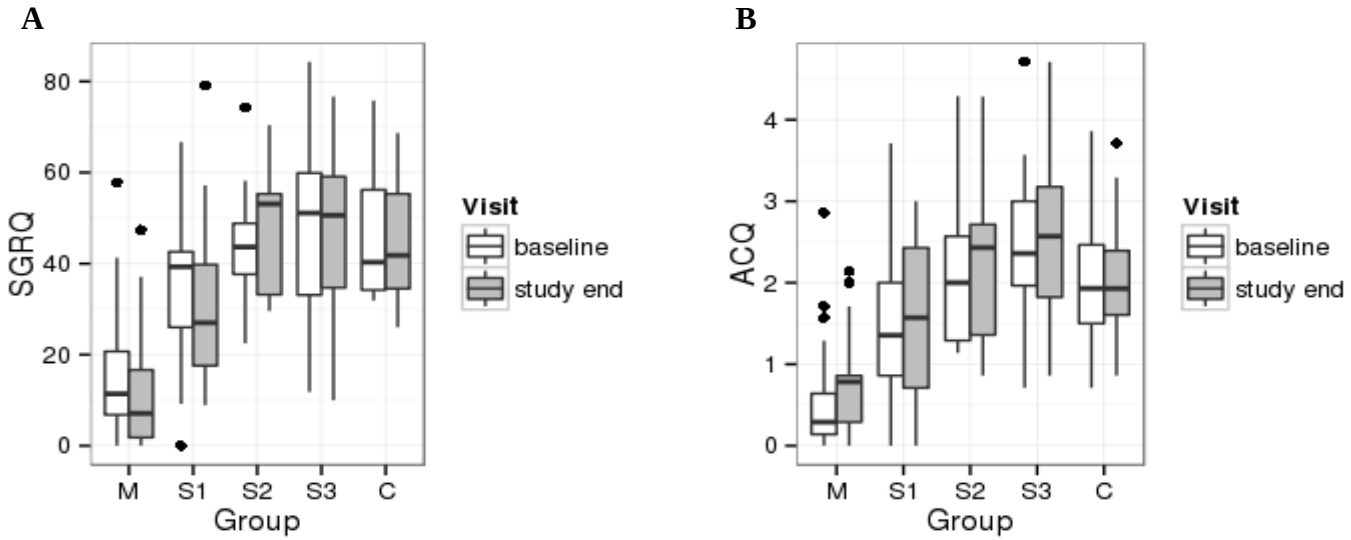

Figure E5.2. Distribution of clinical characteristics at baseline and at study end, according to groups M, S_1, S_2, S_3, and C

(A) Score of St George's Respiratory Questionnaire, (B) Score of Asthma control Questionnaire

Group M includes mild-to-moderates asthmatics, groups S_1, S_2, and S_3 include severe asthmatics, group C includes COPD patients

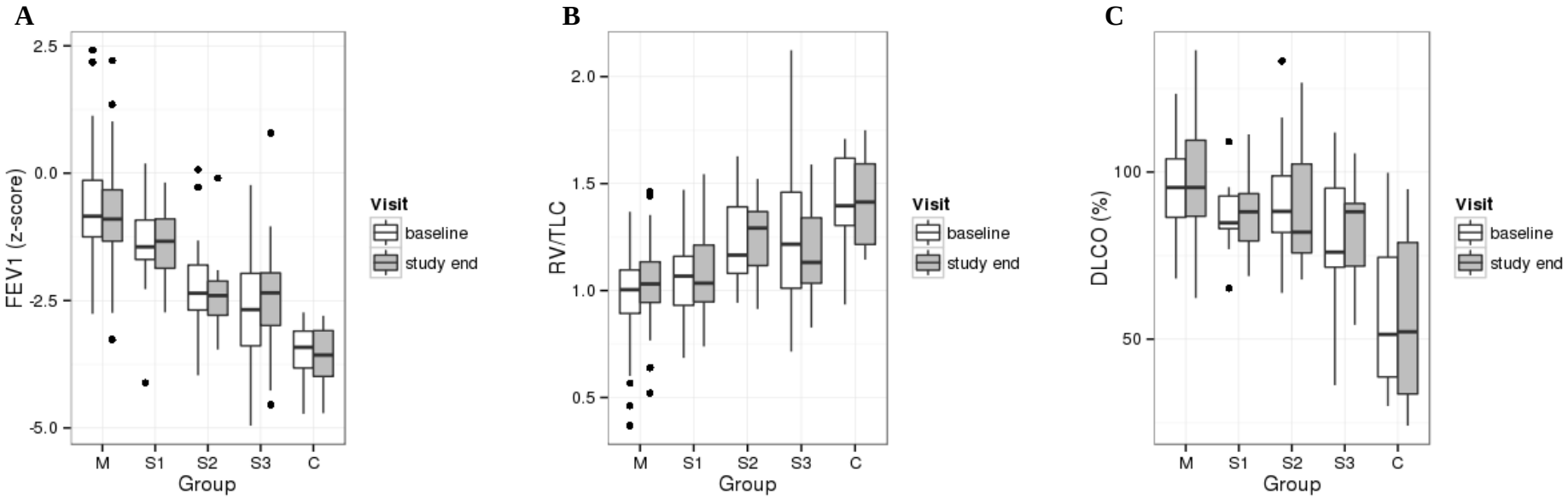

Figure E5.3. Distribution of lung function features at baseline and at study end, according to groups M, S_1, S_2, S_3, and C

(A) FEV_1, (B) RV/TLC, (C) D_{LCO}

Group M includes mild-to-moderates asthmatics, groups S_1, S_2, and S_3 include severe asthmatics, group C includes COPD patients

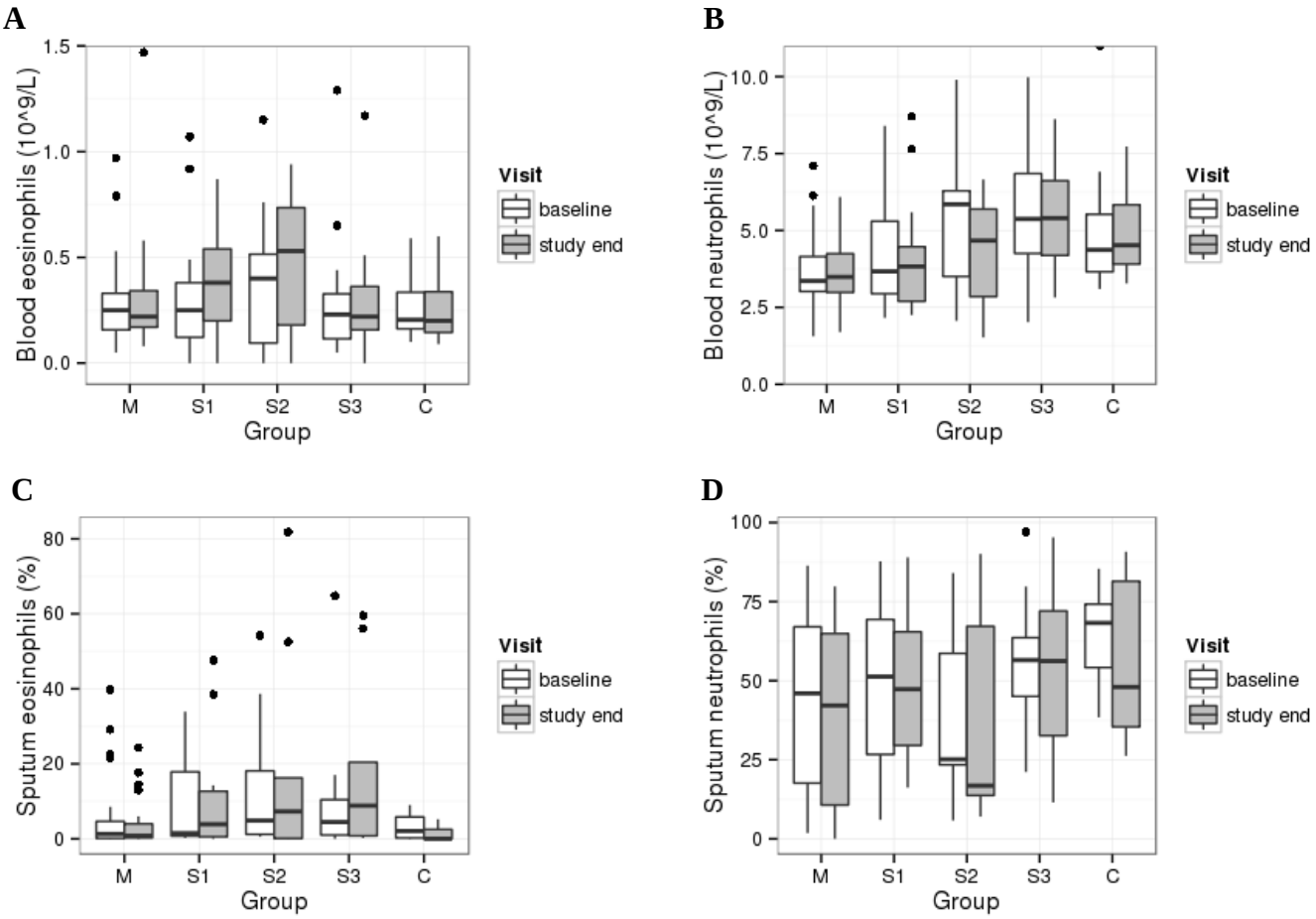

Figure E5.4. Distribution of inflammatory biomarkers at baseline and at study end, according to groups M, S_1, S_2, S_3, and C

(A) Blood eosinophils, (B) Blood neutrophils, (C) Sputum eosinophils, (D) Sputum neutrophils

Group M includes mild-to-moderates asthmatics, groups S_1, S_2, and S_3 include severe asthmatics, group C includes COPD patients

Table E5.3. Characteristics of patients at the end of follow-up (visit 6) according to subgroups M, S_1, S_2, S_3, and C (n=111)

	Group M (N=39)	MD	**Group S_1 (N=18)**	MD	**Group S_2 (N=16)**	MD	**Group S_3 (N=18)**	MD	**Group C (N=20)**	MD	**p-value** *
Airway disease	**Mild-to-moderate asthma**		**Severe asthma**		**Severe asthma**		**Severe asthma**		**COPD**		
Clinical characteristics											
QoL, SGRQ	7.1 [1.8;16.7] ‡	8	27.0 [17.6;39.8]	2	53.1 [33.2;55.3] †	4	50.6 [34.7;59.1] †	1	41.8 [34.5;55.3] †	7	**<0.001**
ACQ, Juniper	0.8 [0.3;0.9]	1	1.6 [0.7;2.5]	1	2.4 [1.4;2.7] †	1	2.6 [1.8;3.2] †	2	NA		**<0.001**
Lung function											
FEV_1, z-score	-0.8±1.1 ‡	1	-0.4±0.8 ‡	1	-2.4±0.8 † ‡	1	-2.4±1.3 † ‡	1	-3.6±0.6 †	2	**<0.001**
FVC, z-score	-0.01±0.8 ‡	1	-0.9±0.8 †	1	-1.0±0.6 †	1	-1.2±1.2 †	1	-1.6±0.6 †	2	**<0.001**
FEV_1/FVC	-1.3±1.2 ‡	1	-0.9±1.2 ‡	1	-2.3±1.0 † ‡	1	-2.2±1.4 † ‡	1	-3.9±0.7 ‡	2	**<0.001**
DLCO, %predicted corrected	95.4 [86.8;109.5]	6	88.1 [79.4;96.5]	6	82.0 [75.8;102.4]	2	88.1 [71.8;90.6]	2	52.2 [33.6;78.9]	4	**<0.001**
FRC, %predicted corrected	101.6±19.9 ‡	6	107.9±30.3	7	95.5±22.0 ‡	3	92.7±20.9 ‡	3	128.4±30.6 †	6	**0.001**
IVC, %predicted corrected	107.9±13.2 ‡	4	96.7±12.5	3	95.3±10.6	6	97.5±18.5	3	90.2±10.6 †	4	**<0.001**
TLC, %predicted corrected	107.6±12.6	3	101.8±14.5	2	1015.0±11.9	3	101.8±15.1	2	114.1±17.1	3	0.07
RV, %predicted corrected	111.2±29.7 ‡	3	114.7±39.1 ‡	2	132.5±30.3	3	118.9±30.5 ‡	2	163.0±45.6 †	3	**<0.001**
RV/TLC	1.0±0.2 ‡	3	1.1±0.2 ‡	2	1.3±0.2 †	3	1.2±0.2 ‡	2	1.4±0.2 †	3	**<0.001**
Inflammatory response											
Sputum cells, $\times 10^6$	0.7 [0.3;1.8]	19	1.2 [0.5;5.3]	8	0.8 [0.5;2.3]	7	1.3 [0.4;2.1]	9	3.3 [1.0;6.3]	14	0.50
Sputum eosinophils, %	0.8 [0.3;4.0]	20	3.9 [0.6;12.7]	8	7.3 [0.1;16.3]	7	8.8 [0.8;20.4]	9	0.0 [0.0;2.5]	15	0.17
Sputum eosinophilia ≥ 2%	6 (31.6%)	20	5 (50.0%)	8	6 (66.7%)	7	6 (66.7%)	9	2 (40.0%)	15	0.33
Sputum neutrophils, %	42.2 [10.7;64.9]	20	47.3 [29.5;65.5]	8	16.8 [13.7;67.2]	7	56.2 [32.6;72.0]	9	48.0 [35.4;81.4]	15	0.35
Sputum neutrophilia ≥ 40%	10 (52.6%)	20	5 (50.0%)	8	3 (33.3%)	7	6 (66.7%)	9	3 (60.0%)	15	0.71
Mixed granulocytic inflammation, %	3 (15.8%)	20	2 (20.0%)	8	2 (22.2%)	7	3 (33.3%)	9	1 (20.0%)	15	0.89
hs-CRP, mg/l	2.4 [1.0;5.4]	4	3.6 [3.1;4.6]	6	3.1 [0.8;6.7]	1	4.7 [2.4;5.7]	2	3.1 [1.1;6.5]	2	0.50
White blood cells, $\times 10^9$/l	6.3±1.3 ‡	2	6.9±2.1	1	8.0±2.5 †	1	9.0±2.8 †	2	7.8±1.8 †	2	**<0.001**
Blood eosinophils, $\times 10^9$/l	0.2 [0.2;0.3]	3	0.4 [0.2;0.5]	1	0.5 [0.2;0.7]	1	0.2 [0.2;0.4]	2	0.2 [0.1;0.3]	2	0.21
Blood neutrophils, $\times 10^9$/l	3.5 [3.0;4.2] ‡	3	3.8 [2.7;4.5]	2	4.7 [2.9;5.7]	1	5.4 [4.2;6.6] †	2	4.5 [3.9;5.8] †	2	**<0.001**

Values shown are mean ± standard deviation, median [25th percentile; 75th percentile], and numbers (percentages)

ACQ, Asthma Control Questionnaire; hs-CRP, high-sensitivity C-reactive protein; D_{LCO}, diffusing capacity of the lung for carbon monoxide; FEV_1, forced expiratory volume in one second; FRC, forced residual volume; FVC, forced vital capacity; IVC, inspiratory vital capacity; MD, missing data; NA, not applicable; QoL, quality of life; RV, residual volume; SGRQ, St George's Respiratory Questionnaire; TLC, total lung capacity

*Comparison between groups using the one-way ANOVA or the Kruskal-Wallis test, as appropriate, for continuous variables, and using the Chi^2 or the Fisher's exact test, as appropriate, for categorical variables; †As compared to group M; ‡As compared to group C

Differences between groups regarding specific biomarkers

Groups M, S_1, S_2, S_3, and C significantly differed regarding DPPIV, MMP-3, sRAGE, Chitinase 3-Like 1. Pair-wise multiple comparisons showed significantly higher chitinase levels in group C (37380 [21420;54580]) compared to group M (15470 [12080;18410], p-value=0.006) and to group S_3 (15400 [10710;25020], p-value=0.049). All the other pair-wise multiple comparisons were not significant.

Discussion

Representativeness of the analysis population as compared to the entire population

Excluded severe asthmatics had significantly lower levels of $zFEV_1$ and zFVC compared to severe asthmatics from the analysis population (-2.7±1.3 *vs.* -2.0±1.3, p=0.01 and -1.8±1.3 *vs.* -1.1±1.2, p=0.02, respectively). That might have balanced out the differences in lung function with the mild-to-moderate asthmatics. However, we still found significantly lower levels of $zFEV_1$ and zFVC in groups S_2 and S_3 compared to group M.

Similarly, excluded COPD patients had significantly lower levels of zFVC, FRC, and IVC compared to COPD patients from the analysis population (-2.3±1.2 *vs.* -1.4±0.8, p=0.003; 94.9% [80.3%;125.4%] *vs.* 126.2 [104.0%;147.1%], p=0.01, and 72.7% [64.0%;90.7%] *vs.* 90.2% [81.7%;96.7%] , p=0.008, respectively). That might have balanced out the differences in lung function with the mild-to-moderate asthmatics. However, we still found significantly lower levels of zFVC, FRC, and IVC in group C compared to group M.

Inflammatory response to OCS

Before OCS intervention, groups significantly differed according to blood neutrophils, sputum neutrophils, FeNO, and CRP. Groups S_2, S_3, and C globally exhibited a higher degree of inflammation compared to groups M and S_1. In response to OCS, all groups exhibited a satisfactory anti-inflammatory response. The decrease in the inflammatory biomarkers levels was even higher while the degree of inflammation before OCS intervention was high. Consequently, there was a normalization of the inflammatory biomarkers levels in each group to that of the group M after the OCS intervention.

Stability of characteristics of phenotypes

It is interesting to note, that the lung mechanical characteristics of phenotypes M, S_1, S_2, S_3, and C are stable over time, whereas their inflammatory characteristics are more variable,

which is in accordance with previous findings documented in BIOAIR (Kupczyk, Dahlen et al. 2014). It was not the purpose of this study to investigate whether FBC performed at baseline would reveal similar clusters when performed at study exit. Nevertheless, our data support the hypothesis that the FBC clusters are characterized by similar lung mechanical properties at the start and end of the observation period. Furthermore, dissociations between airway dynamics and airway inflammation has been seen also in other studies, highlighting that structure and function of airways relates to more than inflammatory cell profiles (Grainge, Lau et al. 2011).

Differences between groups regarding specific biomarkers

As described, four of the serum biomarkers measures showed significantly different levels among groups M, S_1, S_2, S_3 and C. The expression of Dipeptidyl peptidase-4 (DPPIV, or CD26) is increased following IL-13 stimulation in bronchial epithelial cells and its levels in the circulation have therefore been proposed to be a biomarker of type-2 driven inflammation and response to anti-IL-13 therapy (Lancet Respir Med. 2015;3:692). In the current investigation its levels are highest in group M and lowest in group C, yet do not follow the same pattern as other proposed markers of type-2 inflammation such as blood or sputum eosinophils, or exhaled NO. However, although not extensively investigated, previous investigations of DPP4 as a circulating biomarker do suggest that generally, its levels are lower in more severe, steroid dependent asthma compared to milder disease (Ranade et al. ATS poster 2016, James et al. ERS poster 2016) which may be in line with the current observations.

Matrix metalloproteinase-3 is one member of a family of extracellular matrix degrading enzymes which are believed to play a role in airway disease due to effects on tissue remodeling and repair and their ability to regulate the kinetics and function of inflammatory cells. Although circulating MMP-3 has not been thoroughly investigated as a biomarker of airway disease, one characteristic of possible relevance to the current findings is the fact that serum MMP-3 levels are strongly increased by corticosteroid use (Hathout et al. Sci Rep. 2016; 6: 31727). Serum MMP-3 levels were greatest in groups S_2, S_3 and C, which were those taking the highest doses of corticosteroids.

In human and animal studies of airway disease, RAGE and its ligands are often increased, whereas soluble RAGE is decreased. Inverse associations between circulating concentrations of total soluble RAGE, and surrogate markers of disease risk or burden observations in other

chronic inflammatory conditions have led to the view that soluble RAGE is somehow a protective factor (Sukkar, Ullah et al. 2012). It has been shown that a deficiency in sRAGE is specifically associated with neutrophilic airway inflammation in asthma and COPD (Sukkar, Wood et al. 2012), which is in accordance with the neutrophilic-dominant inflammation observed in groups S_3 and C. While it is not known whether deficiency in sRAGE occurs as a consequence of neutrophilic inflammation, or whether it is a causative factor that underlies neutrophilic inflammation, these findings highlight the possibility that sRAGE might be a useful biomarker, and a possible future therapeutic target in severe, neutrophilic asthma.

