POLITECNICO DI TORINO

Master's Degree
in Data Science and Engineering

Master's Thesis

Association of the Capillary Metabolome with Cardiorespiratory Fitness and Metabolic Response to Acute Exercise



Supervisors:

prof. Gabriella Olmo

Candidate

Lorenzo de Gregorio

Academic year 2025-2026

Abstract

Acute exercise metabolomics is commonly based on a single pre-exercise measurement, which limits the ability to determine whether post-exercise changes exceed short-term resting variability. This thesis investigated whether repeated-baseline capillary blood metabolomics can improve the interpretation of acute metabolic responses to maximal exercise and whether baseline metabolomic information contributes to the prediction of cardiorespiratory fitness.

A total of 100 healthy adults completed maximal cardiopulmonary exercise testing. Capillary blood samples were collected at two resting baseline timepoints and immediately after exercise. Targeted metabolomics data were filtered to retain 52 metabolites, including amino acids, amino acid-related metabolites, fatty acids, and lactate. Baseline reproducibility was assessed using within-subject variability, correlations and exploratory dimensionality reduction. Acute exercise responses were evaluated through principal component analysis, conventional single-baseline comparisons and a signal-to-noise ratio framework, in which exercise reactivity was interpreted relative to baseline-derived methodological noise. Participant-level determinant analyses examined the associations of physiological and experimental characteristics with noise-adjusted metabolite responses. In addition, baseline metabolomic and participant-level features were used to predict VO_2peak through regularized regression and nested cross-validation.

The repeated baseline samples showed strong overall agreement, with a global correlation of $R = 0.996$ and a median metabolite-level correlation of $R = 0.830$. Principal component analysis showed a clear separation between resting and post-exercise samples, indicating a broad exercise-induced shift in the capillary metabolome. Forty of the 52 retained metabolites were significantly altered after exercise, with lactate showing the strongest response. However, signal-to-noise stratification showed that response robustness varied substantially across metabolites, with 46 metabolites displaying an average post-exercise response larger than baseline-derived variability.

Exhaustion status and fasting status were the main determinants of inter-individual response variability. Participants classified as potentially not exhausted showed attenuated amino acid-related responses, whereas fasting status was associated with a smaller subset of metabolite responses. The prediction analysis showed that VO_2peak was estimated with moderate accuracy from participant-level clinical features ($R^2 = 0.606$; MAE = 4.322 ml min^{-1} kg^{-1}), whereas baseline metabolomic features provided limited standalone and incremental predictive value.

Overall, repeated-baseline capillary metabolomics provides a useful framework for interpreting acute exercise responses relative to short-term resting variability. The main contribution of this thesis is methodological: repeated baseline sampling supports a noise-aware interpretation of metabolite changes and helps distinguish broad group-level perturbations from responses that more clearly exceed baseline-derived methodological noise.

Ringraziamenti

.

Index

1 Introduction 1

2 Scientific and Methodological Background 4

2.1 Cardiorespiratory Fitness and Cardiopulmonary Exercise Testing 4

2.1.1 VO_2max, VO_2peak and Maximal-Effort Assessment 4

2.1.2 Verification Bouts and Complementary CPET Markers 5

2.2 Acute Exercise Metabolism and Circulating Metabolites 5

2.3 Metabolomics in Exercise Physiology 6

2.3.1 Targeted and Untargeted Metabolomics 7

2.3.2 Analytical and Pre-Analytical Considerations 7

2.3.3 Capillary and Venous Blood as Metabolomic Matrices 8

2.4 Capillary Blood Sampling and Repeated Baseline Designs 9

2.5 Biological Variability and Noise-Aware Response Analysis 10

2.5.1 Sources of Metabolomic Variability 10

2.5.2 Repeated Baselines as an Operational Uncertainty Estimate 10

2.5.3 Individual Reference Ranges and Responder Interpretation 11

2.6 Multivariate and Predictive Modelling in Metabolomic Data 13

2.6.1 Association, Explanation and Prediction 14

2.6.2 Small-Sample Biomedical Machine Learning 15

2.6.3 Leakage, Nested Validation and Stability Selection 15

3 Materials and Methods 16

3.1 Study Design and Cohort Overview 16

3.1.1 Cohort Characteristics and Analytical Subsets 17

3.2 Cardiopulmonary Exercise Testing Protocol 18

3.3 Capillary Blood Sampling and Measurement Timepoints 19

3.4 Targeted Metabolomics Panel and Preprocessing 20

3.5 Data Quality Control and Transformation 21

3.6 Baseline Reliability Analysis 22

3.7 Single-Baseline and Noise-Adjusted Exercise Response Analysis 22

3.8 Participant-Level Determinant Analysis 24

3.9 Directional Response, Baseline-Envelope and Exploratory Responder Analysis 25

3.10 Predictive Modelling of VO_2peak 26

3.11 Computational Reproducibility and Software Environment 28

3.12 Statistical Decision Rules and Sensitivity Analyses 28

4 Baseline Reliability and Individual Metabolic Fingerprints 29

4.1 Baseline Reproducibility Between Repeated Resting Samples 29

4.2 Metabolite-Level Baseline Agreement 31
4.3 Individual Metabolic Fingerprints 32
4.4 Additional Reliability Checks 33
4.5 Chapter Summary 35
5 Acute Exercise Response and Noise-Adjusted Metabolite Ranking 36
5.1 Global Metabolomic Shift After Maximal Exercise 36
5.2 Metabolite Class Contribution to the Exercise Response 38
5.3 Participant Characteristics and PCA Structure 39
5.4 Conventional Single-Baseline Post-Exercise Response 40
5.4.1 Robustness of Exploratory Response-Pattern Clustering 40
5.5 Signal-to-Noise Stratification of Exercise Responses 42
5.6 Chapter Summary 43
6 Participant-Level Determinants and Directional Response Patterns 44
6.1 Rationale for Participant-Level Determinant Analysis 44
6.2 Exhaustion Status and Attenuated Amino-Acid Responses 46
6.3 Fasting Status and Amino-Acid Response Variability 47
6.4 Limited Role of Other Participant Characteristics 48
6.5 Directional Response and Baseline-Envelope Analysis 49
6.6 Exploratory Responder Analysis Based on Post-Exercise Deviations 51
6.7 Chapter Summary 52
7 Prediction of VO_2peak from Baseline Metabolomic and Participant-Level Features 54
7.1 Prediction Task and Analytical Rationale 54
7.2 Feature Sets and Modelling Strategy 54
7.3 Predictive Performance 56
7.4 Feature Contribution and Stable Predictors 57
7.5 Sensitivity Analysis Using the Full Metabolite Panel 59
7.5.1 Repeated Cross-Validation Variability 60
7.6 Interpretation and Limitations of the Prediction Analysis 61
7.7 Chapter Summary 62
8 Discussion 63
8.1 Main Findings 63
8.2 Repeated Baseline Design and Methodological Noise 64
8.3 Acute Exercise Response and Noise-Adjusted Metabolite Ranking 66
8.4 Participant-Level Determinants of Response Variability 68
8.5 Exploratory Responder Patterns 70
8.6 Prediction of VO_2peak 71

8.7 Strengths and Limitations 73
8.8 Future Perspectives 75
9 Conclusion 76
10 References 78

1 Introduction

Cardiorespiratory fitness is one of the most relevant indicators of physical performance and physiological health. It reflects the integrated ability of the cardiovascular, respiratory and muscular systems to transport and utilize oxygen during exercise. In both research and clinical settings, cardiorespiratory fitness is commonly assessed through cardiopulmonary exercise testing, with peak oxygen uptake (VO_2peak) representing one of the most widely used markers of aerobic capacity. Differences in VO_2peak are associated with differences in exercise tolerance, metabolic efficiency and overall functional capacity. For this reason, understanding the biological factors associated with cardiorespiratory fitness remains an important objective in exercise physiology and health-related research.