The chitinase-like protein YKL-40 is known to be increased in both severe asthma and COPD, where it has been associated with neutrophilic inflammation, but also markers of airway remodeling such as bronchial wall thickness (Chupp, Lee et al. 2007, Konradsen, James et al. 2013, Hinks, Brown et al. 2016, James, Reinius et al. 2016). Accordingly, YKL-40 has also been shown to increase the proliferation of cultured human airway smooth muscle cells *in vitro* (Bara, Ozier et al. 2012). Although not statistically significant, it is of interest that the pattern of serum YKL-40 appears to be different among S_1, S_2 and S_3, whereby levels are highest in S_2. Although this is not the group with the highest sputum neutrophils or lowest diffusion capacity, one may speculate that this could be related to the fact that these subjects are older, have high blood neutrophils, have had their respiratory disease for longer, or are the least reversible, possibly suggesting a more fixed airway obstruction, although the underlying reasons require further validation.

6. General discussion

This PhD thesis investigates the fluctuation behavior of heart and respiratory system signals, and how it changes with long-term environmental exposures and chronic diseases. The investigations of the fluctuation behavior of physiological systems using mathematical tools have already provided significant new insights into disease pathogenesis. Until recently, however, such mathematical tools have not been implemented in large epidemiological or clinical datasets, allowing for the control of the most potential confounders. This translation of involved mathematical techniques into the clinical and epidemiological fields underscores the uniqueness of this work. It results from stimulating teamwork, in a group with mixed expertise in computational, clinical physiology, and epidemiology. This valuable exchange enabled us to develop original approaches to unravel the important effects of environmental factors on the cardiovascular system, as well as to unveil relevant phenotypes of patients with severe asthma. With the emergence of new fields of research, such as systems biology and systems medicine, such interdisciplinarity is becoming essential.

6.1. Main findings

This PhD thesis demonstrates original applications in the assessment of dynamics of cardiovascular and respiratory systems in health and disease, and provides relevant new findings.

The first application evaluates the long-term influence of smoking cessation on heart rate variability and heart rate dynamics, in an aging general population, using the subpopulation of lifelong non-smokers as control group. This application is an illustration of how complexity in biological signals can be measured, with the constitution of a “toolkit” of parameters to probe different aspects of the signals dynamics; here the cardiac interbeat interval dynamics. We investigated whether we could objectify perturbations in heart rate dynamics of current smokers as compared to lifelong non-smokers, and whether there was a normalization of the dynamics after smoking cessation. We were able to provide evidence that:

(1) Smoking triggers adverse changes in the regulation of the cardiovascular system, even at low levels of exposure, with a dose-response effect. The effect of current smoking was suggested with standard measures of HRV, and strengthened by measures derived from nonlinear theories. Moreover, we observed that power spectral density, $\alpha_{\text{short-term time scale}}$, multiscale entropy, and largest Lyapunov exponent were significantly modified in current heavy smokers, as compared to lifelong non-smokers, but not in current light smokers. This

finding suggested that more properties of the dynamics got altered when the smoking exposure increased. Namely, heavy exposure might trigger specific alterations in the dynamics of the cardiovascular system, in addition to those triggered at lower levels of exposure.

(2) Light smokers fully recover within the first 15 years of cessation.

(3) Heavy smokers also fully recover, but might need up to 15 to 25 years. Our findings suggested a full normalization of the Lyapunov Largest Exponent after only 25 years of cessation in former heavy smokers. This supports the hypothesis that nonlinear time series analysis techniques may be able to unveil subtle, but important, changes in the regulation of the cardiovascular system; more difficult to detect by traditional analysis methods. To the extent that we may translate perturbations in the heart rate dynamics to an increase in the risk of coronary heart disease, this finding is consistent with Teo et al.'s findings, which identified a still increased risk of acute myocardial infarction in former heavy smokers after 20 years of smoking cessation (Teo, Ounpuu et al. 2006).

The second application evaluates the influence of long-term exposure to TPM_{10} on HRV and heart rate dynamics. While we did not find any overall association in the entire study population, we observed strong significant associations of long-term exposure to TPM_{10} with the HRD parameters in subjects without cardiovascular morbidity, and even stronger associations, with both HRV and HRD parameters, in non-obese subjects without cardiovascular morbidity. These findings suggest that the relative contribution of both the underlying health condition and the countering effects of drug treatments on the TPM_{10}–HRV/HRD relationship might render this relationship so variable that the overall TPM_{10}–HRV/HRD relationship in such subjects might be null. Therefore, adverse effects of TPM_{10}, even if they are present in subjects with comorbidity, might be more visible in healthy subjects. Additionally, our findings are in the line with previous studies that have provided evidences that TPM_{10} might impact in part through oxidative stress pathways. Finally, the fact that adverse effects of TPM_{10} were revealed in subjects without cardiovascular morbidity, only by HRD parameters, supports the hypothesis that, measuring changes in complexity in heart rate dynamics during exposure to environmental elements, might unveil subtle but important changes in the regulatory mechanisms of the cardiovascular system not detectable by traditional analysis methods.

The third application evaluates whether the subgrouping of patients with chronic obstructive airway diseases, including mild-to-moderate asthma, severe asthma, and COPD, according to

their profile of airway dynamics, allows for the identification of phenotypes with specific treatable traits. This application is an illustration of how complexity in biological signals can be compared, and used for disease phenotyping. We investigated how an unlabeled data set with patients with mild-to-moderate asthma, severe asthma, and COPD, organizes into groups, on the basis of patients' lung function fluctuation. Combination of the resulting lung function fluctuation based clusters with the initial clinical classification (i.e., mild-to-moderate asthma, severe asthma, and COPD) allowed for the identification of 5 groups (M, S_1, S_2, S_3, and C) corresponding to relevant phenotypes. Phenotypes were characterized by a gradually increasing clinical severity and functional alteration of the lung from M to C, with a high exacerbation risk in S_2 and S_3. Response to OCS differed according to groups and seemed to result from the relative pathophysiological contributions of airway obstruction, inflammation, and irreversible mechanical impairment. Response in M was weak, probably due to a ceiling effect. Indeed, since this group of patients had minimal airway obstruction, they were probably controlled with inhaled corticosteroids, reducing any benefit of adding OCS. From S_1 to S_3, where degree of obstruction and signs of inflammation gradually increased, we found a gradually increasing response to OCS, with a particularly good response in S_3. In C, irreversible mechanical impairment of the lung might have rendered patients clinically unresponsive to OCS despite a satisfactory anti-inflammatory response. Thus, in these patients the mechanical impairment dominated the clinical picture. Our approach provided evidence that airway dynamics contain substantial information, which enables the identification of phenotypes, in which the functional alteration of the lung translates into specific pathological features and clinically meaningful outcomes.

6.2. Strengths and limitations

Particular strengths of our applications were the original approaches used to answer original research questions, with data from two large and unique datasets.

6.2.1. Effects of long-term environmental exposures on heart rate variability and heart rate dynamics

A unique dataset

The SAPALDIA (Swiss Cohort Study on Air Pollution and Lung and Heart Disease in Adults) study was designed to assess the health effects of long-term exposure to air pollutants in the Swiss adult population. Main strengths of this study include the population-based

design, the large number of participants who underwent a 24-hour electrocardiogram Holter recording to assess HRV, the 10-year follow-up period and detailed information available on participants, allowing for the control of most potential confounders.

Original research questions

The unique design of the SAPALDIA study enabled us to examine, for the first time, the influence of long-term smoking cessation, as well as the influence of long-term exposure to TPM_{10} on the regulation of the cardiovascular system and heart rate dynamics.

An original approach

To answer these research questions, we calculated traditional measures of HRV, and, in addition, we generated a toolkit of parameters derived from nonlinear dynamics methods in order to probe different dynamics properties of heart rate variability. This approach allowed us to strengthen findings from the traditional measures of HRV, to unveil long-term alterations caused by heavy smoking exposure, as well as alterations caused by long-term exposure to TPM_{10} in the subjects without cardiovascular morbidity.

Limitations

A limitation of this work is the absence of a physiological interpretation of the parameters calculated with methods from nonlinear dynamics. Physiological interpretation of such metrics constitutes a major limitation for their use (1996, Goldberger, Amaral et al. 2000, Francesco, Maria Grazia et al. 2012, Manor and Lipsitz 2013). Though it is reasonable to assume that these concepts from mathematics could help gain insight into the regulatory mechanisms of physiological systems, efforts are needed to improve our understanding of their physiological correlates. In the case of the present work, this uncertain knowledge limited the interpretation of associations between parameters and risk of cardiac events.

6.2.2. Lung function fluctuation based phenotypes in asthma and COPD

A unique design

The Pan-European BIOAIR (Longitudinal Assessment of Clinical Course and BIOmarkers in Severe Chronic AIRway Disease) study was designed to characterize the course of severe chronic airway diseases over time. The unique aspect of this study was the twice-daily collection of lung function measurements over a one-year period. Additional

strengths of this study include the mixed population of adults with mild-to-moderate asthma, severe asthma, and COPD; a 2-week double blind placebo-controlled oral corticosteroid intervention, allowing for the assessment of response to treatment; and the detailed information available about the patients.

A currently unmet need

For clinicians, the identification of asthma and COPD phenotypes related to specific treatable traits is of primary concern. Especially in severe asthma (Heaney and Robinson 2005, Wenzel 2012), COPD, and the transition forms between these entities (Chung 2013), in which the unpredictable nature of exacerbations and the heterogeneity of response to drug therapy present a major clinical challenge (Moore and Peters 2006, Donaldson, Seemungal et al. 2012, Kupczyk, Haque et al. 2013, Phipatanakul, Mauger et al. 2016).

An original approach

The BIOAIR study, due to its unique design, enabled, for the first time, a comparison of patients with mild-to-moderate asthma, severe asthma, and COPD, on the basis of their airway dynamics over a one-year period. This was achieved using a novel clustering approach, developed by our group (Delgado-Eckert, Fuchs et al. 2017), called fluctuation-based clustering (FBC). Classical clustering approaches usually rely on a cross-sectional bunch of clinical and biological variables (Haldar, Pavord et al. 2008, Smith, Drake et al. 2008, Weatherall, Travers et al. 2009, Moore, Meyers et al. 2010, Fitzpatrick, Teague et al. 2011, Siroux, Basagana et al. 2011, Just, Gouvis-Echraghi et al. 2012, Boudier, Curjuric et al. 2013, Moore 2013, Prosperi, Sahiner et al. 2013, Schatz, Hsu et al. 2013, Wu, Bleecker et al. 2014). However, characterizing patients with dynamical diseases, such as asthma and COPD, at a single point in time, is prone to misclassification. Instead, serial measurements of a single biomarker, as used in the FBC approach, may enable a more accurate classification of patients, and could better account for the temporal stability of a given phenotype. Furthermore, since fluctuation in FEV_1 describes the patient's response to day-to-day real life stimuli, the FBC approach may account for the interaction with the given environment over the observation time period. This is particularly important for asthma phenotyping, since both intrinsic features of the disease and environmental stimuli might determine disease phenotypes.

Limitations

This work faced some limitations. First, the FBC approach is not based on correlation properties of the lung function measurements, but on their distribution (Delgado-Eckert, Fuchs et al. 2017). Consequently, it neglects the time dimension, but it gains robustness with respect to missing data. Indeed, calculating correlation from data with missing data points would be prone to error, and thus, would not be appropriate in clinical context, where missing data are a frequent issue. Second, we were not able to assess the stability of the phenotypes identified. Namely, to assess whether a similar FBC analysis performed at another time point would generate similar phenotypes and how allocation to clusters would change. Given that the FBC analysis was performed on the entire follow-up period, we were not able to repeat the analysis at another time point.

6.3. Clinical and public health relevance and recommendations

The investigation of the change in complexity dynamics of physiological signals has many possible applications that are of clinical and public health relevance, in a wide range of domains, such as aging, disease, and environment. As part of the present work, our investigation of the effect of environmental exposures on HRV and heart rate dynamics, as well as the investigation of lung function fluctuation behaviour for asthma and COPD phenotyping, provided findings of relevance for public health and clinical research.

6.3.1. Effects of long-term environmental exposures on heart rate variability and heart rate dynamics: applications of relevance for public health

Findings from our study related to smoking cessation and HRV/HRD support the substantial benefits of smoking cessation, but also warn of important alterations caused by heavy smoking. It constitutes a strong argument for health policy makers advocating for more intensive prevention campaigns aimed at discouraging smoking, and underpins the value of public healthcare programs supporting the benefits of smoking cessation. Furthermore, we could show that heavy former smokers might need up to 15-25 years to fully recover after smoking cessation. Thus, former heavy smokers remain exposed longer after cessation to a higher risk of cardiovascular morbidity and cardiovascular-related morbidity. Analogous to the recommendations of the American Cancer Society (2014) related to lung cancer screening, our data suggest that close monitoring of cardiovascular disease in current and former heavy

smokers might be warranted. In such patients, characterization of HRV and heart rate dynamics might be relevant for cardiac events risk stratification after smoking cessation.

Findings from our study related to long-term exposure to TPM_{10} and HRV/HRD provides evidence of adverse effects of air pollution in healthy subjects, believed to be less susceptible than specific subpopulations with comorbidities (e.g., the elderly, patients with preexisting cardiovascular disease or diabetes, obese subjects) though. This constitutes a strong argument for health policy makers advocating for more intensive prevention campaigns aimed at reducing traffic-related pollution. Additionally, our findings suggest that the TPM_{10}-HRV/HRD relationship in subjects with cardiovascular morbidity might be modified by both the underlying cardiovascular condition and the related treatments. Thus, some cardiac therapies, for a given underlying cardiovascular condition, might be protective against the adverse cardiac effects of pollution, whereas some other cardiac therapies/conditions might render subjects particularly susceptible to those effects.

6.3.2. Lung function fluctuation based phenotypes in asthma and COPD: an application of clinical relevance

Research implications

Asthma and COPD patients' phenotyping based on airway dynamics might, in the near future, have relevant research applications. First, further investigations of such phenotypes, and characterization of related endotypes, might help in our understanding of the underlying mechanisms of disease pathogenesis, leading to more targeted therapies and personalized approaches to asthma management.

Future study designs might include phenotypes-based interventions, such as:

- The characterization of structural alterations of the lung, using imaging techniques, in patients with severe phenotypes. Structural alterations may reflect specific underlying pathophysiological mechanisms and their investigation, specifically in severe phenotypes, might improve our understanding of disease pathogenesis. Choi et al. identified four clusters very similar to our phenotypes M, S_1, S_2, and S_3, using an imaging-based clustering approach (Choi, Hoffman et al. 2017). In particular, our group S_2 was similar to their luminal narrowing-dominant cluster, and our group S_3 was similar to their wall thickening-dominant cluster.
- The evaluation of new treatment strategies. The severe phenotypes S_2 and S_3 identified in the BIOAIR study were characterized by patients with high

exacerbation risk and a good response to oral corticosteroids. These patients might particularly benefit from a controller medication, especially from more targeted immunosuppressant treatments.

Clinical implications

In the longer-term, we also see promising clinical applications. In particular, the implementation of lung function measurements in telemonitoring settings for:

- Diagnostic purposes - In a situation of unclear asthma history, unclear graduation of asthma severity, or suspicion of asthma-COPD-overlap-syndrome (ACOS), a twice-daily lung function monitoring for a given period of observation might help determine which diagnosis would support the observed lung function fluctuations.
- Monitoring purposes - In patients with severe phenotypes, especially if there is a high risk exacerbation, periods of closer monitoring could be recommended, for instance after implementing a new treatment strategy.

The implementation of the FBC approach in telemonitoring settings appears feasible. While conventional disease phenotyping usually relies on many characterizing parameters, which tend to be expensive and limited to in-hospital assessment, fluctuation of FEV_1 can be implemented in a simple and cost-effective way in a telemonitoring setting with an appropriate adherence measure (Kupczyk, Haque et al. 2013). Moreover, in order to increase its feasibility and the clinical applicability, the FBC approach includes a data-driven algorithm which determines the tolerable amount of missing measurements. Finally, our lung function based clustering could be repeated in a large database, generated from existing datasets of patients with chronic obstructive airway diseases. After validation of the phenotypes identified, this database could be used as a reference database, to automatize data analysis. Thus, a patient with a complete telemonitoring dataset could be instantaneously attributed to a phenotype.

6.4. Outlook

6.4.1. Effects of long-term environmental exposures on heart rate variability and heart rate dynamics

Regarding the investigation of the influence of smoking cessation on HRV/HRD, further studies are needed:

- to improve our understanding of the physiological correlates of the modifications of dynamics properties triggered by smoking exposure;
- to investigate whether the late normalization in former heavy smokers, especially in the Lyapunov Largest Exponent, suggested by our findings is observed in other datasets, and whether a persistent decrease in Lyapunov Largest Exponent might be associated with an increased risk of coronary heart disease;
- to evaluate the benefits of close monitoring of cardiovascular disease in current and former heavy smokers, and whether, in such patients, characterization of heart rate dynamics might be relevant for cardiac events risk stratification after smoking cessation.

Regarding the investigation of the influence of long-term exposure to TPM_{10} on HRV/HRD, further studies are needed:

- to see whether these alterations in HRV/HRD in healthy people lead to increased mortality and morbidity later in life;
- to investigate how the TPM_{10}-HRV/HRD relationship in subjects with cardiovascular morbidity is modified depending on the underlying cardiovascular condition and the related drug treatments.

6.4.2. Lung function fluctuation based phenotypes in asthma and COPD

Regarding the lung function fluctuation phenotyping in asthma and COPD, our group is planning the following future investigations as part of the BIOAIR study:

- The examination of the minimal window of observation needed to ensure the correct phenotyping of patients in order to facilitate implementation into clinical practice and telemonitoring settings;
- The exploration of the long-term stability of the phenotypes. So far, long-term stability of asthma phenotypes is poorly understood. The determination of the minimal window of observation needed for the correct phenotyping of patients would allow for the phenotyping of patients and the assessment of the temporal evolution of the phenotypes, using a gliding window along the follow-up period. The size of the gliding window will be determined by the size of the minimal observation window needed to ensure the correct phenotyping of patients. The

stability of the clusters will be assessed using Jaccard's similarity coefficient, which is a measure of overlap between groups. Moreover, the cluster membership of individual participants will be traced as a function of time.

- The identification of factors likely to influence phenotypting. Especially, the observation period (e.g., influence of seasonality), the time and frequency of the measurements (e.g., influence of the circadian rhythm).

6.5. Conclusion

This thesis attempts to demonstrate the importance of multidimensional approaches to understand the complex functioning of our physiological system and of diseases process. Characterization of the complexity in the fluctuation behavior of system signals holds enormous promise for providing new understandings of the regulatory mechanisms of physiological systems and how they change with diseases. However, it is important to combine this kind of approach with classical epidemiological approaches in order to disentangle the various contributions of the intrinsic physiological dynamics, aging, diseases and comorbidities, lifestyle, and environment. In the SAPALDIA cohort study, we were able to disentangle the influence of specific environmental exposures, such as particulate matter air pollution and smoking exposure, on the HRV and heart rate dynamics, and thus to unveil long-term alterations in former heavy smokers, as well as adverse effects of low level, but long-term, exposure to TPM_{10} in healthy subjects and in subjects with homozygous GSTM1 gene deletion. In the BIOAIR study, we provide evidence that airway dynamics contain substantial information, which enables the identification of clinically meaningful phenotypes, in which the functional alteration of the lung translates into specific treatable traits.

7. References

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

(1996). "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology." Eur Heart J **17**(3): 354-381.

(1999). "Recommendations for standardized procedures for the on-line and off-line measurement of exhaled lower respiratory nitric oxide and nasal nitric oxide in adults and children-1999. This official statement of the American Thoracic Society was adopted by the ATS Board of Directors, July 1999." Am J Respir Crit Care Med **160**(6): 2104-2117.

(2014). "Lung cancer screening recommended for heavy smokers." Cancer Discov **doi: 10.1158/2159-8290.CD-NB2014-007**.

(2015). "Global Strategy for Asthma Management and Prevention, Global Initiative for Asthma (GINA)."