Acute physical exercise represents a strong physiological stimulus that rapidly alters energy metabolism. During intense exercise skeletal muscle increases its demand for ATP, leading to changes in carbohydrate metabolism, lactate production, amino acid turnover and lipid mobilization. These processes are reflected in circulating metabolites, which can therefore provide a molecular snapshot of the systemic response to exercise[1–6]. Metabolomics has become a valuable tool in this context because it allows the simultaneous quantification of multiple small molecules involved in substrate utilization, energy production and recovery. Targeted metabolomics is particularly useful when the aim is to focus on metabolite classes with direct physiological relevance, such as amino acids, amino-acid-related compounds, fatty acids and lactate [7,8].

Capillary blood sampling provides a practical and minimally invasive approach for studying exercise-induced metabolic changes. Compared with conventional venous blood sampling, capillary sampling requires a smaller blood volume and can be repeated within short time intervals, making it suitable for exercise protocols that require multiple measurements before and after physical effort [9–13]. This is especially relevant in acute exercise studies, where metabolic changes may occur rapidly and repeated sampling can provide a more detailed picture of the physiological response. However, the interpretation of capillary blood metabolomics data requires careful attention to variability. Circulating metabolite concentrations are not completely stable, even under resting conditions, and may fluctuate because of biological, nutritional, technical and sampling-related factors.

This issue is particularly important when exercise-induced changes are interpreted from a single pre-exercise baseline. A conventional single-baseline approach can identify metabolites that differ significantly after exercise at the group level, but it does not indicate whether the observed post-exercise change is larger than the short-term variability of that metabolite at rest. As a result, statistical significance does not necessarily imply that a metabolite shows a reliable individual-level physiological response. This distinction is important because two metabolites may both be significantly altered after exercise, while only one of them may show a response that clearly exceeds resting variability. Therefore, a robust interpretation of acute exercise metabolomics requires not only detecting post-exercise changes, but also comparing these changes against baseline fluctuation [1–6].

Repeated baseline sampling provides a way to address this limitation. By collecting two independent resting samples before exercise, it becomes possible to estimate short-term within-subject variability for each metabolite [14,15]. This repeated-baseline information can then be used as a reference to evaluate the magnitude of the post-exercise response. In this framework, exercise-induced changes can be interpreted not only through conventional statistical significance, but also through noise-aware metrics such as signal-to-noise ratios. A signal-to-noise approach can rank metabolites according to whether their exercise response is comparable to, moderately larger than, or clearly larger than the estimated methodological noise. This allows a more informative distinction between broad group-level metabolic shifts and metabolite responses that are more reliable at the individual level.

The present thesis is based on a capillary blood metabolomics study in which 100 healthy adults underwent maximal cardiopulmonary exercise testing. Two resting baseline samples were collected before exercise, followed by an immediate post-exercise sample. This design allows the resting capillary metabolome to be evaluated for reproducibility and individuality before investigating the acute metabolic response to maximal exercise. The repeated baseline is therefore not treated only as a quality-control procedure, but as a central analytical component of the study. It provides the basis for estimating resting variability, interpreting post-exercise changes and identifying metabolites whose response exceeds baseline noise.

The first part of the thesis focuses on the reliability of the capillary metabolome at rest. The comparison between the two baseline samples is used to evaluate intra-individual variability, inter-individual variability and baseline agreement across the retained metabolite panel. In addition, dimensionality reduction techniques are used to assess whether repeated samples preserve stable individual metabolic profiles, or metabolic fingerprints. This step is essential because the interpretation of post-exercise responses depends on the stability of the resting reference state. If baseline measurements show good reproducibility, then the repeated-baseline design can be used to distinguish true exercise-related changes from short-term resting fluctuation.

The second part of the thesis examines the acute metabolic response to maximal exercise. Multivariate analyses are used to evaluate whether post-exercise samples separate from resting samples and to identify the metabolite classes that drive this separation. Particular attention is given to amino acids, amino-acid-related compounds, fatty acids and lactate, because these metabolites are closely connected to substrate use and exercise metabolism. A conventional single-baseline analysis is first used to describe the overall post-exercise response and to identify metabolites that change significantly after exercise. This provides a useful group-level overview, but it does not fully account for resting variability.