Ackermann-Liebrich, U., B. Kuna-Dibbert, N. M. Probst-Hensch, C. Schindler, D. Felber Dietrich, E. Z. Stutz, L. Bayer-Oglesby, F. Baum, O. Brandli, M. Brutsche, S. H. Downs, D. Keidel, M. W. Gerbase, M. Imboden, R. Keller, B. Knopfli, N. Kunzli, L. Nicod, M. Pons, P. Staedele, J. M. Tschopp, J. P. Zellweger and P. Leuenberger (2005). "Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991-2003: methods and characterization of participants." Soz Praventivmed **50**(4): 245-263.

Adam, M., D. Felber Dietrich, E. Schaffner, D. Carballo, J. C. Barthelemy, J. M. Gaspoz, M. Y. Tsai, R. Rapp, H. C. Phuleria, C. Schindler, J. Schwartz, N. Kunzli and N. M. Probst-Hensch (2012). "Long-term exposure to traffic-related PM(10) and decreased heart rate variability: is the association restricted to subjects taking ACE inhibitors?" Environ Int **48**: 9-16.

Adam, M., M. Imboden, E. Boes, E. Schaffner, N. Kunzli, H. C. Phuleria, F. Kronenberg, J. M. Gaspoz, D. Carballo and N. Probst-Hensch (2014). "Modifying effect of a common

polymorphism in the interleukin-6 promoter on the relationship between long-term exposure to traffic-related particulate matter and heart rate variability." PLoS One **9**(8): e104978.

Adam, M., M. Imboden, E. Schaffner, E. Boes, F. Kronenberg, M. Pons, R. Bettschart, J. C. Barthelemy, C. Schindler and N. Probst-Hensch (2017). "The adverse impact of obesity on heart rate variability is modified by a NFE2L2 gene variant: The SAPALDIA cohort." Int J Cardiol **228**: 341-346.

Agresti, A. (1992). "A survey of exact inference for contingency tables." Statistical Science 131-153.

Alvarez-Ramirez, J., E. Rodriguez, and J. Carlos Echeverría (2005). "Detrending fluctuation analysis based on moving average filtering." Physica A: Statistical Mechanics and its Applications **5**(1): 82-87.

Baccarelli, A., P. A. Cassano, A. Litonjua, S. K. Park, H. Suh, D. Sparrow, P. Vokonas and J. Schwartz (2008). "Cardiac autonomic dysfunction: effects from particulate air pollution and protection by dietary methyl nutrients and metabolic polymorphisms." Circulation **117**(14): 1802-1809.

Bara, I., A. Ozier, P. O. Girodet, G. Carvalho, J. Cattiaux, H. Begueret, M. Thumerel, O. Ousova, R. Kolbeck, A. J. Coyle, J. Woods, J. M. Tunon de Lara, R. Marthan and P. Berger (2012). "Role of YKL-40 in bronchial smooth muscle remodeling in asthma." Am J Respir Crit Care Med **185**(7): 715-722.

Barclay, J. L., B. G. Miller, S. Dick, M. Dennekamp, I. Ford, G. S. Hillis, J. G. Ayres and A. Seaton (2009). "A panel study of air pollution in subjects with heart failure: negative results in treated patients." Occup Environ Med **66**(5): 325-334.

Beuther, D. A. and E. R. Sutherland (2007). "Overweight, obesity, and incident asthma: a meta-analysis of prospective epidemiologic studies." Am J Respir Crit Care Med **175**(7): 661-666.

Bigger, J. T., Jr., J. L. Fleiss, R. C. Steinman, L. M. Rolnitzky, R. E. Kleiger and J. N. Rottman (1992). "Frequency domain measures of heart period variability and mortality after myocardial infarction." Circulation **85**(1): 164-171.

Boeing, G. (2016). "Visual analysis of nonlinear dynamical systems: chaos, fractals, self-similarity and the limits of prediction." Systems **4**(37): 1-18.

Boudier, A., I. Curjuric, X. Basagana, H. Hazgui, J. M. Anto, J. Bousquet, P. O. Bridevaux, E. Dupuis-Lozeron, J. Garcia-Aymerich, J. Heinrich, C. Janson, N. Kunzli, B. Leynaert, R. de Marco, T. Rochat, C. Schindler, R. Varraso, I. Pin, N. Probst-Hensch, J. Sunyer, F. Kauffmann and V. Siroux (2013). "Ten-year follow-up of cluster-based asthma phenotypes in adults. A pooled analysis of three cohorts." Am J Respir Crit Care Med **188**(5): 550-560.

Brook, R. D., S. Rajagopalan, C. A. Pope, 3rd, J. R. Brook, A. Bhatnagar, A. V. Diez-Roux, F. Holguin, Y. Hong, R. V. Luepker, M. A. Mittleman, A. Peters, D. Siscovick, S. C. Smith, Jr., L. Whitsel, J. D. Kaufman, E. American Heart Association Council on, C. o. t. K. i. C. D. Prevention, P. A. Council on Nutrition and Metabolism (2010). "Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association." Circulation **121**(21): 2331-2378.

Burney, P. G., C. Luczynska, S. Chinn and D. Jarvis (1994). "The European Community Respiratory Health Survey." Eur Respir J **7**(5): 954-960.

Buteau, S. and M. S. Goldberg (2016). "A structured review of panel studies used to investigate associations between ambient air pollution and heart rate variability." Environ Res **148**: 207-247.

Carolan, B. J. and E. R. Sutherland (2013). "Clinical phenotypes of chronic obstructive pulmonary disease and asthma: recent advances." J Allergy Clin Immunol **131**(3): 627-634; quiz 635.

Choi, S., E. A. Hoffman, S. E. Wenzel, M. Castro, S. Fain, N. Jarjour, M. L. Schiebler, K. Chen, C. L. Lin, L. National Heart and P. Blood Institute's Severe Asthma Research (2017). "Quantitative computed tomography imaging-based clustering differentiates asthmatic subgroups with distinctive clinical phenotypes." J Allergy Clin Immunol.

Chuang, K. J., C. C. Chan, N. T. Chen, T. C. Su and L. Y. Lin (2005). "Effects of particle size fractions on reducing heart rate variability in cardiac and hypertensive patients." Environ Health Perspect **113**(12): 1693-1697.

Chung, K. F. (2013). "New treatments for severe treatment-resistant asthma: targeting the right patient." Lancet Respir Med **1**(8): 639-652.

Chung, K. F., S. E. Wenzel, J. L. Brozek, A. Bush, M. Castro, P. J. Sterk, I. M. Adcock, E. D. Bateman, E. H. Bel, E. R. Bleecker, L. P. Boulet, C. Brightling, P. Chanez, S. E. Dahlen, R. Djukanovic, U. Frey, M. Gaga, P. Gibson, Q. Hamid, N. N. Jajour, T. Mauad, R. L. Sorkness and W. G. Teague (2014). "International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma." Eur Respir J **43**(2): 343-373.

Chupp, G. L., C. G. Lee, N. Jarjour, Y. M. Shim, C. T. Holm, S. He, J. D. Dziura, J. Reed, A. J. Coyle, P. Kiener, M. Cullen, M. Grandsaigne, M. C. Dombret, M. Aubier, M. Pretolani and J. A. Elias (2007). "A chitinase-like protein in the lung and circulation of patients with severe asthma." N Engl J Med **357**(20): 2016-2027.

Collins, L. C., P. D. Hoberty, J. F. Walker, E. C. Fletcher and A. N. Peiris (1995). "The effect of body fat distribution on pulmonary function tests." Chest **107**(5): 1298-1302.

Costa, M., A. L. Goldberger and C. K. Peng (2002). "Multiscale entropy analysis of complex physiologic time series." Phys Rev Lett **89**(6): 068102.

Costa, M., A. L. Goldberger and C. K. Peng (2005). "Multiscale entropy analysis of biological signals." Phys Rev E Stat Nonlin Soft Matter Phys **71**(2 Pt 1): 021906.

Cusenza, M., et al (2010). "Relationship between fractal dimension and power-law exponent of heart rate variability in normal and heart failure subjects." IEEE

Delgado-Eckert, E., O. Fuchs, N. Kumar, J. Pekkanen, J. C. Dalphin, J. Riedler, R. Lauener, M. Kabesch, M. Kupczyk, D. S.E., E. Mutius and U. Frey (2017). "Functional Phenotypes Determined by Fluctuation Based Clustering of Lung Function Measurements in Healthy and Asthmatic Cohort Participants." Thorax **Accepted**.

Delgado, E., N. Kumar, O. Fuchs, J. Pekkanen, J. C. Dalphin, J. Riedler, R. Lauene, A. M. Karvonen, J. Genuneit, E. Von Mutius and U. Frey (2015). "Asthma phenotypes determined by a novel fluctuation based clustering method using a time window of lung function observations." European Respiratory Journal **46**(suppl 59).

Dinas, P. C., Y. Koutedakis and A. D. Flouris (2013). "Effects of active and passive tobacco cigarette smoking on heart rate variability." Int J Cardiol **163**(2): 109-115.

Dobson, A. J., H. M. Alexander, R. F. Heller and D. M. Lloyd (1991). "How soon after quitting smoking does risk of heart attack decline?" J Clin Epidemiol **44**(11): 1247-1253.

Doll, R. and R. Peto (1976). "Mortality in relation to smoking: 20 years' observations on male British doctors." Br Med J **2**(6051): 1525-1536.

Doll, R., R. Peto, J. Boreham and I. Sutherland (2004). "Mortality in relation to smoking: 50 years' observations on male British doctors." BMJ **328**(7455): 1519.

Donaldson, G. C., T. A. Seemungal, J. R. Hurst and J. A. Wedzicha (2012). "Detrended fluctuation analysis of peak expiratory flow and exacerbation frequency in COPD." Eur Respir J **40**(5): 1123-1129.

Felber Dietrich, D., C. Schindler, J. Schwartz, J. C. Barthelemy, J. M. Tschopp, F. Roche, A. von Eckardstein, O. Brandli, P. Leuenberger, D. R. Gold, J. M. Gaspoz and U. Ackermann-Liebrich (2006). "Heart rate variability in an ageing population and its association with lifestyle and cardiovascular risk factors: results of the SAPALDIA study." Europace **8**(7): 521-529.

Felber Dietrich, D., C. Schindler, J. Schwartz, J. C. Barthelemy, J. M. Tschopp, F. Roche, A. von Eckardstein, O. Brandli, P. Leuenberger, D. R. Gold, J. M. Gaspoz, U. Ackermann-Liebrich and S. Team (2006). "Heart rate variability in an ageing population and its association with lifestyle and cardiovascular risk factors: results of the SAPALDIA study." Europace **8**(7): 521-529.

Felber Dietrich, D., J. Schwartz, C. Schindler, J. M. Gaspoz, J. C. Barthelemy, J. M. Tschopp, F. Roche, A. von Eckardstein, O. Brandli, P. Leuenberger, D. R. Gold and U. Ackermann-Liebrich (2007). "Effects of passive smoking on heart rate variability, heart rate and blood pressure: an observational study." Int J Epidemiol **36**(4): 834-840.

Feng, Y., X. Huang, H. Sun, C. Liu, B. Zhang, Z. Zhang, V. Sharma Tengur, W. Chen, T. Wu, J. Yuan and X. Zhang (2015). "Framingham risk score modifies the effect of PM10 on heart rate variability." Sci Total Environ **523**: 146-151.

Fitzpatrick, A. M., W. G. Teague, D. A. Meyers, S. P. Peters, X. Li, H. Li, S. E. Wenzel, S. Aujla, M. Castro, L. B. Bacharier, B. M. Gaston, E. R. Bleecker, W. C. Moore, L. National Institutes of Health/National Heart and P. Blood Institute Severe Asthma Research (2011). "Heterogeneity of severe asthma in childhood: confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program." J Allergy Clin Immunol **127**(2): 382-389 e381-313.

Francesco, B., B. Maria Grazia, G. Emanuele, F. Valentina, C. Sara, F. Chiara, M. Riccardo and F. Francesco (2012). "Linear and nonlinear heart rate variability indexes in clinical practice." Comput Math Methods Med **2012**: 219080.

Frey, U., T. Brodbeck, A. Majumdar, D. R. Taylor, G. I. Town, M. Silverman and B. Suki (2005). "Risk of severe asthma episodes predicted from fluctuation analysis of airway function." Nature **438**(7068): 667-670.

Frey, U., P. Latzin, J. Usemann, J. Maccora, U. Zumsteg and S. Kriemler (2015). "Asthma and obesity in children: current evidence and potential systems biology approaches." Allergy **70**(1): 26-40.

Frey, U. and B. Suki (2008). "Complexity of chronic asthma and chronic obstructive pulmonary disease: implications for risk assessment, and disease progression and control." Lancet **372**(9643): 1088-1099.

Gac, P. and M. Sobieszczanska (2014). "Effects of cigarette smoke on Holter ECG recordings in patients with arterial hypertension. Part 1: Time domain parameters of heart rate variability." Environ Toxicol Pharmacol **37**(1): 404-413.

Girard, D., E. Delgado-Eckert, E. Schaffner, C. Hacki, M. Adam, G. L. Stern, N. Kumar, D. Felber Dietrich, A. Turk, M. Pons, N. Kunzli, J. M. Gaspoz, T. Rochat, C. Schindler, N. Probst-Hensch and U. Frey (2015). "Long-term smoking cessation and heart rate dynamics in an aging healthy cohort: Is it possible to fully recover?" Environ Res **143**(Pt A): 39-48.

Gold, D. R., A. Litonjua, J. Schwartz, E. Lovett, A. Larson, B. Nearing, G. Allen, M. Verrier, R. Cherry and R. Verrier (2000). "Ambient pollution and heart rate variability." Circulation **101**(11): 1267-1273.

Goldberger, A. L. (1997). "Fractal variability versus pathologic periodicity: complexity loss and stereotypy in disease." Perspect Biol Med **40**(4): 543-561.

Goldberger, A. L., L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng and H. E. Stanley (2000). "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals." Circulation **101**(23): E215-220.

Goldberger, A. L., L. A. Amaral, J. M. Hausdorff, P. Ivanov, C. K. Peng and H. E. Stanley (2002). "Fractal dynamics in physiology: alterations with disease and aging." Proc Natl Acad Sci U S A **99 Suppl 1**: 2466-2472.

Goldberger, A. L., C. K. Peng and L. A. Lipsitz (2002). "What is physiologic complexity and how does it change with aging and disease?" Neurobiol Aging **23**(1): 23-26.

Goldberger, A. L. and B. J. West (1987). "Applications of nonlinear dynamics to clinical cardiology." Ann N Y Acad Sci **504**: 195-213.

Grainge, C. L., L. C. Lau, J. A. Ward, V. Dulay, G. Lahiff, S. Wilson, S. Holgate, D. E. Davies and P. H. Howarth (2011). "Effect of bronchoconstriction on airway remodeling in asthma." N Engl J Med **364**(21): 2006-2015.

Hakala, K., B. Stenius-Aarniala and A. Sovijarvi (2000). "Effects of weight loss on peak flow variability, airways obstruction, and lung volumes in obese patients with asthma." Chest **118**(5): 1315-1321.

Haldar, P., I. D. Pavord, D. E. Shaw, M. A. Berry, M. Thomas, C. E. Brightling, A. J. Wardlaw and R. H. Green (2008). "Cluster analysis and clinical asthma phenotypes." Am J Respir Crit Care Med **178**(3): 218-224.

Han, J. and M. Kamber (2001). Data Mining: Concepts and Techniques. San Francisco.

Harte, C. B. and C. M. Meston (2013). "Effects of Smoking Cessation on Heart Rate Variability Among Long-Term Male Smokers." Int J Behav Med.

Health, O. o. t. S. G. a. O. o. S. a. (2004). The Health Consequences of Smoking: A Report of the Surgeon General. Atlanta (GA).

Heaney, L. G. and D. S. Robinson (2005). "Severe asthma treatment: need for characterising patients." Lancet **365**(9463): 974-976.

Hegger, R., H. Kantz and T. Schreiber (1999). "Practical implementation of nonlinear time series methods: The TISEAN package." Chaos **9**(2): 413-435.

Hinks, T. S., T. Brown, L. C. Lau, H. Rupani, C. Barber, S. Elliott, J. A. Ward, J. Ono, S. Ohta, K. Izuhara, R. Djukanovic, R. J. Kurukulaaratchy, A. Chauhan and P. H. Howarth (2016). "Multidimensional endotyping in patients with severe asthma reveals inflammatory heterogeneity in matrix metalloproteinases and chitinase 3-like protein 1." J Allergy Clin Immunol **138**(1): 61-75.

Holguin, F., M. M. Tellez-Rojo, M. Hernandez, M. Cortez, J. C. Chow, J. G. Watson, D. Mannino and I. Romieu (2003). "Air pollution and heart rate variability among the elderly in Mexico City." Epidemiology **14**(5): 521-527.

Honjo, K., H. Iso, S. Tsugane, A. Tamakoshi, H. Satoh, K. Tajima, T. Suzuki and T. Sobue (2010). "The effects of smoking and smoking cessation on mortality from cardiovascular disease among Japanese: pooled analysis of three large-scale cohort studies in Japan." Tob Control **19**(1): 50-57.

James, A. J., L. E. Reinius, M. Verhoek, A. Gomes, M. Kupczyk, U. Hammar, J. Ono, S. Ohta, K. Izuhara, E. Bel, J. Kere, C. Soderhall, B. Dahlen, R. G. Boot, S. E. Dahlen and B. Consortium (2016). "Increased YKL-40 and Chitotriosidase in Asthma and Chronic Obstructive Pulmonary Disease." Am J Respir Crit Care Med **193**(2): 131-142.

Jones, P. W., F. H. Quirk, C. M. Baveystock and P. Littlejohns (1992). "A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire." Am Rev Respir Dis **145**(6): 1321-1327.

Juniper, E. F., P. M. O'Byrne, G. H. Guyatt, P. J. Ferrie and D. R. King (1999). "Development and validation of a questionnaire to measure asthma control." Eur Respir J **14**(4): 902-907.

Just, J., R. Gouvis-Echraghi, S. Rouve, S. Wanin, D. Moreau and I. Annesi-Maesano (2012). "Two novel, severe asthma phenotypes identified during childhood using a clustering approach." Eur Respir J **40**(1): 55-60.

Kaminsky, D. A., L. L. Wang, J. H. Bates, C. Thamrin, D. M. Shade, A. E. Dixon, R. A. Wise, S. Peters and C. G. Irvin (2016). "Fluctuation Analysis of Peak Expiratory Flow and its Association with Treatment Failure in Asthma." Am J Respir Crit Care Med.

Kantz, H. and T. schreiber (2004). "Nonlinear time series analysis." Cambridge university press.

Kerstjens, H. A., P. L. Brand, M. D. Hughes, N. J. Robinson, D. S. Postma, H. J. Sluiter, E. R. Bleecker, P. N. Dekhuijzen, P. M. de Jong, H. J. Mengelers and et al. (1992). "A comparison of bronchodilator therapy with or without inhaled corticosteroid therapy for obstructive airways disease. Dutch Chronic Non-Specific Lung Disease Study Group." N Engl J Med **327**(20): 1413-1419.

Kleiger, R. E., J. P. Miller, J. T. Bigger, Jr. and A. J. Moss (1987). "Decreased heart rate variability and its association with increased mortality after acute myocardial infarction." Am J Cardiol **59**(4): 256-262.

Konradsen, J. R., A. James, B. Nordlund, L. E. Reinius, C. Soderhall, E. Melen, A. M. Wheelock, K. C. Lodrup Carlsen, M. Lidegran, M. Verhoek, R. G. Boot, B. Dahlen, S. E. Dahlen and G. Hedlin (2013). "The chitinase-like protein YKL-40: a possible biomarker of inflammation and airway remodeling in severe pediatric asthma." J Allergy Clin Immunol **132**(2): 328-335 e325.

Konstantellou, E., A. I. Papaioannou, S. Loukides, G. Patentalakis, A. Papaporfyriou, G. Hillas, S. Papiris, N. Koulouris, P. Bakakos and K. Kostikas (2015). "Persistent airflow obstruction in patients with asthma: Characteristics of a distinct clinical phenotype." Respir Med **109**(11): 1404-1409.

Kupczyk, M., B. Dahlen, P. J. Sterk, E. Nizankowska-Mogilnicka, A. Papi, E. H. Bel, P. Chanez, P. H. Howarth, S. T. Holgate, G. Brusselle, N. M. Siafakas, M. Gjomarkaj, S. E. Dahlen and B. investigators (2014). "Stability of phenotypes defined by physiological variables and biomarkers in adults with asthma." Allergy **69**(9): 1198-1204.