The third part of the thesis introduces the repeated-baseline information into the response analysis. Signal-to-noise ratios are used to compare the magnitude of each metabolite's exercise response with its estimated baseline fluctuation. This allows metabolites to be ranked according to the strength of their response relative to resting noise. In this way, the thesis moves beyond the simple question of whether a metabolite changes significantly after exercise and addresses a more specific question: whether the response is large enough to be considered reliable compared with the variability observed before exercise. This distinction represents one of the central methodological contributions of the work.

The fourth part of the thesis investigates inter-individual variability in the noise-adjusted metabolic response. Even when a metabolite shows a clear post-exercise change at the group level, the magnitude of that response may differ across participants. Participant-level characteristics such as VO_2peak, body composition, age, sex, fasting status and exhaustion status may contribute to this variability. In particular, exhaustion status is relevant because incomplete attainment of maximal effort during cardiopulmonary exercise testing may attenuate the metabolic response. Fasting status is also important because pre-exercise nutritional state can influence substrate availability and amino acid metabolism. By modelling the association between participant characteristics and noise-adjusted metabolite responses, the thesis aims to identify which physiological or experimental factors partly explain individual response patterns.

As a complementary step, individual response directions and exploratory responder patterns can be evaluated relative to baseline-derived reference ranges. This allows post-exercise changes to be interpreted not only at the group level, but also in terms of whether individual participants show responses that exceed their expected resting variability. This approach classifies post-exercise metabolite values as increased, decreased or unchanged compared with the interval defined by the two baseline samples. Although this analysis is not the main focus of the central paper, it provides an additional way to investigate heterogeneous or bidirectional response patterns. Metabolites that show

mixed response directions may be particularly interesting because they could reflect differences in nutritional state, substrate preference, metabolic flexibility or other individual factors not fully captured by standard participant characteristics.

Finally, the thesis extends the study toward a more data-science-oriented objective: the prediction of VO_2peak from metabolomic and participant-level variables. Since VO_2peak is a key marker of cardiorespiratory fitness, predicting it from baseline metabolomic profiles and physical characteristics can help assess whether the capillary metabolome contains information related to individual fitness status. Different feature sets can be compared, including anthropometric and physiological variables, metabolomic variables and their combination. This predictive analysis allows the thesis to move from descriptive and inferential metabolomics toward model-based interpretation, evaluating both prediction accuracy and the relative contribution of individual variables.

Overall, this thesis aims to show that repeated-baseline capillary blood metabolomics provides a robust framework for studying acute exercise responses. The central idea is that post-exercise metabolic changes should not be interpreted only in terms of group-level statistical significance, but also in relation to the variability observed under resting conditions. By combining baseline reliability analysis, multivariate exploration, signal-to-noise stratification, participant-level modelling and predictive analysis of VO_2peak, the thesis connects exercise physiology with data science. The final objective is to improve the interpretation of the capillary metabolome as both a marker of acute metabolic response and a potential source of information about cardiorespiratory fitness.

2 Scientific and Methodological Background

2.1 Cardiorespiratory Fitness and Cardiopulmonary Exercise Testing

Cardiorespiratory fitness describes the ability of the body to take up, transport and use oxygen during physical exercise. It depends on the combined function of the respiratory system, the cardiovascular system and the muscles. One of the most common indicators of cardiorespiratory fitness is peak oxygen uptake, usually reported as VO_2peak. VO_2peak represents the highest oxygen uptake reached during an exercise test and is commonly expressed relative to body mass as $mL \cdot min^{-1} \cdot kg^{-1}$.

In this thesis, VO_2peak is important for two reasons. First, it is one of the main physiological variables used to describe the participants. Second, it is later used as the target variable in the predictive modelling analysis. For this reason, VO_2peak provides a connection between the physiological part of the study and the data science objective of predicting cardiorespiratory fitness from metabolomic and participant-level features.