Kupczyk, M., S. Haque, R. J. Middelveld, B. Dahlen, S. E. Dahlen and BIOAIR investigators (2013). "Phenotypic predictors of response to oral glucocorticosteroids in severe asthma." Respir Med **107**(10): 1521-1530.

Kupczyk, M., S. Haque, P. J. Sterk, E. Nizankowska-Mogilnicka, A. Papi, E. H. Bel, P. Chanez, B. Dahlen, M. Gaga, M. Gjomarkaj, P. H. Howarth, S. L. Johnston, G. F. Joos, F. Kanniess, E. Tzortzaki, A. James, R. J. Middelveld, S. E. Dahlen and BIOAIR investigators (2013). "Detection of exacerbations in asthma based on electronic diary data: results from the 1-year prospective BIOAIR study." Thorax **68**(7): 611-618.

Lewis, M. J., G. Balaji, H. Dixon, Y. Syed and K. E. Lewis (2010). "Influence of smoking abstinence and nicotine replacement therapy on heart rate and QT time-series." Clin Physiol Funct Imaging **30**(1): 43-50.

Lightwood, J. M. and S. A. Glantz (1997). "Short-term economic and health benefits of smoking cessation: myocardial infarction and stroke." Circulation **96**(4): 1089-1096.

Lipsitz, L. A. and A. L. Goldberger (1992). "Loss of 'complexity' and aging. Potential applications of fractals and chaos theory to senescence." JAMA **267**(13): 1806-1809.

Liu, L. J., I. Curjuric, D. Keidel, J. Heldstab, N. Kunzli, L. Bayer-Oglesby, U. Ackermann-Liebrich, C. Schindler and S. team (2007). "Characterization of source-specific air pollution exposure for a large population-based Swiss cohort (SAPALDIA)." Environ Health Perspect **115**(11): 1638-1645.

Mannan, H., C. Stevenson, A. Peeters, H. Walls and J. McNeil (2010). "Framingham risk prediction equations for incidence of cardiovascular disease using detailed measures for smoking." Heart Int **5**(2): e11.

Manor, B. and L. A. Lipsitz (2013). "Physiologic complexity and aging: implications for physical function and rehabilitation." Prog Neuropsychopharmacol Biol Psychiatry **45**: 287-293.

Martin, B. W., U. Ackermann-Liebrich, P. Leuenberger, N. Kunzli, E. Z. Stutz, R. Keller, J. P. Zellweger, B. Wuthrich, C. Monn, K. Blaser, G. Bolognini, J. P. Bongard, O. Brandli, P. Braun, C. Defila, G. Domenighetti, L. Grize, W. Karrer, H. Keller-Wossidlo, T. C. Medici, A. Peeters, A. P. Perruchoud, C. Schindler, M. H. Schoeni, B. Villiger and et al. (1997). "SAPALDIA: methods and participation in the cross-sectional part of the Swiss Study on Air Pollution and Lung Diseases in Adults." Soz Praventivmed **42**(2): 67-84.

Meyer, M. and O. Stiedl (2003). "Self-affine fractal variability of human heartbeat interval dynamics in health and disease." Eur J Appl Physiol **90**(3-4): 305-316.

Minami, J., T. Ishimitsu and H. Matsuoka (1999). "Effects of smoking cessation on blood pressure and heart rate variability in habitual smokers." Hypertension **33**(1 Pt 2): 586-590.

Moeckel, R. and B. Murray (1997). "Measuring the distance between time series." Physica D **102**: 187-194.

Moore, W. C. (2013). "The natural history of asthma phenotypes identified by cluster analysis. Looking for chutes and ladders." Am J Respir Crit Care Med **188**(5): 521-522.

Moore, W. C., D. A. Meyers, S. E. Wenzel, W. G. Teague, H. Li, X. Li, R. D'Agostino, Jr., M. Castro, D. Curran-Everett, A. M. Fitzpatrick, B. Gaston, N. N. Jarjour, R. Sorkness, W. J. Calhoun, K. F. Chung, S. A. Comhair, R. A. Dweik, E. Israel, S. P. Peters, W. W. Busse, S. C. Erzurum, E. R. Bleecker, L. National Heart and P. Blood Institute's Severe Asthma Research (2010). "Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program." Am J Respir Crit Care Med **181**(4): 315-323.

Moore, W. C. and S. P. Peters (2006). "Severe asthma: an overview." J Allergy Clin Immunol **117**(3): 487-494; quiz 495.

Mordukhovich, I., B. Coull, I. Kloog, P. Koutrakis, P. Vokonas and J. Schwartz (2015). "Exposure to sub-chronic and long-term particulate air pollution and heart rate variability in an elderly cohort: the Normative Aging Study." Environ Health **14**: 87.

Mosen, D. M., M. Schatz, D. J. Magid and C. A. Camargo, Jr. (2008). "The relationship between obesity and asthma severity and control in adults." J Allergy Clin Immunol **122**(3): 507-511 e506.

Munjal, S., T. Koval, R. Muhammad, Y. Jin, V. Demmel, H. J. Roethig, P. Mendes and M. Unverdorben (2009). "Heart rate variability increases with reductions in cigarette smoke exposure after 3 days." J Cardiovasc Pharmacol Ther **14**(3): 192-198.

Muskulus, M. and S. Verduyn-Lunel (2011). "Wasserstein distances in the analysis of time series and dynamical systems." Physica D **240**: 45-58.

Negri, E., C. La Vecchia, B. D'Avanzo, A. Nobili and R. G. La Malfa (1994). "Acute myocardial infarction: association with time since stopping smoking in Italy. GISSI-EFRIM Investigators. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto. Epidemiologia dei Fattori di Rischio dell'Infarto Miocardico." J Epidemiol Community Health **48**(2): 129-133.

Novello, A. C. (1990). "Surgeon General's report on the health benefits of smoking cessation." Public Health Rep **105**(6): 545-548.

Ockene, J. K., L. H. Kuller, K. H. Svendsen and E. Meilahn (1990). "The relationship of smoking cessation to coronary heart disease and lung cancer in the Multiple Risk Factor Intervention Trial (MRFIT)." Am J Public Health **80**(8): 954-958.

Paggiaro, P. L., P. Chanez, O. Holz, P. W. Ind, R. Djukanovic, P. Maestrelli and P. J. Sterk (2002). "Sputum induction." Eur Respir J Suppl **37**: 3s-8s.

Park, S. K., M. S. O'Neill, P. S. Vokonas, D. Sparrow and J. Schwartz (2005). "Effects of air pollution on heart rate variability: the VA normative aging study." Environ Health Perspect **113**(3): 304-309.

Park, S. K., M. S. O'Neill, R. O. Wright, H. Hu, P. S. Vokonas, D. Sparrow, H. Suh and J. Schwartz (2006). "HFE genotype, particulate air pollution, and heart rate variability: a gene-environment interaction." Circulation **114**(25): 2798-2805.

Pascual, R. M. and S. P. Peters (2009). "The irreversible component of persistent asthma." J Allergy Clin Immunol **124**(5): 883-890; quiz 891-882.

Patrick, D. L., A. Cheadle, D. C. Thompson, P. Diehr, T. Koepsell and S. Kinne (1994). "The validity of self-reported smoking: a review and meta-analysis." Am J Public Health **84**(7): 1086-1093.

Peng, C. K., S. Havlin, J. M. Hausdorff, J. E. Mietus, H. E. Stanley and A. L. Goldberger (1995). "Fractal mechanisms and heart rate dynamics. Long-range correlations and their breakdown with disease." J Electrocardiol **28 Suppl**: 59-65.

Peters-Golden, M., A. Swern, S. S. Bird, C. M. Hustad, E. Grant and J. M. Edelman (2006). "Influence of body mass index on the response to asthma controller agents." Eur Respir J **27**(3): 495-503.

Phipatanakul, W., D. T. Mauger, R. L. Sorkness, J. M. Gaffin, F. Holguin, P. G. Woodruff, N. P. Ly, L. B. Bacharier, N. R. Bhakta, W. C. Moore, E. R. Bleecker, A. T. Hastie, D. A. Meyers, M. Castro, J. Fahy, A. Fitzpatrick, B. M. Gaston, N. N. Jarjour, B. D. Levy, S. P. Peters, W. G. Teague, M. Fajt, S. E. Wenzel, S. C. Erzurum, E. Israel and P. and the Severe Asthma Research (2016). "Effects of Age and Disease Severity on Systemic Corticosteroid Responses in Asthma." Am J Respir Crit Care Med.

Pieters, N., M. Plusquin, B. Cox, M. Kicinski, J. Vangronsveld and T. S. Nawrot (2012). "An epidemiological appraisal of the association between heart rate variability and particulate air pollution: a meta-analysis." Heart **98**(15): 1127-1135.

Pikkujamsa, S. M., T. H. Makikallio, K. E. Airaksinen and H. V. Huikuri (2001). "Determinants and interindividual variation of R-R interval dynamics in healthy middle-aged subjects." Am J Physiol Heart Circ Physiol **280**(3): H1400-1406.

Pincus, S. M. (1991). "Approximate entropy as a measure of system complexity." Proc Natl Acad Sci U S A **88**(6): 2297-2301.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar and R Core Team (2014). "nlme: Linear and Nonlinear Mixed Effects Models."

Pope, C. A., 3rd, R. T. Burnett, G. D. Thurston, M. J. Thun, E. E. Calle, D. Krewski and J. J. Godleski (2004). "Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease." Circulation **109**(1): 71-77.

Probst-Hensch, N. M., M. Imboden, D. Felber Dietrich, J. C. Barthelemy, U. Ackermann-Liebrich, W. Berger, J. M. Gaspoz and J. Schwartz (2008). "Glutathione S-transferase polymorphisms, passive smoking, obesity, and heart rate variability in nonsmokers." Environ Health Perspect **116**(11): 1494-1499.

Prosperi, M. C., U. M. Sahiner, D. Belgrave, C. Sackesen, I. E. Buchan, A. Simpson, T. S. Yavuz, O. Kalayci and A. Custovic (2013). "Challenges in identifying asthma subgroups using unsupervised statistical learning techniques." Am J Respir Crit Care Med **188**(11): 1303-1312.

Quanjer, P. H., S. Stanojevic, T. J. Cole, X. Baur, G. L. Hall, B. H. Culver, P. L. Enright, J. L. Hankinson, M. S. Ip, J. Zheng, J. Stocks and E. R. S. G. L. F. Initiative (2012). "Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations." Eur Respir J **40**(6): 1324-1343.

Quanjer, P. H., G. J. Tammeling, J. E. Cotes, O. F. Pedersen, R. Peslin and J. C. Yernault (1993). "Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society." Eur Respir J Suppl **16**: 5-40.

Rajendra Acharya, U., K. Paul Joseph, N. Kannathal, C. M. Lim and J. S. Suri (2006). "Heart rate variability: a review." Med Biol Eng Comput **44**(12): 1031-1051.

Raupach, T., K. Schafer, S. Konstantinides and S. Andreas (2006). "Secondhand smoke as an acute threat for the cardiovascular system: a change in paradigm." Eur Heart J **27**(4): 386-392.

Richman, J. S. and J. R. Moorman (2000). "Physiological time-series analysis using approximate entropy and sample entropy." Am J Physiol Heart Circ Physiol **278**(6): H2039-2049.

Rosenberg, L., J. R. Palmer and S. Shapiro (1990). "Decline in the risk of myocardial infarction among women who stop smoking." N Engl J Med **322**(4): 213-217.

Rosenstein, M. T., J. J. Collins and C. J. De Luca (1993). "A practical method for calculating largest Lyapunov exponents from small data sets." Physica D **65**: 117-134.

Saint-Pierre, P., A. Bourdin, P. Chanez, J. P. Daures and P. Godard (2006). "Are overweight asthmatics more difficult to control?" Allergy **61**(1): 79-84.

Sarkar, D. (2008). "Lattice: Multivariate Data Visualization with R."

Schatz, M., J. W. Hsu, R. S. Zeiger, W. Chen, A. Dorenbaum, B. E. Chipps and T. Haselkorn (2013). "Phenotypes determined by cluster analysis in severe or difficult-to-treat asthma." J Allergy Clin Immunol.

Schumacher, A. (2004). "Linear and nonlinear approaches to the analysis of R-R interval variability." Biol Res Nurs **5**(3): 211-221.

Schwartz, J., A. Litonjua, H. Suh, M. Verrier, A. Zanobetti, M. Syring, B. Nearing, R. Verrier, P. Stone, G. MacCallum, F. E. Speizer and D. R. Gold (2005). "Traffic related pollution and heart rate variability in a panel of elderly subjects." Thorax **60**(6): 455-461.

Schwartz, J., S. K. Park, M. S. O'Neill, P. S. Vokonas, D. Sparrow, S. Weiss and K. Kelsey (2005). "Glutathione-S-transferase M1, obesity, statins, and autonomic effects of particles: gene-by-drug-by-environment interaction." Am J Respir Crit Care Med **172**(12): 1529-1533.

Seaton, A., A. Soutar, V. Crawford, R. Elton, S. McNerlan, J. Cherrie, M. Watt, R. Agius and R. Stout (1999). "Particulate air pollution and the blood." Thorax **54**(11): 1027-1032.

Shaper, A. G., S. J. Pocock, M. Walker, A. N. Phillips, T. P. Whitehead and P. W. Macfarlane (1985). "Risk factors for ischaemic heart disease: the prospective phase of the British Regional Heart Study." J Epidemiol Community Health **39**(3): 197-209.

Shields, M., R. E. Garner and K. Wilkins (2013). "Dynamics of smoking cessation and health-related quality of life among Canadians." Health Rep **24**(2): 3-11.

Shields, M. and K. Wilkins (2013). "Smoking, smoking cessation and heart disease risk: A 16-year follow-up study." Health Rep **24**(2): 12-22.

Shore, S. A. (2008). "Obesity and asthma: possible mechanisms." J Allergy Clin Immunol **121**(5): 1087-1093; quiz 1094-1085.

Siroux, V., X. Basagana, A. Boudier, I. Pin, J. Garcia-Aymerich, A. Vesin, R. Slama, D. Jarvis, J. M. Anto, F. Kauffmann and J. Sunyer (2011). "Identifying adult asthma phenotypes using a clustering approach." Eur Respir J **38**(2): 310-317.

Siroux, V. and J. Garcia-Aymerich (2011). "The investigation of asthma phenotypes." Curr Opin Allergy Clin Immunol **11**(5): 393-399.

Smith, J. A., R. Drake, A. Simpson, A. Woodcock, A. Pickles and A. Custovic (2008). "Dimensions of respiratory symptoms in preschool children: population-based birth cohort study." Am J Respir Crit Care Med **177**(12): 1358-1363.

Sodlerlund, A., A. Fischer and T. Johansson (2009). "Physical activity, diet and behaviour modification in the treatment of overweight and obese adults: a systematic review." Perspect Public Health **129**(3): 132-142.

Sorkness, R. L., E. R. Bleecker, W. W. Busse, W. J. Calhoun, M. Castro, K. F. Chung, D. Curran-Everett, S. C. Erzurum, B. M. Gaston, E. Israel, N. N. Jarjour, W. C. Moore, S. P. Peters, W. G. Teague, S. E. Wenzel, L. National Heart and P. Blood Institute Severe Asthma Research (2008). "Lung function in adults with stable but severe asthma: air trapping and incomplete reversal of obstruction with bronchodilation." J Appl Physiol (1985) **104**(2): 394-403.

Stein, P. K., J. N. Rottman and R. E. Kleiger (1996). "Effect of 21 mg transdermal nicotine patches and smoking cessation on heart rate variability." Am J Cardiol **77**(9): 701-705.

Stenius-Aarniala, B., T. Poussa, J. Kvarnstrom, E. L. Gronlund, M. Ylikahri and P. Mustajoki (2000). "Immediate and long term effects of weight reduction in obese people with asthma: randomised controlled study." BMJ **320**(7238): 827-832.

Stern, G., J. de Jongste, R. van der Valk, E. Baraldi, S. Carraro, C. Thamrin and U. Frey (2011). "Fluctuation phenotyping based on daily fraction of exhaled nitric oxide values in asthmatic children." J Allergy Clin Immunol **128**(2): 293-300.

Stevens, K. R. and L. R. Munoz (2004). "Cigarette smoking: Evidence to guide measurement." Res Nurs Health **27**(4): 281-292.

Sukkar, M. B., M. A. Ullah, W. J. Gan, P. A. Wark, K. F. Chung, J. M. Hughes, C. L. Armour and S. Phipps (2012). "RAGE: a new frontier in chronic airways disease." Br J Pharmacol **167**(6): 1161-1176.

Sukkar, M. B., L. G. Wood, M. Tooze, J. L. Simpson, V. M. McDonald, P. G. Gibson and P. A. Wark (2012). "Soluble RAGE is deficient in neutrophilic asthma and COPD." Eur Respir J **39**(3): 721-729.

Takens, F. (1981). Detecting strange attractors in turbulence, Springer-Verlag.

Taylor, B., D. Mannino, C. Brown, D. Crocker, N. Twum-Baah and F. Holguin (2008). "Body mass index and asthma severity in the National Asthma Survey." Thorax **63**(1): 14-20.

Teo, K. K., S. Ounpuu, S. Hawken, M. R. Pandey, V. Valentin, D. Hunt, R. Diaz, W. Rashed, R. Freeman, L. Jiang, X. Zhang and S. Yusuf (2006). "Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study." Lancet **368**(9536): 647-658.

Thamrin, C., U. Frey, D. A. Kaminsky, H. K. Reddel, A. J. Seely, B. Suki and P. J. Sterk (2016). "Systems Biology and Clinical Practice in Respiratory Medicine. The Twain Shall Meet." Am J Respir Crit Care Med **194**(9): 1053-1061.

Thamrin, C., R. Nydegger, G. Stern, P. Chanez, S. E. Wenzel, R. A. Watt, S. FitzPatrick, D. R. Taylor and U. Frey (2011). "Associations between fluctuations in lung function and asthma control in two populations with differing asthma severity." Thorax **66**(12): 1036-1042.

Thamrin, C., J. Zindel, R. Nydegger, H. K. Reddel, P. Chanez, S. E. Wenzel, S. FitzPatrick, R. A. Watt, B. Suki and U. Frey (2011). "Predicting future risk of asthma exacerbations using individual conditional probabilities." J Allergy Clin Immunol **127**(6): 1494-1502 e1493.

Tschumperlin, D. J. and J. M. Drazen (2006). "Chronic effects of mechanical force on airways." Annu Rev Physiol **68**: 563-583.

Tsuji, H., M. G. Larson, F. J. Venditti, Jr., E. S. Manders, J. C. Evans, C. L. Feldman and D. Levy (1996). "Impact of reduced heart rate variability on risk for cardiac events. The Framingham Heart Study." Circulation **94**(11): 2850-2855.

Tverdal, A., D. Thelle, I. Stensvold, P. Leren and K. Bjartveit (1993). "Mortality in relation to smoking history: 13 years' follow-up of 68,000 Norwegian men and women 35-49 years." J Clin Epidemiol **46**(5): 475-487.

Umetsu, D. T. (2016). "Mechanisms by which obesity impacts asthma." Thorax.

Vandeput, S., B. Verheyden, A. E. Aubert and S. Van Huffel (2012). "Nonlinear heart rate dynamics: circadian profile and influence of age and gender." Med Eng Phys **34**(1): 108-117.

Voss, A., S. Schulz, R. Schroeder, M. Baumert and P. Caminal (2009). "Methods derived from nonlinear dynamics for analysing heart rate variability." Philos Trans A Math Phys Eng Sci **367**(1887): 277-296.

Wannamethee, S. G., A. G. Shaper, P. H. Whincup and M. Walker (1995). "Smoking cessation and the risk of stroke in middle-aged men." JAMA **274**(2): 155-160.

Warren-Liao, T. (2005). "Clustering of time series data - a survey." Pattern Recognition **38**: 1857-1874.

Weatherall, M., J. Travers, P. M. Shirtcliffe, S. E. Marsh, M. V. Williams, M. R. Nowitz, S. Aldington and R. Beasley (2009). "Distinct clinical phenotypes of airways disease defined by cluster analysis." Eur Respir J **34**(4): 812-818.

Wenzel, S. E. (2012). "Asthma phenotypes: the evolution from clinical to molecular approaches." Nat Med **18**(5): 716-725.

Wickham, H. (2009). "ggplot2: elegant graphics for data analysis."

Woo, M. A., W. G. Stevenson, D. K. Moser, R. B. Trelease and R. M. Harper (1992). "Patterns of beat-to-beat heart rate variability in advanced heart failure." Am Heart J **123**(3): 704-710.

Wu, W., E. Bleecker, W. Moore, W. W. Busse, M. Castro, K. F. Chung, W. J. Calhoun, S. Erzurum, B. Gaston, E. Israel, D. Curran-Everett and S. E. Wenzel (2014). "Unsupervised phenotyping of Severe Asthma Research Program participants using expanded lung data." J Allergy Clin Immunol **133**(5): 1280-1288.

Yotsukura, M., Y. Koide, K. Fujii, Y. Tomono, A. Katayama, H. Ando, J. Suzuki and K. Ishikawa (1998). "Heart rate variability during the first month of smoking cessation." Am Heart J **135**(6 Pt 1): 1004-1009.

Zeki, A. A., M. Schivo, A. Chan, T. E. Albertson and S. Louie (2011). "The Asthma-COPD Overlap Syndrome: A Common Clinical Problem in the Elderly." J Allergy (Cairo) **2011**: 861926.

Zerah, F., A. Harf, L. Perlemuter, H. Lorino, A. M. Lorino and G. Atlan (1993). "Effects of obesity on respiratory resistance." Chest **103**(5): 1470-1476.

8. Appendix

8.1. Appendix 1: Handling missing data

The FBC approach was conceived with the aim of a clinical application. Consequently, special attention was given to the fact that this approach should be adapted for times series containing missing values. This constraint motivated the use of the Earth mover's distance to quantify the similarity between individuals. Thus, comparison of the signals is not based on correlation properties of signals, but on the probability distribution of the data points. Consequently, it neglects the time dimension, but it is quite robust with respect to missing data. Therefore, with the FBC approach, no extensive handling of missing data is required.

In the context of lung function measured twice-daily by subjects using a lung function meter, time series are typically characterized by short continuous segments of missing values (gaps ≤ 3 data points), which correspond to patients intermittent/punctual omissions. In this case, the local mean imputation allows for a substantial increase of the signal continuity, in a simple manner, without distorting the distribution of the variable. Method for local mean imputation is given in **Table A8.1**.

Table A8.1. Method for local mean imputation

Gap length	Representation of the gap	Steps for local mean imputation
1 missing value	X_1, X_2, X_3, M_1, X_4, X_5, X_6	M_1 = mean(X_3,X_4)
2 missing values	X_1, X_2, X_3, M_1, M_2, X_4, X_5, X_6	
	X_1, X_2, X_3, M, X_4, X_5, X_6	M = mean(X_2,X_3,X_4,X_5)
	X_1, X_2, X_3, M_1, M, M_2, X_4, X_5, X_6	M_1 = mean(X_3,M)
		M_2 = mean(X_4,M)
3 missing values	X_1, X_2, X_3, M_1, M_2, M_3, X_4, X_5, X_6	M_2 = mean(X_2,X_3,X_4,X_5)
		M_1 = mean(X_3,M_2)
		M_3 = mean(X_4,M_2)

M, Missing value; X, measurement

In order to examine the missing values contained in the time series of z-score FEV_1 used for the FBC analysis, and to evaluate whether we should perform the local mean imputation, we defined three fragmentation indexes.

Index 1: Distribution of length of uninterrupted segments of missing values

Median segments length was 1 [min: 1, 25thquartile: 1, 75thquartile: 2, maximum: 73] (**Figure A8.1**). Therefore, the uninterrupted segments of missing values were mostly smaller than 3 missing values.

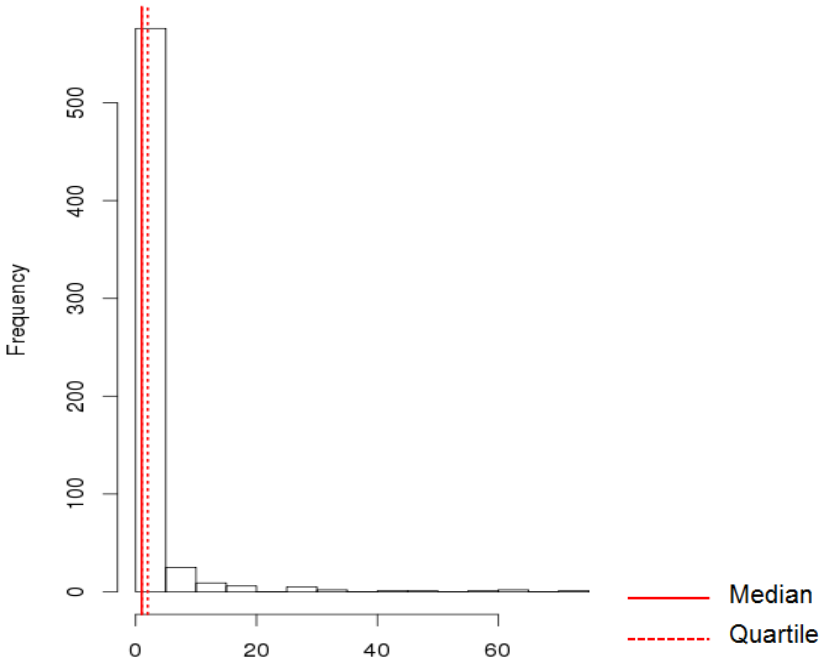

Figure A8.1. Distribution of length of uninterrupted segments of missing values

Index 2: Distribution of relative length of uninterrupted segments of missing values

$$\text{Relative gap lenght} = \frac{\text{gap length}}{\text{total number of data points in the time series (measurements + missing values)}}$$

A gap corresponds to an uninterrupted segment of missing values.

Median segments relative length was 2% [min: 0.4%, 25thquartile: 1%, 75thquartile: 4%, maximum: 100%] (**Figure A8.2)**. Therefore, the uninterrupted segments of missing values were mostly very short.

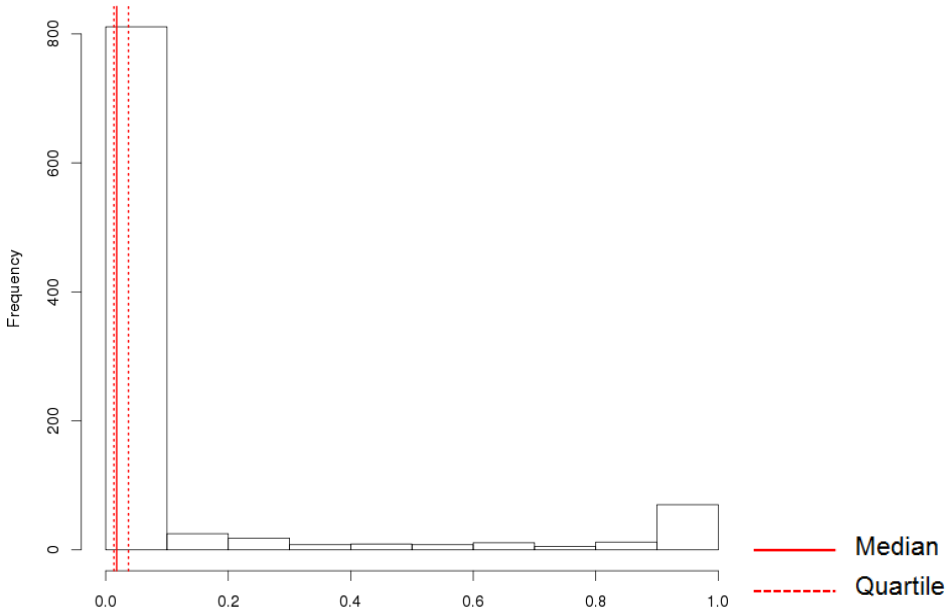

Figure A8.2. Distribution of relative length of uninterrupted segments with missing values

Index 3: Index of fragmentation (for X ≤ 3)

$$\text{Index of fragmentation} = \frac{\text{number of gaps whose length} \leq 3}{\text{total number of gaps in the time series}}$$

A gap corresponds to an uninterrupted segment with missing values.

For X≤ 3, median index of fragmentation was 100% [min: 0%, 25thquartile: 77%, 75thquartile: 100%, maximum: 100%] (**Figure A8.3)**. Therefore, most of the uninterrupted segments of missing values counted ≤ 3 missing values.

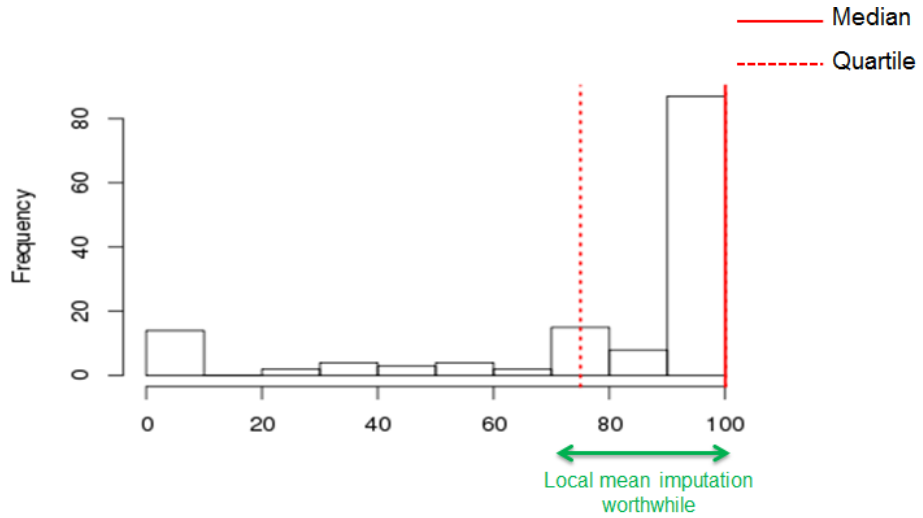

Figure A8.3. Distribution of index of fragmentation, for X ≤ 3

According to the three indexes, the local mean imputation should substantially improve the continuity of the time series.

Finally, in order to evaluate how the continuity of the time series improved using the local mean imputation, we calculated the three fragmentation indexes for the uninterrupted segments of measurement in the time series before and after imputation.

Index 1: Distribution of length of uninterrupted segments of measurements

Table A8.2. Distribution of length of uninterrupted segments of measurements before and after imputation

	Minimum	First quartile	Median	Third quartile	Maximum
Before imputation	1	1	2	6	58
After imputation	1	1	8	27	91

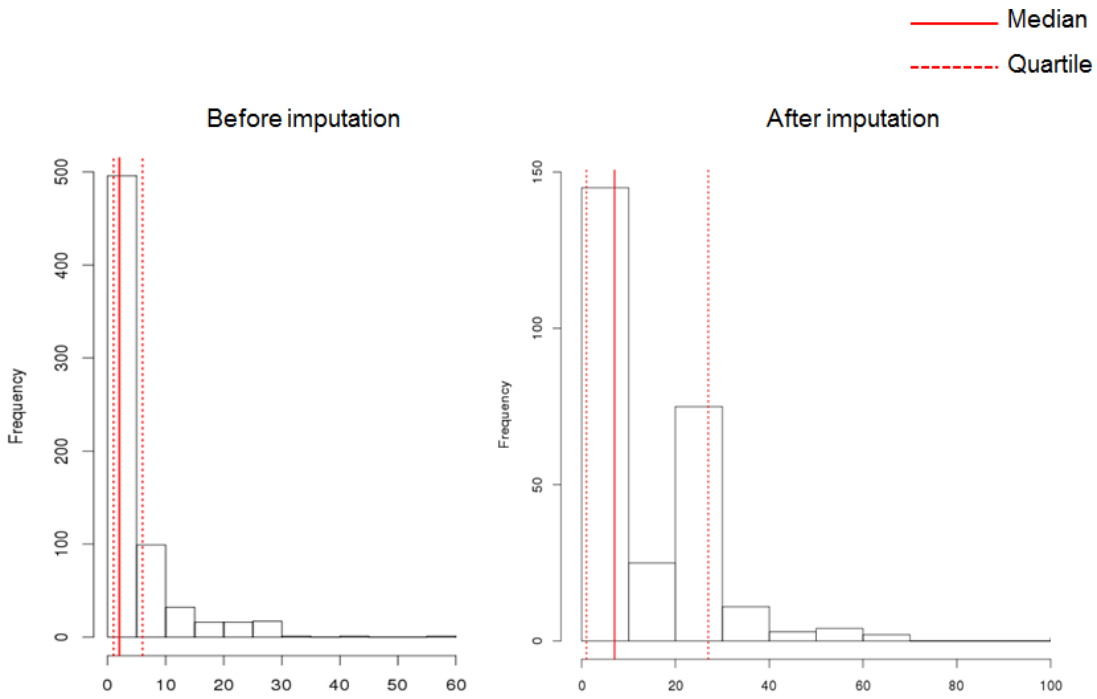

Figure A8.4. Distribution of length of uninterrupted segments of measurements before and after local mean imputation

Index 2: Distribution of relative length of uninterrupted segments of measurements

Table A8.3. Distribution of relative length of uninterrupted segments of measurements before and after imputation

	Minimum	First quartile	Median	Third quartile	Maximum
Before imputation	0.7%	3%	7%	20%	100%
After imputation	1%	4%	30%	100%	100%

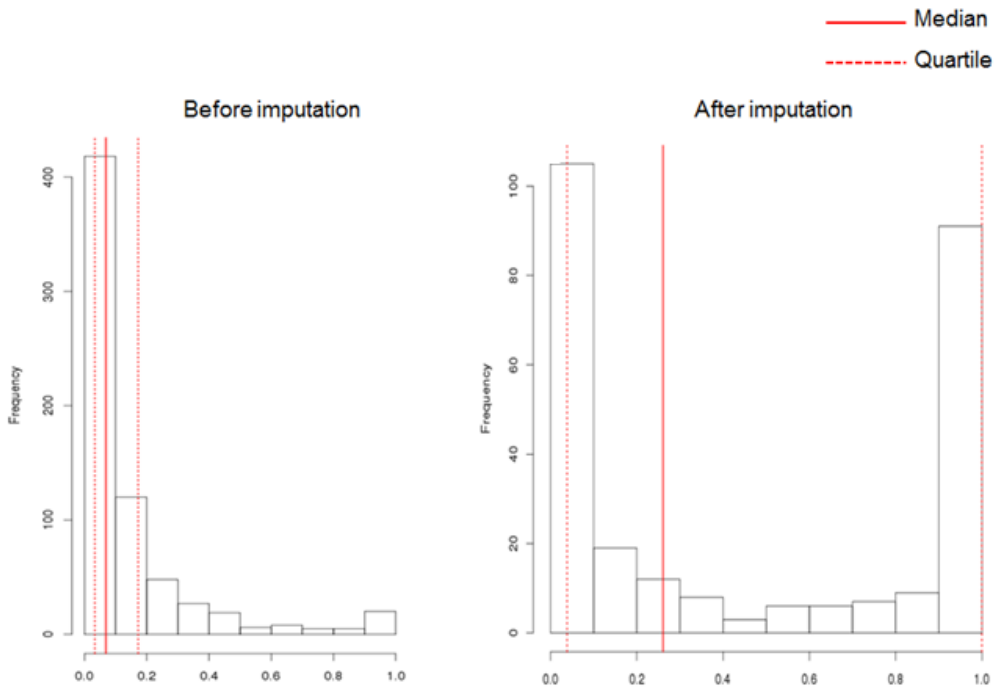

Figure A8.5. Distribution of relative length of uninterrupted segments of measurements before and after local mean imputation

Index 3: Index of fragmentation (X ≥ 10)

Table A8.4. Distribution of index of fragmentation, for X ≥ 10, before and after local mean imputation

	Minimum	First quartile	Median	Third quartile	Maximum
Before imputation	0%	15%	48%	100%	100%
After imputation	0%	50%	100%	100%	100%

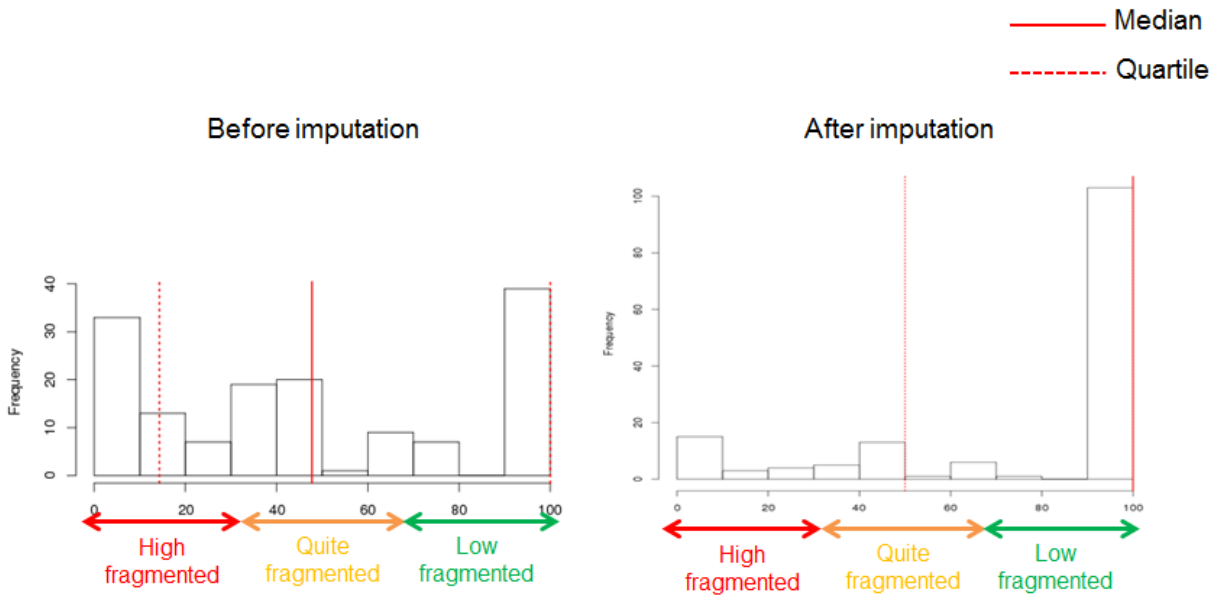

Figure A8.6. Distribution of index of fragmentation, for X ≥ 10, before and after local mean imputation

According to the three indexes, the local mean imputation substantially improved the continuity of the time series.

8.2. Appendix 2: Time delay embedding

Given a time delay t=2, and an embedding dimension d=3, the embedding procedure consists in starting at the very first value X_1 of the time series and grouping three (d=3) consecutive values that are separated by t into a vector (here, given that t=2 and d=3, that would be the vector (X_1,X_3,X_5)), then moving to the next entry in the time series, that is X_2, and repeating the grouping procedure resulting in the vector (X_2,X_4,X_6). The embedded time series is the series of vectors (X_1,X_3,X_5), (X_2,X_4,X_6), etc.

8.3. Appendix 3: Additional manuscript

Another stimulating aspect of my PhD work was using the knowledge gained as a part of my studies to contribute to other studies from our interdisciplinary research group. The manuscript of a successful project is provided in this appendix.

Physiological phenotyping of pediatric chronic obstructive airway diseases

S. Nyilas, F. Singer, N. Kumar, S. Yammine, D. Meier-Girard, C. Koerner-Rettberg, C. Casaulta, U. Frey, P. Latzin.

Journal of Applied Physiology 2016 Jul 1;121(1):324-32

J Appl Physiol 121: 324–332, 2016.
First published May 26, 2016; doi:10.1152/japplphysiol.00086.2016.