VO_2peak is usually measured through cardiopulmonary exercise testing. During CPET, exercise intensity is progressively increased while oxygen uptake, carbon dioxide production and ventilation are measured. This allows the evaluation of both maximal and submaximal responses to exercise. Besides VO_2peak, CPET provides additional variables such as ventilatory thresholds and respiratory exchange ratio. Ventilatory thresholds describe changes in the physiological response to increasing exercise intensity, while respiratory exchange ratio gives information about the balance between oxygen uptake and carbon dioxide production.

Peak RER is also useful to evaluate whether a participant reached a high level of effort during the test. This aspect is relevant because an incomplete maximal effort may lead to a weaker physiological and metabolic response. In this thesis, exhaustion status is therefore considered as a participant-level variable when analysing metabolite responses after exercise. This is important because differences in the metabolomic response may depend not only on individual biology, but also on how completely the participant reached maximal effort during the test [16–18].

2.1.1 VO_2max, VO_2peak and Maximal-Effort Assessment

The distinction between VO_2max and VO_2peak is important when interpreting cardiopulmonary exercise testing. VO_2max is traditionally defined as the highest rate of oxygen uptake that can be achieved during progressive exercise, ideally identified by a plateau in oxygen uptake despite a further increase in external workload[19]. In practice, however, a clear plateau is not observed in all participants, particularly in heterogeneous cohorts, clinical populations or studies in which participants differ substantially in training status and familiarity with maximal exercise testing[19,20].

For this reason, VO_2peak is frequently used as an operational measure of maximal aerobic capacity. VO_2peak refers to the highest oxygen uptake value observed during the test, regardless of whether a formal plateau criterion is achieved[19,20]. This does not make VO_2peak less informative, but it requires careful interpretation because the measured value depends on the exercise protocol, participant motivation, tolerance of discomfort and the degree to which maximal effort was attained. In the present thesis, VO_2peak was used as the main marker of cardiopulmonary fitness because it represented the highest oxygen uptake achieved during the incremental CPET and was directly available for all participants.

The interpretation of VO_2peak should therefore not rely on a single physiological criterion. A plateau in oxygen uptake, peak respiratory exchange ratio, peak heart rate, subjective exhaustion and the duration of the test can all provide complementary information regarding effort attainment. However, each of these markers has limitations when considered in isolation. For example, peak RER may be

influenced by hyperventilation and buffering of exercise-induced acidosis, while heart-rate thresholds based on age-predicted values may show substantial inter-individual variation. A multi-criterion interpretation is therefore more appropriate than defining maximal effort from one threshold alone.

This distinction is particularly relevant in the present study because VO_2peak was used not only as a descriptive physiological variable, but also as the target of the predictive modelling analysis. If the measured VO_2peak does not adequately represent the highest attainable oxygen uptake for a participant, its use as a prediction target may introduce additional uncertainty. The verification bout was included in the study protocol to reduce this uncertainty and to provide an additional physiological quality-control step[21].

2.1.2 Verification Bouts and Complementary CPET Markers

A verification bout is an additional exercise phase performed after incremental CPET, usually at a severe or supra-maximal workload. The highest oxygen uptake achieved during the verification bout can be compared with the value observed during the incremental test. When the two values are sufficiently similar, this supports the interpretation that the incremental test elicited the highest attainable oxygen uptake for that participant[21].

Verification procedures have been proposed because traditional secondary criteria for maximal effort are imperfect[19]. In particular, the absence of an oxygen uptake plateau does not necessarily indicate that maximal aerobic capacity was not reached, while the presence of a high RER or high heart rate does not guarantee that the measured VO_2peak reflects the highest physiologically attainable value. The verification phase therefore provides an alternative strategy based on the reproducibility of peak oxygen uptake under a second high-intensity exercise stimulus[21].

However, verification bouts should not be interpreted as a universally mandatory procedure. Their protocol can differ with respect to recovery duration, workload, exercise mode and the criterion used to compare incremental and verification values. Their usefulness may therefore depend on the study population and research objective. In this thesis, the verification bout was used as a study-specific quality-control component rather than as a universal diagnostic test of VO_2max. Participants were classified as potentially not exhausted when the verification-bout VO_2peak exceeded the incremental-test VO_2peak by more than 120 $mL \cdot min^{-1}$ or when peak RER during the incremental test was below 1.10.