Physiological phenotyping of pediatric chronic obstructive airway diseases

Sylvia Nyilas,[1,2] Florian Singer,[1,3] Nitin Kumar,[2] Sophie Yammine,[1,2] Delphine Meier-Girard,[2] Cordula Koerner-Rettberg,[4] Carmen Casaulta,[1] Urs Frey,[2] and Philipp Latzin[1,2]

[1]*Division of Respiratory Medicine, Department of Paediatrics, University Children's Hospital of Bern, University of Bern, Bern, Switzerland;* [2]*Department of Paediatric Pulmonology, University Children's Hospital Basel (UKBB), Basel, Switzerland;* [3]*Division of Respiratory Medicine, University Children's Hospital Zurich, Zurich, Switzerland; and* [4]*Department of Paediatric Pulmonology, University Children's Hospital of Ruhr University Bochum at St. Josef-Hospital, Bochum, Germany*

Submitted 27 January 2016; accepted in final form 24 May 2016

Nyilas S, Singer F, Kumar N, Yammine S, Meier-Girard D, Koerner-Rettberg C, Casaulta C, Frey U, Latzin P. Physiological phenotyping of pediatric chronic obstructive airway diseases. *J Appl Physiol* 121: 324–332, 2016. First published May 26, 2016; doi:10.1152/japplphysiol.00086.2016.—Inert tracer gas washout (IGW) measurements detect increased ventilation inhomogeneity (VI) in chronic lung diseases. Their suitability for different diseases, such as cystic fibrosis (CF) and primary ciliary dyskinesia (PCD), has already been shown. However, it is still unclear if physiological phenotypes based on different IGW variables can be defined independently of underlying disease. Eighty school-age children, 20 with CF, 20 with PCD, 20 former preterm children, and 20 healthy children, performed nitrogen multiple-breath washout, double-tracer gas (DTG) single-breath washout, and spirometry. Our primary outcome was the definition of physiological phenotypes based on IGW variables. We applied principal component analysis, hierarchical Ward's clustering, and enrichment analysis to compare clinical characteristics between the clusters. IGW variables used for clustering were lung clearance index (LCI) and convection-dependent [conductive ventilation heterogeneity index (Scond)] and diffusion-convection-dependent variables [acinar ventilation heterogeneity index (Sacin) and carbon dioxide and DTG phase III slopes]. Three main phenotypes were identified. Phenotype I (n = 38) showed normal values in all IGW outcome variables. Phenotype II (n = 21) was characterized by pronounced global and convection-dependent VI while diffusion-dependent VI was normal. Phenotype III (n = 21) was characterized by increased global and diffusion- and convection-dependent VI. Enrichment analysis revealed an overrepresentation of healthy children and former preterm children in phenotype I and of CF and PCD in phenotypes II and III. Patients in phenotype III showed the highest proportion and frequency of exacerbations and hospitalization in the year prior to the measurement. IGW techniques allow identification of clinically meaningful, disease-independent physiological clusters. Their predictive value of future disease outcomes remains to be determined.

gas washout; spirometry; phenotypes; clustering; lung disease

NEW & NOTEWORTHY

Clustering signals from different single- and multiple-breath gas washout tests in children with various lung diseases (e.g., cystic fibrosis and primary ciliary dyskinesia) results in the identification of three different physiological phenotypes. This novel application of the hierarchical Ward's clustering method allows the characterization of lung disease independent of the underlying disease entity and thus seems a promising tool for personalized medicine.

Address for reprint requests and other correspondence: S. Nyilas, Children's University Hospital Bern, Freiburgstrasse, 3010 Bern, Switzerland (e-mail: sylvia.nyilas@ukbb.ch).

THE FRACTAL ARCHITECTURE OF small airways enables homogeneous ventilation with gas transport into the gas-exchanging acinar compartments by convection and diffusion (44, 45). Depending on the generation and distribution of affected airways, respective function is impaired to a varying degree and often subclinical in many chronic lung diseases. Chronic lung diseases such as cystic fibrosis (CF) and primary ciliary dyskinesia (PCD) significantly differ in etiology but may share similar physiological characteristics. In both diseases, inhomogeneous patchy distribution of airway obstruction is associated with inhomogeneous ventilation distribution (16, 25). During early stages of lung diseases, small airways in particular are most affected. This is best captured by inert gas washout (IGW) tests while ventilation capacity measured by spirometry is usually less affected (8, 17, 27). During the disease course, alterations in airway structure occur and ultimately lead to a changing picture of functional limitations. Despite possible similarities between diseases and changes in physiological properties over the disease course, treatment usually depends on underlying disease entity.

Especially with regard to precision and individualized medicine a better characterization of the underlying physiology seems worthwhile to establish; this would enable tailored diagnosis and therapy. One possibility to do so would be to cluster the compound information from different outcome variables from various IGW tests. Statistical methods such as hierarchical Ward's clustering can then be used to combine different physiological variables in an unsupervised way (12, 35).

In our study we hypothesized that clustering may identify physiological phenotypes that do not necessarily relate to different disease etiologies but rather to common physiological information and that resulting clusters show differences in enrichment of other characteristics.

METHODS

Study Design

In this proof-of-principle study we used IGW data from 80 children with different lung diseases. We first performed conventional (biased) analyses comparing IGW outcome variables between disease groups. Second, we performed principal component analysis (PCA) to determine whether or not all IGW outcome variables are necessary for clustering. In a third step, we performed Ward's hierarchical clustering using the outcome variables from IGW. Primary outcome was the definition of physiological phenotypes based on IGW variables calculated by hierarchical Ward's clustering. Secondary outcome was

differences between clusters assessed by enrichment analysis. For this analysis we compared disease etiology, demographics, IGW outcomes, and disease course previous to the measurement between the groups (7, 28).

Study population. We included 80 children from 3 disease groups (CF, PCD, and former preterm children) as well as healthy children with 20 children from each group. Children were 7-18 yr old and consecutively recruited at the outpatient clinics at the University Children's Hospital Bern, Switzerland, and Children's Hospital Bochum, Germany, between January 2012 and May 2015 independent of disease activity or medication. PCD patients were enrolled at University Children's Hospital Bochum ($n = 17$) and Bern ($n = 3$); all other children were enrolled at the University Children's Hospital Bern. Exclusion criteria for all children were respiratory infection within the last 3 wk, acute pulmonary exacerbation at the time of measurements (marked increase in cough, fever, or malaise), or history of lung disease in healthy children. All children underwent two different established gas washout techniques, single- and multiple-breath washout tests (SBW or MBW). All except healthy children underwent subsequent spirometry. Measurements were performed on the same day.

Ethics Statement

The study was approved by the Ethics Committee of the Canton of Bern, Bern, Switzerland, and Ethics Committee of the Ruhr University, Bochum, Germany. We obtained written informed consent from parents and participants older than 16 yr.

Lung Function Measurements

Multiple-breath washout. Tidal N_2-MBW tests were performed in triplicate with a validated setup (Exhalyzer D; Eco Medics, Duernten, Switzerland) (32) according to the guidelines (31). The main outcome variable was the lung clearance index (LCI) calculated from the ratio of cumulative expired volume divided by functional residual capacity (FRC). Resulting LCI units are lung turnovers. LCI reflects the lung turnover measured at one-fortieth of initial starting N_2 concentration as recommended. Alveolar phase III slopes (S_{III}) from washout breaths were automatically calculated between 65 and 95% of the expired volume with manual adjustment as appropriate. Standard corrections for tracer gas concentration and tidal volume were done automatically. Conductive ventilation heterogeneity index (Scond) was calculated from the evolution of S_{III} between lung turnovers 1.5 and 6. The first washout breath's S_{III} was used to derive acinar ventilation heterogeneity index (Sacin; 18). Please see below for detailed physiological explanations.

Single-breath washout. The tidal SBW tests were performed in triplicate using the same setup (Exhalyzer D). After established relaxed tidal breathing, measurements took one tidal inspiration and expiration from and back to FRC while the tracer gases were washed in and out. The double-tracer gas (DTG) mixture contained 26.3% He, 5% SF_6, 21% oxygen (O_2), and balanced N_2 (33). The total molar mass of this gas mixture was equal to air; therefore molar mass changes during washout reflected ventilation distribution of the tracer gases. The S_{III} quantified from the molar mass expirogram (S_{III}-DTG) was the primary outcome variable (34). Capnography was derived from the DTG-SBW, with S_{III} quantified from the CO_2 expirogram (S_{III}-CO_2) as outcome variable. We used LungSim 4.6.0 (NM Numerical Modelling, Thalwil, Switzerland) for signal processing and analyses as described (1, 29, 33).

Spirometry. Spirometry was performed according to the guidelines (24, 36) using the MasterScreen (Jaeger, Würzburg, Germany). Outcome was forced expiratory volume in 1 s (FEV_1).

Physiological Meaning of Gas Washout Signals

Both MBW and SBW tests are established methods to quantify the extent of impaired ventilation distribution efficiency and potential areas of the airway tree where ventilation inhomogeneity (VI) may predominantly arise. LCI from N_2-MBW reflects global VI, which is a mixture of VI arising in central and peripheral airways as well as in dead space (39). To characterize more specifically the location in the airway tree at which VI arises, additional outcome variables were assessed. One is the division of MBW into Scond and Sacin. Scond represents VI generated in convection-dependent preacinar airways. Sacin estimates inhomogeneity generated in the region of the diffusion- and convection-dependent front close to the entrance of acinar airways. The third method used to more specifically characterize VI is by analyzing S_{III} that are simultaneously obtained from inert gases of similar convection but different diffusion properties. For a heavy gas such as SF_6, the diffusion-convection front approximates to the mouth of the acinus and stretches into the proximal portion of the acinus. Concerning He, this front is more proximal. The DTG-SBW signal reflects a composite signal of He, SF_6, and N_2 (38). Elevated S_{III}-DTG seems to mainly reflect VI arising due to structural changes between the different diffusion-convection fronts of all three gases. Capnography is less specific and depends on perfusion, the blood air barrier, and diffusion- and convection-dependent VI as well as dead space.

Statistical Analysis

The z-scores for washout outcomes were calculated from healthy children. Upper limit of normal (ULN) and lower limit of normal (LLN) were defined as means ± 1.96 (SD) from healthy children. The z-scores for FEV_1, height, weight, and body mass index (BMI) were derived from the current standard reference equations (10, 36). For the following outcomes, higher absolute values and z-scores indicate greater disease: LCI, Scond, Sacin, and SIII-CO_2. Lower absolute values and z-scores indicate worse disease for FEV_1 and S_{III}-DTG.

For classical biased analysis, comparisons were done using Student's *t*-test, Wilcoxon rank sum tests, chi-square test, and one-way ANOVA tests, as appropriate. Post hoc tests for pairwise multiple comparisons were performed using the ANOVA with Bonferroni correction, as appropriate. All tests were two-sided with a significance level of 0.05 and performed using STATA R, Version 2.10 (Stata Statistical Software: Release 13; StataCorp, College Station, TX) (27a).

Principal component analysis (PCA) was utilized as a dimension reduction procedure to reduce the large number of variables into interpretable combination of the data. The resulting linear combination corresponds to a principal component (42). We applied PCA to identify a combination of VI outcomes that would explain more than 80% of overall variation in data. PCA performed on correlation matrix of the five IGW outcomes was used as a dimensionality reduction technique to identify which combination of five IGW outcomes might be most relevant for diagnostic purposes. For each condition (healthy, former preterm, CF, and PCD), the obtained eigenvalues and principal components (PCs) of the matrices were considered. The first PC (PC_1) accounts for the majority of the data variance; PC_1 and the corresponding values from IGW outcome variables for all children and each disease group separately were used to reflect the dominant value. PC_1 and its corresponding loading coefficients were evaluated to determine the dominant value in the lung function measurement. The data for *1)* all children and *2)* each disease group separately were plotted into the first principal axis to obtain the variance explained by PC_1 for all children and each disease group separately (21, 42).

Ward's hierarchical clustering was performed to identify physiological phenotypes based on the IGW outcome variables. We performed the clustering until the next number of clusters resulted in less than five patients in one of the clusters (43). An enrichment analysis was performed using the hypergeometric test (2) to assess overrepresentation of clinical variables within any phenotype.

RESULTS

Eighty children were enrolled (Table 1). We used 400 IGW outcomes to define different clusters.

Classical (Biased) Analysis

We performed classical analysis of pulmonary function outcome variables for the different disease groups on the group level (details in Table 2) and on an individual level (Table 3).

Lung function in former preterm children was normal on the group level apart from elevated S_{III}-CO_2. S_{III}-CO_2 was 0.7 ± 0.2 (mean ± SD) and thus slightly increased compared with healthy children (0.5 ± 0.2, $P = 0.047$). On an individual level, prevalence of pathological values did not exceed 10% of patients in any of the lung function indexes (Table 3).

CF showed significantly impaired IGW results compared with healthy children on the group level. LCI was significantly higher in CF than in healthy children (10.8 ± 2.4 vs. 7.4 ± 0.7, $P < 0.001$). Scond also differed significantly, with 0.07 ± 0.03 compared with 0.01 ± 0.02, $P < 0.001$. On an individual level, pathological values were found in 16/20 patients (80%) for LCI and Scond. Despite that, interindividual heterogeneity of VI was high, with very different z-scores for the different IGW outcome variables as shown for two specific patients with CF in Figs. 1–3.

In patients with PCD, in addition to convection-dependent VI, also diffusion- and convection-dependent VI measured by S_{III}-DTG was significantly elevated on the group level compared with healthy children (−0.2 ± 0.1 vs. −0.1 ± 0.09, $P < 0.001$). A comparable elevation was found for S_{III}-CO_2 with 0.8 ± 0.3 vs. 0.5 ± 0.2, $P = 0.004$. On an individual level, up to 30% of patients with PCD showed pathological values for peripheral VI, compared with 85% of patients with pathological values for global VI (Table 3).

Unsupervised Analysis

Principal component analysis. We applied PCA to IGW outcome variables expressed in z-scores from MBW and SBW. The first three PCs explaining 87% of total variance were selected. PC_1 reflected the comparison between S_{III}-DTG and all other outcome variables. PC_2 had a higher Scond loading. PC_3 reflected S_{III}-DTG and Sacin vs. LCI, Scond, and S_{III}-CO_2. The presence of all five IGW outcomes in the first three principal components suggests the importance of all these outcomes to explain the variability of the data and to differentiate between diseases. Therefore we included all five IGW outcomes in the cluster analysis (Tables 4 and 5 and Fig. 4, *A* and *B*).

Hierarchical Ward's clustering. Unsupervised analysis was feasible and identified three physiological phenotypes. The number of phenotypes was determined by a satisfying sample size in each phenotype as described above. The heat map containing the dendrogram obtained using clustering of IGW outcome variables is shown in Fig. 5. Enrichment analysis revealed an overrepresentation of healthy children and former preterm children in the first phenotype ($P < 0.001$) and of CF ($P = 0.03$) and PCD ($P = 0.008$) in the second phenotype. The third phenotype consisted only of patients with CF ($P < 0.001$) and PCD ($P < 0.001$). Age and other anthropometric variables were not different between phenotypes. The following IGW outcomes were different between the first and second phe-

Table 1. *Demographics of the healthy, former preterm, CF, and PCD subjects*

Characteristic	Healthy ($n = 20$)	Former Preterm ($n = 20$)	CF ($n = 20$)	PCD ($n = 20$)
Age, yr	13.5 ± 2.1	9.0 ± 1.7*	11.4 ± 2.6	13 ± 2.7
Gender, male	9 (45%)	7 (35%)	8 (40%)	8 (40%)
Weight, kg	49.5 ± 12.9	27.4 ± 4.9	36.6 ± 11.6	48.2 ± 13.6
Weight, z-score	0.1 ± 1.0	−0.4 ± 1.0	−0.5 ± 1.4	0.2 ± 1.0
BMI, kg/m²	19.4 ± 3.0	15.8 ± 1.4	17.7 ± 2.7	20.0 ± 3.3
BMI, z-score	−0.02 ± 0.9	−0.4 ± 0.9	−0.1 ± 1.0	0.2 ± 1.3
Height, cm	158.6 ± 12.6	131.5 ± 9.0	142.0 ± 16.0	153.8 ± 16.2
Height, z-score	0.2 ± 1.1	−0.2 ± 0.9	−0.5 ± 1.5	−0.08 ± 0.9

Values are means ± SD or number (percentage). CF, cystic fibrosis; PCD, primary ciliary dyskinesia. *Statistically significant differences ($P < 0.05$) in the demographic (z-scores for weight, BMI, and height) compared with the healthy group. Comparisons were done using Student's *t*-test or Wilcoxon rank sum tests, as appropriate. One-way ANOVA showed no significance for group comparison.

Table 2. *Lung function outcomes*

Index	Healthy ($n = 20$)	Former Preterm ($n = 20$)	CF ($n = 20$)	PCD ($n = 20$)
		N_2-MBW		
LCI, turnover	7.4 ± 0.7	7.5 ± 0.6	10.8 ± 2.4	11.3 ± 2.9
LCI, z-score		0.2 ± 0.9	5.0 ± 3.5	5.7 ± 4.1
Scond, %/turnover	0.01 ± 0.02	0.02 ± 0.02	0.07 ± 0.03	0.06 ± 0.02
Scond, z-score		0.5 ± 1.2	4.2 ± 2.3	3.3 ± 1.6
Sacin, %	0.09 ± 0.05	0.07 ± 0.05	0.1 ± 0.1	0.1 ± 0.1
Sacin, z-score		−0.3 ± 1.0	0.4 ± 2.0	0.5 ± 2.1
		DTG-SBW		
S_{III}-DTG, g/mol	−0.1 ± 0.09	−0.1 ± 0.09	−0.2 ± 0.1	−0.2 ± 0.1
S_{III}-DTG, z-score		0.0 ± 1.0	−0.7 ± 1.6	−1.3 ± 1.4
S_{III}-CO_2, %	0.5 ± 0.2	0.7 ± 0.2	0.9 ± 0.3	0.8 ± 0.3
S_{III}-CO_2, z-score		0.6 ± 0.8	1.4 ± 1.2	1.1 ± 1.4
		Spirometry		
FEV_1, z-score		−0.4 ± 0.9	−1 ± 1.3	−0.9 ± 1.5

Values are means ± SD for absolute values and z-scores. lung clearance index (LCI), outcome for global VI; Scond, outcome for convection-dependent VI; Sacin, outcome for the diffusion- and convection-dependent VI; S_{III}-DTG and S_{III}-CO_2, diffusion- and convection-dependent VI; forced expiratory volume in 1 s (FEV_1), outcome for central airway VI.

Table 3. *Prevalence of abnormal lung function*

Index	Healthy ($n = 20$)	Former Preterm ($n = 20$)	CF ($n = 20$)	PCD ($n = 20$)
Global ventilation inhomogeneity				
LCI, turnover	1 (5%)	2 (10%)	16 (80%)	14 (70%)
Convection-dependent ventilation inhomogeneity				
Scond, %/turnover	1 (5%)	1 (5%)	16 (80%)	17 (85%)
Peripheral ventilation inhomogeneity				
Sacin, %	1 (5%)	1 (5%)	1 (5%)	4 (20%)
S_{III}-CO_2, %	2 (10%)	2 (10%)	6 (30%)	4 (20%)
S_{III}-DTG, g/mol	1 (5%)	0 (0%)	4 (20%)	6 (30%)

Data are presented as number (percentage). Upper limit of normality was defined as mean ± 1.96 (SD) from the healthy children. CF, cystic fibrosis; PCD, primary ciliary dyskinesia. lung clearance index (LCI), outcome for global VI; Scond, outcome for convection-dependent VI; Sacin, outcome for diffusion- and convection-dependent VI; S_{III}-DTG and S_{III}-CO_2, diffusion- and convection-dependent VI.

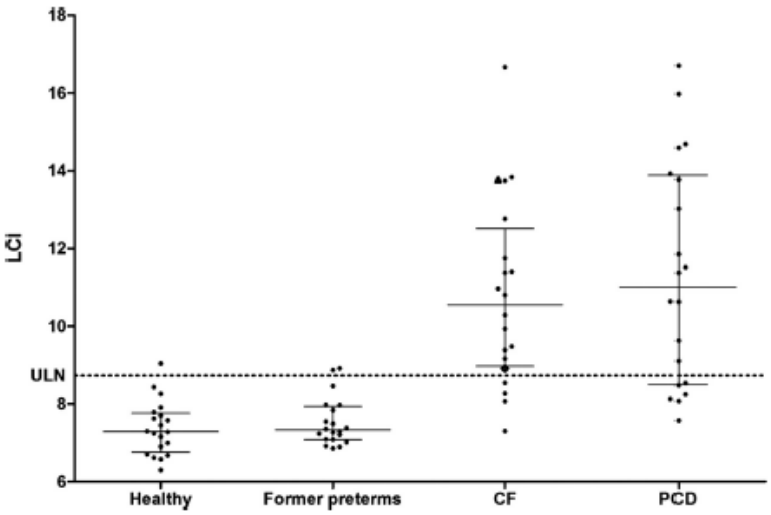

Fig. 1. LCI values in the four different groups. CF, cystic fibrosis; PCD, primary ciliary dyskinesia. ▲ and ◆ show two specific patients with CF. LCI values are given as absolute mean values. Dashed line denotes upper limit of normality, which was defined as mean ± 1.96 (SD) from the healthy children. Horizontal lines represent the median and interquartile range.

notypes: LCI ($P < 0.001$), Scond ($P < 0.001$), S_{III}-DTG ($P = 0.016$), and S_{III}-CO_2 ($P = 0.008$). Sacin did not significantly differ between the first two phenotypes. Interestingly, FEV_1 was comparable between the first and second phenotypes. All IGW outcome variables and also FEV_1 were significantly different in the third phenotype compared with the other two phenotypes (Table 6 and Fig. 5). Although patients in the third phenotype showed a higher rate and frequency of exacerbations and hospitalization, differences between phenotypes did not reach statistical significance (Table 6).

DISCUSSION

Using different IGW outcomes, we identified three main phenotypes of peripheral airway disease in children with different lung diseases independent of the underlying disease entity. While IGW outcomes were different between those phenotypes, interestingly, spirometry did not differ between the first two phenotypes. Hierarchical Ward's clustering is easily applicable to IGW outcomes and seems to be a suitable method to better characterize physiological lung disease to a large degree independent of the clinical diagnosis. Phenotype I showed normal VI. Phenotype II showed pronounced global and convection-dependent VI while diffusion-dependent VI was normal. Phenotype III was characterized by increased global and diffusion- and convection-dependent VI. The physiological clusters appear clinically meaningful. Comparing clinical characteristics, we found an increase in occurrence and frequency of exacerbations and hospitalization for intravenous antibiotic treatment from phenotypes I to III. The clusters'

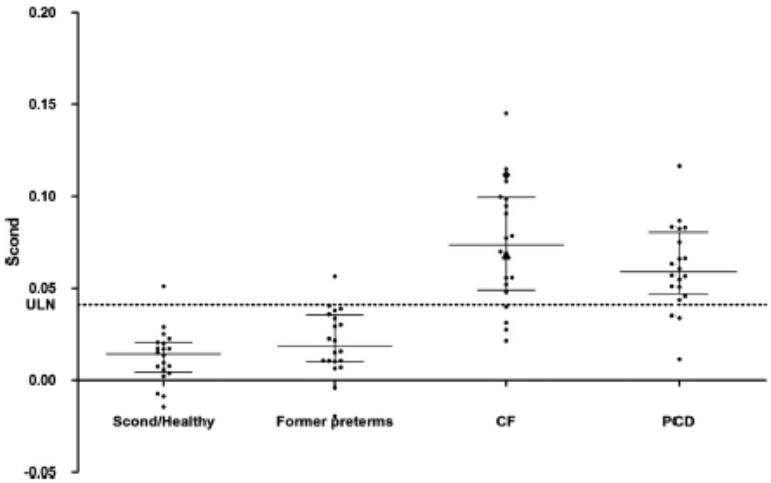

Fig. 2. Scond values in the four different groups. CF, cystic fibrosis; PCD, primary ciliary dyskinesia. ▲ and ◆ show the same two specific patients with CF as in Fig. 1. Scond values are given as absolute mean values. Dashed line denotes upper limit of normality, which was defined as mean ± 1.96 (SD) from the healthy children. Horizontal lines represent the median and interquartile range.

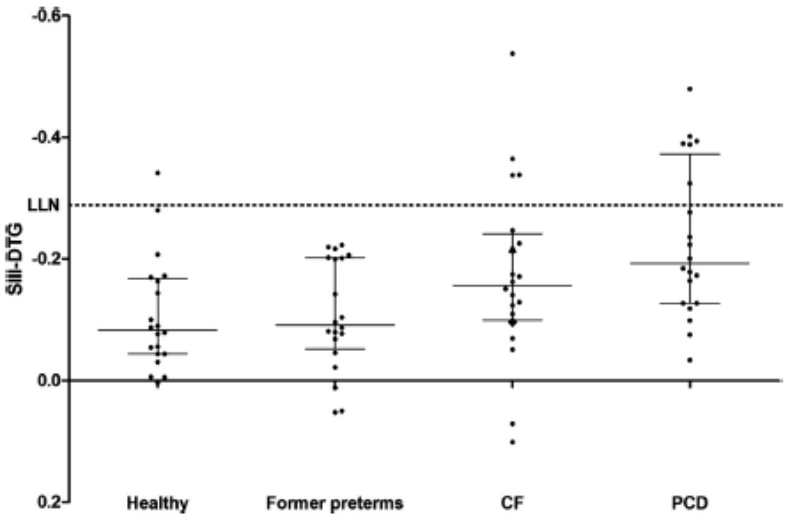

Fig. 3. S_{III}-DTG values in the four groups. CF, cystic fibrosis; PCD, primary ciliary dyskinesia. ▲ and ◆ show the same two specific patients with CF as in Figs. 1 and 2. S_{III}-DTG values are given as absolute mean values. Dashed line denotes upper limit of normality, which was defined as mean ± 1.96 (SD) from the healthy children. Horizontal lines represent the median and interquartile range.

predictive value of future disease outcomes and stability over time remain to be determined.

The first three PCs explained 87% of total variance on the basis of all five IGW outcomes. This underlines the different information obtained from all IGW outcomes. This may in part confirm previous numeric lung model work (40, 41) suggesting that indexes of global and specific VI relate to the full range of airway calibers across all generations. Classical descriptive comparison between groups did not distinguish well between individual differences in lung function parameters. This high individual heterogeneity in IGW outcome variables requires a different and independent approach to determine individual VI. We believe that unsupervised analysis using hierarchical Ward's clustering represents such an approach, especially as it is easily feasible and provides physiologically meaningful phenotypes.

These phenotypes are not necessarily specific for disease entities in children but reflect physiological relations, which are similar in CF and PCD from a functional outcome perspective. As VI in chronic lung disease is complex and dynamic over time and current methods do not provide, for example, spatial resolution (4, 5), characterization into physiological phenotypes may indeed help to personalize diagnostic procedures and therapeutic approaches in the future. Notably, convection-dependent and diffusion- and convection-dependent VI is rather independent from airway resistance formed by larger airway spaces (22).

On the basis of hierarchical Ward's clustering we identified three main phenotypes. While the majority of children in the first phenotype were healthy children and former preterm children, still a few patients with CF and PCD were included in this phenotype. The second cluster is mainly represented by CF and PCD patients, but also healthy children and former preterm children were included. A clear delineation between healthy children and patients with chronic lung disease depends on disease severity and phenotype. In agreement with that, previous studies in adult patients with chronic obstructive pulmonary disease (COPD) could also distinguish three phenotypes by using the hierarchical Ward's clustering (12). They found an overlap between healthy subjects and patients with mild COPD.

FEV_1 did not differ between the first two clusters, which is no surprise as FEV_1 is rather insensitive for structural pathology in small airways or central bronchiectasis detected in CT scans (8). This may, however, prompt the question of whether previously described associations between LCI and FEV_1 were directly related or rather an epiphenomenon of advanced airway obstruction impairing both ventilatory capacity and efficiency (8, 14, 15, 17, 23, 26). Patients in the third cluster with

Table 4. *Principal component analysis*

Lung Function Outcome	PC_1	PC_2	PC_3	PC_4	PC_5
S_{III}-DTG	0.292	0.218	0.902	0.193	−0.128
S_{III}-CO_2	−0.513	−0.270	0.065	0.811	0.056
LCI	−0.559	0.235	0.065	−0.229	−0.759
Scond	−0.409	0.726	0.048	−0.059	0.548
Sacin	−0.415	−0.546	0.419	−0.500	0.323

The five extracted principal components (PCs) and their loading coefficients for all children. S_{III}-DTG and S_{III}-CO_2, diffusion- and convection-dependent VI; lung clearance index (LCI), outcome for global VI; Scond, outcome for convection-dependent VI; Sacin, outcome for diffusion- and convection-dependent VI.

Table 5. *Percentage of total variance*

	PC_1	PC_2	PC_3	PC_4	PC_5
All	54	19	18	7	3
Healthy	34	29	20	10	7
Former preterm	50	26	17	4	3
CF	48	35	12	4	2
PCD	53	21	17	6	3

Values are given in percent. The percentages of variance accounted by each PC are derived for all subjects, healthy subjects, former preterm subjects, patients with cystic fibrosis (CF), and patients with primary ciliary dyskinesia (PCD).

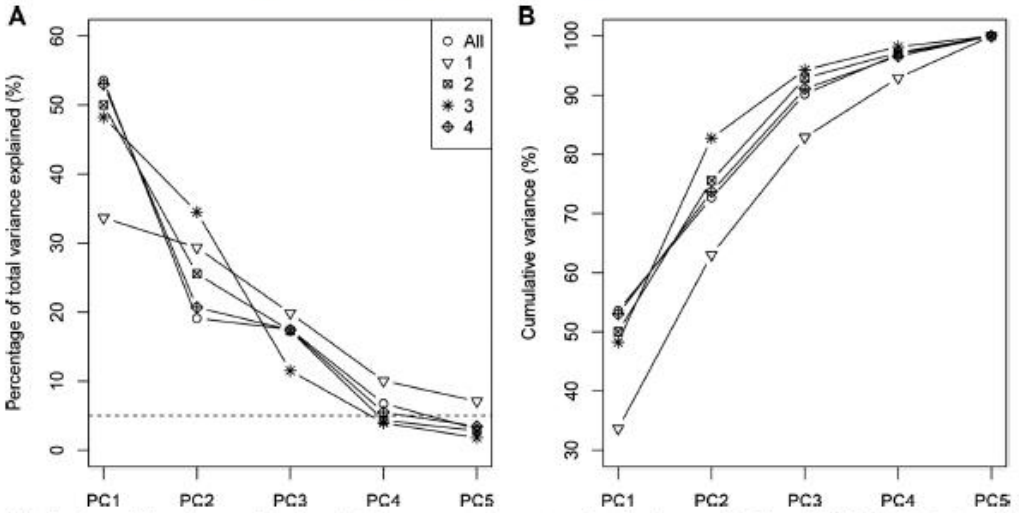

Fig. 4. Total and cumulative variance explained by principal components. *A*: percentage of total variance explained by each PC derived for the four different groups: ○, all children; ▽, only healthy children; squares with x, former preterm children; *, patients with CF; diamonds with cross, only patients with PCD. PC_1 to PC_5, the first to the fifth principal components. *B*: cumulative percentages of variance accounted by the PCs for all children.

CF and PCD were mainly characterized by marked overall and diffusion- and convection-dependent VI and a trend for a higher rate and frequency of exacerbations and hospitalization. This has also been reported in patients with severe CF lung disease (3, 15, 17, 18).

Implications of the Study

This is the first study showing that different physiological phenotypes for pediatric obstructive small airway diseases can be derived by applying established clustering methods using individual washout variables. The applied IGW seem useful for routine application because of their ease and reliability of data collection in both children and adults (20). Patients with distinct VI phenotypes may benefit from, for example, VI specific particle size of inhaled drugs to improve deposition. This seems to be a further step toward personalized medicine.

Spirometry and disease entities did not add to these phenotypes. Phenotyping the functional deficits may open up new

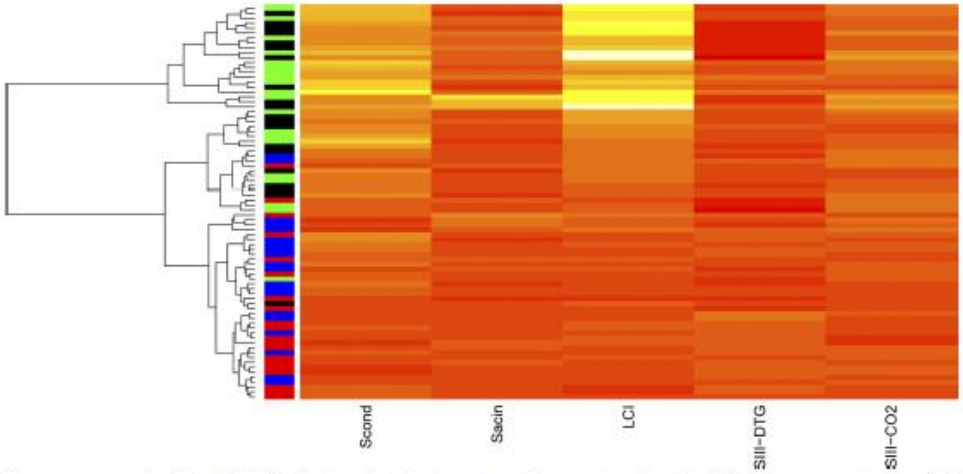

Fig. 5. Heat map representing hierarchical Ward's clustering. Scond, outcome for convection-dependent VI; Sacin, outcome for the diffusion- and convection-dependent VI; lung clearance index (LCI), outcome for global VI; S_{III}-DTG and S_{III}-CO_2, diffusion- and convection-dependent VI. The left color bar denotes individual subject grouping and their related cluster. Column on the left: red, healthy children; blue, former preterm children; green, cystic fibrosis; black, primary ciliary dyskinesia. Color gradients: brighter red tones indicate a higher z-score; darker red tones indicate a lower z-score.

Table 6. *Phenotype characteristics*

Characteristic	Phenotype I (n = 38)	Phenotype II (n = 21)	Phenotype III (n = 21)
Group			
Healthy	18 (47%)	2 (10%)	0 (0%)
Former preterm	18 (47%)	2 (10%)	0 (0%)
CF	1 (3%)	8 (38%)	11 (52%)
PCD	1 (3%)	9 (42%)	10 (48%)
Age, yr	11 ± 3	11 ± 3	13 ± 3
Weight, kg	39 ± 14	40 ± 13	44 ± 16
Height, cm	145 ± 17	145 ± 16	150 ± 18
BMI, kg/m²	18 ± 3	19 ± 3	19 ± 3
Gender, male	16 (42%)	9 (43%)	7 (33%)
LCI	−0.03 ± 0.8, $P < 0.001$*	2.2 ± 1.3, $P < 0.001$†	8.2 ± 2.8, $P < 0.001$‡
Scond	0.2 ± 1.1, $P < 0.001$*	2.4 ± 1.5, $P < 0.001$†	4.9 ± 1.6, $P < 0.001$‡
Sacin	−0.2 ± 1.0	−0.5 ± 0.6, $P < 0.001$†	1.4 ± 2.4, $P < 0.001$‡
S_{III}-DTG	0.05 ± 0.9, $P < 0.016$*	−0.9 ± 1.2	−1.1 ± 1.8, $P < 0.003$‡
S_{III}-CO_2	0.1 ± 0.8, $P < 0.008$*	1.0 ± 1.0, $P < 0.02$†	1.8 ± 1.4, $P < 0.001$‡
FEV_1	−0.3 ± 1.0	−0.2 ± 0.8, $P < 0.001$†	−1.7 ± 1.3, $P < 0.001$‡
Children with CF and PCD that had an exacerbation in the year prior to the measurement			
Exacerbation in CF and PCD	1	8	12
Children with CF and PCD that required hospitalization with antibiotic intravenous therapy in the year prior to the measurement			
Hospitalization in CF and PCD	0	1, $P = 0.07$†	6

Values are means ± SD or number (percentage). The following are given in z-scores: lung clearance index (LCI), outcome for global VI; Scond, outcome for convection-dependent VI; Sacin, outcome for the diffusion- and convection-dependent VI; S_{III}-DTG and S_{III}-CO_2, diffusion- and convection-dependent VI; forced expiratory volume in 1 s (FEV_1), outcome for central airway VI. FEV_1 is not included in cluster analysis. Groups are as follows: healthy children, former preterm children, patients with cystic fibrosis (CF), and patients with primary ciliary dyskinesia (PCD). For the following outcomes, higher absolute values and z-scores indicate greater disease: LCI, Scond, Sacin, and S_{III}-CO_2. Lower absolute values and z-scores indicate worse disease for FEV_1 and S_{III}-DTG. *Difference between phenotypes I and II. †Difference between phenotypes II and III. ‡Differences between phenotypes I and III using the ANOVA with Bonferroni correction or χ^2 test, as appropriate.

treatment areas. Previous literature points out that the success of personalized medicine depends on precise diagnostic tests to determine patients who benefit most from the specific therapy (19). We used gas wash lung function tests applicable in clinical routine (3, 20, 33). CF and PCD patients within phenotype III suffer from global and diffusion- and convection-dependent VI, which may hamper deposition and thus efficacy of inhaled drugs. Compatible with this, patients in phenotype III showed the highest occurrence and frequency of exacerbations and hospitalization in the year prior to the measurement. Taking into account that personalized medicine extends beyond only targeting therapies for patients who are already ill, aiming to identify individuals at elevated risk of clinical exacerbation that could benefit most (9), our approach seems very promising to be applied in the daily clinical routine of personalized medicine. Future studies are needed to establish if these phenotypes may be used to select particle size of inhaled drugs and predict response to treatment and later outcomes (30).

Further, we show that using commonly classical (biased) analysis does not enable one to distinguish between different physiological clusters. This has also been shown in children with severe asthma (11).

Unanswered Questions and Outlook

We did not apply computed tomography (CT) scans and thus association with structure cannot be derived from this study. Whether or not adding data from CT scans will allow even better characterization of the phenotypes is thus unknown. Previous studies found associations between global VI and acinary VI and CT scores (13, 27). We acknowledge the relatively small sample size of each disease group, which constrained internal validation; thus further external validation is warranted.

Longitudinal data are required to assess the stability of identified clusters and their ability to differentiate specific disease-independent phenotypes from disease-dependent stages.

We did not integrate analysis of fast and slowly ventilated compartments in our study (18). Thus it is unclear whether compartment analysis would have exhibited colinearity with those VI indexes assessed in this study. Another drawback is that the upper limit of normal for all outcomes is based on a sample of only 20 healthy subjects. This limits the precision and general applicability of this upper limit. However, all measurements were performed according to current guidelines (6, 29) and use of the setup without further modification. The technical error of the setup is as low as 5%. Thus our findings can be easily reproduced. Notably, the upper limits of normal are comparable with previous studies from different groups (8).

For our proof-of-principle study we only included few children and well-defined airway diseases. However, this approach is certainly applicable to other chronic lung diseases with even greater heterogeneity and larger patient groups, such as asthma or COPD, where classical diagnosis may overlap.

Conclusions

Taken together, we can easily and precisely phenotype patients independently of underlying disease entities using established clustering methods. In our proof-of-principle study, unsupervised VI analysis identified three different physiological phenotypes. Classical comparison of groups between CF and PCD was informative in a different way but did not reveal individual physiological differences independent of the disease

group. Longitudinal larger studies may establish the clusters' stability over time and potential predictive value for later outcome.

ACKNOWLEDGMENTS

The authors thank all the children and their families for their participation in the study. We express our gratitude to Anne Schlegtendal, Edgar Delgado-Eckert, Bin Xu, Gisela Wirz, and Sharon Schmid (University Children's Hospital Bochum, Basel, and Bern) for their valuable work and support.

GRANTS

The work for this report was funded by the Swiss Cystic Fibrosis Foundation (CFCH), Fondation Botnar, the Linde Group, and the Swiss National Foundation (SNF Grant 149576). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

S.N., F.S., N.K., S.Y., and P.L. conception and design of research; S.N. and S.Y. performed experiments; S.N. and N.K. analyzed data; S.N., F.S., N.K., D.M.-G., and P.L. interpreted results of experiments; S.N. and N.K. prepared figures; S.N. and F.S. drafted manuscript; S.N., F.S., N.K., S.Y., D.M.-G., C.K.-R., C.C., U.F., and P.L. edited and revised manuscript; S.N., F.S., N.K., S.Y., D.M.-G., C.K.-R., C.C., U.F., and P.L. approved final version of manuscript.