Ventilatory thresholds and peak RER were also retained as complementary physiological descriptors. The first ventilatory threshold reflects the transition from lower to progressively greater metabolic demand, whereas the second ventilatory threshold is associated with a further disproportionate increase in ventilation relative to oxygen uptake. These thresholds provide submaximal information that complements VO_2peak and may help describe differences in aerobic performance, exercise tolerance and physiological response to increasing workload. Peak RER, in contrast, provides an indication of the degree of metabolic and ventilatory strain achieved during the test. Together, VO_2peak, ventilatory thresholds, RER, test duration and verification-bout information provide a more complete characterization of exercise performance than any individual variable alone[20].

Overall, CPET provides the physiological framework of the study. It defines the exercise stimulus, provides variables related to cardiorespiratory fitness and allows the metabolomic response to be interpreted in relation to exercise intensity and participant characteristics.

2.2 Acute Exercise Metabolism and Circulating Metabolites

Acute exercise causes rapid changes in energy metabolism [1–6]. During intense exercise, skeletal muscle requires more ATP, and the body adapts by modifying the use of

carbohydrates, fats and other substrates. These changes can be observed indirectly through circulating metabolites, which provide information about the systemic response to exercise.

Lactate is one of the most well-known markers of intense exercise. When exercise intensity increases, glycolytic metabolism becomes more active and lactate concentration in blood usually rises. For this reason, lactate can be considered a useful positive marker of a strong exercise stimulus. In the context of this thesis, a clear post-exercise increase in lactate is expected and helps confirm that the metabolomic data capture a physiological response to maximal exercise [22].

Amino acids and amino-acid-related metabolites are also relevant in exercise metabolism. Their blood concentrations can change after exercise because of muscle activity, protein turnover, tissue exchange and changes in energy metabolism. Compared with lactate, their response is usually more complex. Some amino acids may increase, while others may decrease, depending on the balance between release, uptake and utilization. This makes amino-acid-related metabolites particularly interesting for studying individual differences in the response to exercise [23].

Fatty acids represent another important group of metabolites. They are linked to lipid mobilization and substrate availability. During exercise, the body can release fatty acids into the blood and use them as an energy source, especially at lower and moderate intensities. During maximal exercise, carbohydrate metabolism becomes more dominant, but fatty acids may still provide useful information about differences between participants, especially in relation to fasting status, body composition and metabolic state [24–29].

In this thesis, the analysed metabolite panel includes lactate, amino acids, amino-acid-related metabolites and fatty acids. These metabolite classes were selected because they are directly connected to exercise metabolism and can provide complementary information. Lactate reflects the glycolytic response to intense exercise, amino-acid-related metabolites describe broader changes in systemic metabolism, and fatty acids provide information about lipid-related variation. Together, they allow the acute response to maximal exercise to be studied from a molecular point of view.

However, interpreting these metabolites is not straightforward. A metabolite may change after exercise, but the change must be compared with its normal variability at rest. This is one of the main reasons why the repeated-baseline design is important in this thesis. It allows post-exercise changes to be interpreted not only as group-level differences, but also in relation to resting variability.

2.3 Metabolomics in Exercise Physiology

Metabolomics is the study of small molecules, called metabolites, that are present in biological samples such as blood, urine or tissue. These molecules are involved in many physiological processes, including energy production, substrate use, cellular signalling and recovery. Since metabolites can change rapidly in response to external stimuli, metabolomics is particularly useful for studying acute physiological conditions such as physical exercise [1–7,30].

In exercise physiology, metabolomics allows researchers to observe how the body responds to an exercise stimulus at the molecular level. Instead of focusing on a single marker, such as lactate, metabolomics makes it possible to analyse several metabolites at the same time and to describe the response as a global metabolic profile. This is useful because exercise does not affect only one

pathway, but produces coordinated changes in carbohydrate metabolism, amino acid metabolism, lipid metabolism and other biological processes [1–6].