REFERENCES

1. **Abbas C, Singer F, Yammine S, Casaulta C, Latzin P.** Treatment response of airway clearance assessed by single-breath washout in children with cystic fibrosis. *J Cyst Fibros* 12: 567–574, 2013.
2. **Agresti A.** A survey of exact inference for contingency tables. *Stat Sci* 7: 131–153, 1992.
3. **Aurora P, Gustafsson P, Bush A, Lindblad A, Oliver C, Wallis CE, Stocks J.** Multiple breath inert gas washout as a measure of ventilation distribution in children with cystic fibrosis. *Thorax* 59: 1068–1073, 2004.
4. **Bauman G, Lutzen U, Ullrich M, Gaass T, Dinkel J, Elke G, Meybohm P, Frerichs I, Hoffmann B, Borggrefe J, Knuth HC, Schupp J, Prum H, Eichinger M, Puderbach M, Biederer J, Hintze C.** Pulmonary functional imaging: qualitative comparison of Fourier decomposition MR imaging with SPECT/CT in porcine lung. *Radiology* 260: 551–559, 2011.
5. **Bauman G, Puderbach M, Heimann T, Kopp-Schneider A, Fritzsching E, Mall MA, Eichinger M.** Validation of Fourier decomposition MRI with dynamic contrast-enhanced MRI using visual and automated scoring of pulmonary perfusion in young cystic fibrosis patients. *Eur J Radiol* 82: 2371–2377, 2013.
6. **Beydon N, Davis SD, Lombardi E, Allen JL, Arets HG, Aurora P, Bisgaard H, Davis GM, Ducharme FM, Eigen H, Gappa M, Gaultier C, Gustafsson PM, Hall GL, Hantos Z, Healy MJ, Jones MH, Klug B, Lodrup Carlsen KC, McKenzie SA, Marchal F, Mayer OH, Merkus PJ, Morris MG, Oostveen E, Pillow JJ, Seddon PC, Silverman M, Sly PD, Stocks J, Tepper RS, Vilozni D, Wilson NM.** An official American Thoracic Society/European Respiratory Society statement: pulmonary function testing in preschool children. *Am J Respir Crit Care Med* 175: 1304–1345, 2007.
7. **Bilton D, Canny G, Conway S, Dumcius S, Hjelte L, Proesmans M, Tummler B, Vavrova V, De Boeck K.** Pulmonary exacerbation: towards a definition for use in clinical trials. Report from the EuroCare CF Working Group on outcome parameters in clinical trials. *J Cyst Fibros* 10, *Suppl* 2: S79–S81, 2011.
8. **Boon M, Vermeulen FL, Gysemans W, Proesmans M, Jorissen M, De Boeck K.** Lung structure-function correlation in patients with primary ciliary dyskinesia. *Thorax* 70: 339–345, 2015.
9. **Dzau VJ, Ginsburg GS, Van Nuys K, Agus D, Goldman D.** Aligning incentives to fulfil the promise of personalised medicine. *Lancet* 385: 2118–2119, 2015.
10. **Flegal KM, Cole TJ.** Construction of LMS parameters for the Centers for Disease Control and Prevention 2000 growth charts. *Natl Health Stat Report* 63: 1–3, 2013.
11. **Fleming L, Murray C, Bansal AT, Hashimoto S, Bisgaard H, Bush A, Frey U, Hedlin G, Singer F, van Aalderen WM, Vissing NH, Zolkipli Z, Selby A, Fowler S, Shaw D, Chung KF, Sousa AR, Wagers S, Corfield J, Pandis I, Rowe A, Formaggio E, Sterk PJ, Roberts G.** The burden of severe asthma in childhood and adolescence: results from the paediatric U-BIOPRED cohorts. *Eur Respir J* 46: 1322–1333, 2015.
12. **Gagnon P, Casaburi R, Saey D, Porszasz J, Provencher S, Milot J, Bourbeau J, O'Donnell DE, Maltais F.** Cluster analysis in patients with GOLD 1 chronic obstructive pulmonary disease. *PLoS One* 10: e0123626, 2015.
13. **Gonem S, Hardy S, Buhl N, Hartley R, Soares M, Kay R, Costanza R, Gustafsson P, Brightling CE, Owers-Bradley J, Siddiqui S.** Characterization of acinar airspace involvement in asthmatic patients by using inert gas washout and hyperpolarized helium magnetic resonance. *J Allergy Clin Immunol* 137: 417–425, 2016.
14. **Green K, Buchvald FF, Marthin JK, Hanel B, Gustafsson PM, Nielsen KG.** Ventilation inhomogeneity in children with primary ciliary dyskinesia. *Thorax* 67: 49–53, 2012.
15. **Gustafsson PM.** Peripheral airway involvement in CF and asthma compared by inert gas washout. *Pediatr Pulmonol* 42: 168–176, 2007.
16. **Gustafsson PM, Aurora P, Lindblad A.** Evaluation of ventilation maldistribution as an early indicator of lung disease in children with cystic fibrosis. *Eur Respir J* 22: 972–979, 2003.
17. **Gustafsson PM, De Jong PA, Tiddens HA, Lindblad A.** Multiple-breath inert gas washout and spirometry versus structural lung disease in cystic fibrosis. *Thorax* 63: 129–134, 2008.
18. **Gustafsson PM, Robinson PD, Gilljam M, Lindblad A, Houltz BK.** Slow and fast lung compartments in cystic fibrosis measured by nitrogen multiple-breath washout. *J Appl Physiol (1985)* 117: 720–729, 2014.
19. **Hamburg MA, Collins FS.** The path to personalized medicine. *N Engl J Med* 363: 301–304, 2010.
20. **Husemann K, Berg N, Engel J, Port J, Joppek C, Tao Z, Singer F, Schulz H, Kohlhaufl M.** Double tracer gas single-breath washout: reproducibility in healthy subjects and COPD. *Eur Respir J* 44: 1210–1222, 2014.
21. **Johnstone IM.** On the distribution of the largest eigenvalue in principal components analysis. *Ann Stat* 29: 295–327, 2001.
22. **Macklem PT.** The physiology of small airways. *Am J Respir Crit Care Med* 157: S181–S183, 1998.
23. **Macleod KA, Horsley AR, Bell NJ, Greening AP, Innes JA, Cunningham S.** Ventilation heterogeneity in children with well controlled asthma with normal spirometry indicates residual airways disease. *Thorax* 64: 33–37, 2009.
24. **Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Jensen R, Johnson DC, MacIntyre N, McKay R, Navajas D, Pedersen OF, Pellegrino R, Viegi G, Wanger J.** Standardisation of spirometry. *Eur Respir J* 26: 319–338, 2005.
25. **Noone PG, Leigh MW, Sannuti A, Minnix SL, Carson JL, Hazucha M, Zariwala MA, Knowles MR.** Primary ciliary dyskinesia: diagnostic and phenotypic features. *Am J Respir Crit Care Med* 169: 459–467, 2004.
26. **Nyilas S, Schlegtendal A, Yammine S, Casaulta C, Latzin P, Koerner-Rettberg C.** Further evidence for an association between LCI and FEV1 in patients with PCD. *Thorax* 70: 896, 2015.
27. **Ramsey KA, Rosenow T, Turkovic L, Skoric B, Banton G, Adams AM, Simpson SJ, Murray C, Ranganathan SC, Stick SM, Hall GL.** Lung clearance index and structural lung disease on computed tomography in early cystic fibrosis. *Am J Respir Crit Care Med* 193: 60–67, 2016.

27a. **R Development Core Team.** R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2008. [available at http://www.R-project.org]

28. **Regelmann WE, Schechter MS, Wagener JS, Morgan WJ, Pasta DJ, Elkin EP, Konstan MW; Investigators of the Epidemiologic Study of Cystic Fibrosis.** Pulmonary exacerbations in cystic fibrosis: young children with characteristic signs and symptoms. *Pediatr Pulmonol* 48: 649–657, 2013.
29. **Robinson PD, Latzin P, Verbanck S, Hall GL, Horsley A, Gappa M, Thamrin C, Arets HG, Aurora P, Fuchs SI, King GG, Lum S, Macleod K, Paiva M, Pillow JJ, Ranganathan S, Ratjen F, Singer F, Sonnappa S, Stocks J, Subbarao P, Thompson BR, Gustafsson PM.** Consensus statement for inert gas washout measurement using multiple- and single- breath tests. *Eur Respir J* 41: 507–522, 2013.
30. **Shaw DE, Sousa AR, Fowler SJ, Fleming LJ, Roberts G, Corfield J, Pandis I, Bansal AT, Bel EH, Auffray C, Compton CH, Bisgaard H,**

Bucchioni E, Caruso M, Chanez P, Dahlen B, Dahlen SE, Dyson K, Frey U, Geiser T, Gerhardsson de Verdier M, Gibeon D, Guo YK, Hashimoto S, Hedlin G, Jeyasingham E, Hekking PP, Higenbottam T, Horvath I, Knox AJ, Krug N, Erpenbeck VJ, Larsson LX, Lazarinis N, Matthews JG, Middelveld R, Montuschi P, Musial J, Myles D, Pahus L, Sandstrom T, Seibold W, Singer F, Strandberg K, Vestbo J, Vissing N, von Garnier C, Adcock IM, Wagers S, Rowe A, Howarth P, Wagener AH, Djukanovic R, Sterk PJ, Chung KF. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. *Eur Respir J* 46: 1308–1321, 2015.

31. **Sileo C, Corvol H, Boelle PY, Blondiaux E, Clement A, Ducou Le Pointe H.** HRCT and MRI of the lung in children with cystic fibrosis: comparison of different scoring systems. *J Cyst Fibros* 13: 198–204, 2014.
32. **Singer F, Houltz B, Latzin P, Robinson P, Gustafsson P.** A realistic validation study of a new nitrogen multiple-breath washout system. *PLoS One* 7: e36083, 2012.
33. **Singer F, Stern G, Thamrin C, Abbas C, Casaulta C, Frey U, Latzin P.** A new double-tracer gas single-breath washout to assess early cystic fibrosis lung disease. *Eur Respir J* 41: 339–345, 2013.
34. **Singer F, Stern G, Thamrin C, Fuchs O, Riedel T, Gustafsson P, Frey U, Latzin P.** Tidal volume single breath washout of two tracer gases–a practical and promising lung function test. *PLoS One* 6: e17588, 2011.
35. **Spycher BD, Silverman M, Pescatore AM, Beardsmore CS, Kuehni CE.** Comparison of phenotypes of childhood wheeze and cough in 2 independent cohorts. *J Allergy Clin Immunol* 132: 1058–1067, 2013.
36. **Stanojevic S, Wade A, Cole TJ, Lum S, Custovic A, Silverman M, Hall GL, Welsh L, Kirkby J, Nystad W, Badier M, Davis S, Turner S, Piccioni P, Vilozni D, Eigen H, Vlachos-Mayer H, Zheng J, Tomalak W, Jones M, Hankinson JL, Stocks J.** Spirometry centile charts for young Caucasian children: the Asthma UK Collaborative Initiative. *Am J Respir Crit Care Med* 180: 547–552, 2009.
38. **Verbanck S, Paiva M.** Dual gas techniques for peripheral airway function: diffusing the issues. *Eur Respir J* 45: 1491–1494, 2015.
39. **Verbanck S, Paiva M, Schuermans D, Hanon S, Vincken W, Van Muylem A.** Relationships between the lung clearance index and conductive and acinar ventilation heterogeneity. *J Appl Physiol (1985)* 112: 782–790, 2012.
40. **Verbanck S, Paiva M, Schuermans D, Malfroot A, Vincken W, Vanderhelst E.** Acinar and conductive ventilation heterogeneity in severe CF lung disease: back to the model. *Respir Physiol Neurobiol* 188: 124–132, 2013.
41. **Verbanck S, Schuermans D, Van Muylem A, Melot C, Noppen M, Vincken W, Paiva M.** Conductive and acinar lung-zone contributions to ventilation inhomogeneity in COPD. *Am J Respir Crit Care Med* 157: 1573–1577, 1998.
42. **Wang TY, Pao JL, Yang RS, Jang JS, Hsu WL.** The adaptive changes in muscle coordination following lumbar spinal fusion. *Hum Mov Sci* 40: 284–297, 2015.
43. **Ward JH Jr.** Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58: 236–244, 1963.
44. **Weibel ER.** A retrospective of lung morphometry: from 1963 to present. *Am J Physiol Lung Cell Mol Physiol* 305: L405–L408, 2013.
45. **Weibel ER.** What makes a good lung? *Swiss Med Wkly* 139: 375–386, 2009.

9. Curriculum Vitae

Personal data

Name	Delphine Meier-Girard
Place, Country - Date of birth	Clamart, France – 07.12.1984
Marital status	Married
Nationality	French
E-mail	delphine.gd@gmail.com

Education and work experience

08/2014-to date	**PhD student in epidemiology,** University of Basel, Switzerland PhD project: "The fluctuation behavior of heart and respiratory system signals as a quantitative tool for studying long-term environmental exposures and chronic diseases"
28/09/2012	**National Diploma of Pharmacy**, Université Paris 11, France
09/2012	**M.Sc. Public Health** with honours, Université Paris 7, France "Methods in Therapeutic Evaluation: biostatistics, epidemiology, clinical research"
2008-2012	**Residency**, Ile-de-France region, France - National entry exam: ranked in the first 15%
2007-2008	**Residency,** Rhône-Alpes region, France - National entry exam: ranked in the first 20%
2002-2007	**University of Pharmacy**, Université Paris 11, France - entry exam: ranked in the first 5%
2002	**Scientific baccalaureate** (specialty mathematics, with honours), Antony, France

Dr. Pharm. Thesis

28/09/2012	"Les difficultés du démarrage des études cliniques pédiatriques institutionnelles: une approche qualitative et quantitative" (Obstacles to the implementation of academic paediatric clinical studies: a qualitative and quantitative approach). Under the supervision of Prof. D. Porquet, Université Paris 11, France

Professional experience

08/2014-to date	**PhD student, Computational Physiology and Biostatistics Group,** University Children's hospital (Basel, Switzerland)
11/2013-08/2014	**Research assistant, Computational Physiology and Biostatistics Group,** University Children's hospital (Basel, Switzerland)
11/2102-11/2013	**Research fellow** (Assistante hospitalo-universitaire), **Clinical Research Unit**, Hôpital Universitaire Robert Debré, AP-HP (Paris, France)
2007-2012	**Residency**

05/2012–11/2012	**Department of Infectious Diseases**, Institut National de Veille Sanitaire - InVS (Saint Maurice, France)
11/2011–05/2012	**Unit of Biostatistics and Epidemiology**, Hôpital Universitaire Necker Enfants Malades, AP-HP (Paris, France)
05/2011–11/2011	**Clinical Investigation Center (**CIC), Hôpital Universitaire Pitié Salpêtrière, AP-HP (Paris, France)
11/2010–05/2011	**Clinical Research Unit**, Hôpital Universitaire Robert Debré, AP-HP (Paris, France)
05/2010–11/2010	**Public Health Unit**, Hôpital Universitaire Lariboisière, AP-HP (Paris, France)
05/2009–05/2010	**Clinical Research Unit**, Hôpital Universitaire Robert Debré, AP-HP (Paris, France)
11/2008–05/2009	**Department of Pharmacy**, Hôpital Universitaire Tenon, AP-HP (Paris, France)
11/2007–11/2008	**Medical Information Unit - Centre de Documentation et d'Informations Médicales**, Pharmacie Centrale, Hospices Civils de Lyon (Lyon, France)

Publications peer-reviewed

1. Assessment of the effect of chest physiotherapy with the increased-exhalation technique on the respiratory parameters of infants with a first episode of bronchiolitis. D. Evenou, S. Sebban, C. Fausser, **D. Girard**. Kinesitherapie, la Revue 2017;17(187):3-8
2. Physiological phenotyping of pediatric chronic obstructive airway diseases. S. Nyilas, F. Singer, N. Kumar, S. Yammine, **D. Meier-Girard**, C. Koerner-Rettberg, C. Casaulta, U. Frey, P. Latzin. Journal of Applied Physiology 2016 Jul 1;121(1):324-32.
3. Academic pediatric clinical research: factors associated with study implementation duration. **D. Meier-Girard**, A. Tibi, H. Abdoul, S. Prot-Labarthe, F. Brion, O. Bourdon, C. Alberti. BMC Medical Research Methodology 2016 Mar 29;16(1):36.
4. Improvement of medical care in a cohort of newborns with sickle-cell disease in North Paris: impact of national guidelines. N. Couque, **D. Girard**, R. Ducrocq, P. Boizeau, Z.Haouari, F. Missud, L. Holvoet, G. Ithier, M. Belloy, M.-H. Odièvre, M. Benemou, P. Benhaim, B. Retali, P. Bensaid, B. Monier, V. Brousse, R. Amira, C. Orzechowski, E. Lesprit, L. Mangyanda, N. Garrec, J. Elion, C. Alberti, A. Baruchel, M. Benkerrou. British Journal of Heamatology. 2016 Jun;173(6):927-37.
5. Evaluation of frequency of pediatric oral liquid medication dosing errors by caregivers: amoxicillin and josamycin examples. A. Berthe-Aucejo, **D. Girard**, M. Lorrot, X. Bellettre, A. Faye, J.C. Mercier, F. Brion, O. Bourdon, S. Prot-Labarthe. Archives of Disease in Childhood 2016 Jan.
6. Smoking cessation and heart rate dynamics: Is it possible to fully recover? **D. Girard**, E. Delgado-Eckert, E. Schaffner, C. Häcki, M. Adam, G. Stern, N. Kumar, D. Felber Dietrich, A. Turk, M., N. Künzli , J.-M. Gaspoz, T. Rochat, C. Schindler, N. Probst-

Hensch, U. Frey. Environmental Research 2015; 143(2015):39–48.

7. Long-term intestinal failure and home parenteral nutrition in children: changes over the last ten years. L-M. Petit, **D. Girard**, S. Irtan, C. Elie, V. Colomb. Journal of Pediatric Gastroenterology and Nutrition 2015 Sep 17.
8. Prevalence and predictors of early cardiovascular events after kidney transplantation: evaluation of pre-transplant cardiovascular work-up. M. Delville, L. Sabbah, **D. Girard**, C. Elie, F. Martine, A. Méjean, C. Legendre, R. Sberro-Soussan. PLOS One 2015 June 24; 10(6):e0131237.
9. Epidemiology of pulmonary tuberculosis in France. Can the hospital discharge database be a reliable source of information? **D. Girard**, D. Antoine, D. Che. Médecine et Maladies Infectieuses. 2014 Nov 13;44(11-12):509-514.
10. Pregnancy-related listeriosis in France: 1984 to 2011, with a summary of 606 cases occurring from 1999 to 2011. **D. Girard**, A. Leclercq, E. Laurent, M. Lecuit, H. de Valk, V. Goulet. Euro Surveillance. 2014 Sep 25;19(38)
11. Improvement of radiology requisition. P. Troude, A. Dozol, P. Soyer, **D. Girard**, F. Martinez, B. Montagne, C. Segouin. Diagnostic and Interventional Imaging 2014. 95: 69-75
12. Vitamin D, Renal Function and Outcome in Renal Transplantation. F. Bienaimé, **D. Girard**, D. Anglicheau, G. Canaud, J-C. Souberbielle, H. Kreis, G. Friedlander, C. Elie, C. Legendre, D. Prié. Journal of the American Society of Nephrology 2013 Apr; 24(5):831-41
13. How to improve the implementation of academic pediatric clinical trials involving drug therapy: a qualitative study. **D. Girard**, O. Bourdon, H. Abdoul, S. Prot-Labarthe, F. Brion, A. Tibi, C. Alberti. PLOS One 2013 May 28; 8(5):e64516.

Manuscripts in preparation

1. Evidence of adverse effects of long-term exposure to traffic-related PM_{10} on heart rate variability and heart rate dynamics in healthy subjects. **D. Meier-Girard**, E. Delgado-Eckert, E. Schaffner, U. Frey, N. Probst-Hensch
2. Novel phenotyping based on lung function fluctuation clusters in asthma and COPD (BIOAIR study). **D. Meier-Girard**, E. Delgado-Eckert, M. Kupczyk, A. James, R. Middelveld, S.E. Dahlén, U. Frey

Oral communications

09/2015	**European Respiratory Society (Amsterdam, Netherlands)** Detecting specific clusters of asthma and COPD patients using fluctuation based clustering of daily lung function recordings in the BIOAIR study
05/2012	**Clinical Epidemiology Congress - EPICLIN 6 (Lyon, France)** Academic Pediatric research: Factors Associated with the Delay of Studies Implementation
03/2011	**Public Health Congress - EMOIS (Nancy, France)**

Standardized and Computerized Radiology Requisition Improves the Radiology Process

Poster discussions

09/2014 **European Respiratory Society (Munich, Germany)**
Smoking cessation and heart rate dynamics: is it possible to fully recover? **D. Girard**, E. Delgado-Eckert, M. Adam, C. Häcki, E. Schaffner, C. Autenrieth, N. Künzli, J.-M. Gaspoz, T. Rochat, N. Probst-Hensch, U. Frey

Poster presentations

10/2012 **12th Annual Congress of the French Speaking Society of Transplantation (Nantes, France)**
Prevalence and prediction of cardiovascular events after kidney transplantation. M. Delville, **D. Girard**, L. Sabbah, C. Elie, C. Legendre

05/2011 **Clinical Epidemiology Congress - EPICLIN 5 (Marseille, France)**
Pediatric clinical trials involving drug therapy: running the gauntlet. **D. Girard**, O. Bourdon, A. Tibi, S. Prot-Labarthe, H. Abdoul, C. Alberti

Awards

Best paper of 2015 in Environmental Research
Smoking cessation and heart rate dynamics: Is it possible to fully recover? D. Girard et al. Environmental Research 2015; 143(2015):39–48