There are two main approaches in metabolomics: untargeted and targeted metabolomics. Untargeted metabolomics aims to measure as many metabolites as possible and is mainly used for discovery. Targeted metabolomics, instead, focuses on a predefined panel of metabolites that are measured with higher specificity and usually with better comparability across samples [7,8]. In this thesis, the targeted approach is particularly suitable because the analysed metabolites belong to classes that are directly related to exercise metabolism, including amino acids, amino-acid-related compounds, fatty acids and lactate.

2.3.1 Targeted and Untargeted Metabolomics

Targeted and untargeted metabolomics answer different research questions[7,8]. Untargeted metabolomics aims to detect a broad range of measurable molecular features without restricting the analysis to a predefined list of compounds. This approach is particularly useful for discovery-oriented studies because it can reveal previously unrecognized metabolic signals. However, the large number of detected features may create challenges related to metabolite identification, multiple testing, analytical reproducibility and the interpretation of unknown compounds[7,8].

Targeted metabolomics instead focuses on a predefined panel of metabolites selected according to analytical reliability and biological relevance. This strategy typically provides more standardized measurement conditions, more direct compound annotation and a clearer connection between the observed metabolites and known biochemical pathways. Depending on the analytical platform and validation procedures, targeted assays may provide absolute or semi-quantitative concentration estimates and may include internal quality-control procedures, calibration standards and predefined limits of detection[8].

The two approaches should therefore not be considered alternatives in terms of quality. Untargeted metabolomics is advantageous when the aim is broad discovery, whereas targeted metabolomics is particularly suitable when the research question concerns a defined physiological system and requires interpretable metabolite-level results. In the present thesis, targeted metabolomics was appropriate because the analysis focused on amino acids, amino acid-related compounds, fatty acids and lactate, which are metabolite classes directly relevant to acute exercise metabolism, nutritional state and substrate utilization.

An additional advantage of targeted metabolomics in repeated-measures studies is that the same defined metabolites can be compared consistently across baseline and post-exercise samples. This is particularly important when the analysis does not only ask whether a metabolite differs between timepoints, but also whether its post-exercise response is large relative to its own short-term baseline variability[8].

2.3.2 Analytical and Pre-Analytical Considerations

Metabolomic measurements are influenced not only by biology, but also by analytical and pre-analytical conditions. Analytical factors include instrument calibration, batch effects, ionization efficiency, signal drift, internal-standard performance and the limit of detection for each analyte. Pre-analytical factors include the sampling matrix, collection method, anticoagulant, blood-to-processing time, temperature, centrifugation procedure, storage duration and freeze-thaw cycles[31,32].

These sources of variability are particularly relevant for blood metabolomics because metabolite concentrations can change after collection. Cellular metabolism may continue ex vivo, and delays in sample processing can alter the concentration of selected metabolites. In addition, the magnitude and

direction of pre-analytical effects are metabolite-specific. A procedure that has limited impact on one metabolite class may affect another class more substantially[31,32].

For this reason, analytical quality control and pre-analytical standardization should be considered part of the biological interpretation of metabolomics data rather than purely technical procedures[31,32]. When a study investigates repeated measurements over a short period of time, the observed difference between two samples may reflect a combination of physiological fluctuation and variation introduced during collection, processing or measurements. The repeated-baseline design used in this thesis does not fully separate these components. Instead, it provides an operational estimate of their combined short-term effect before exercise.

The use of a targeted assay and metabolite-level quality filtering reduced the contribution of unreliable measurements in the present study. Nevertheless, the remaining repeated-baseline variation should not be interpreted as exclusively technical error. It represents the combined uncertainty that is relevant when deciding whether a post-exercise change is sufficiently large to be interpreted as a robust physiological response.

2.3.3 Capillary and Venous Blood as Metabolomic Matrices

Capillary blood sampling offers practical advantages for exercise metabolomics because it requires a small blood volume, can be collected rapidly and is compatible with repeated sampling during the same experimental session. These features are particularly valuable when the physiological response is expected to change rapidly after exercise and when repeated venous blood draws would increase participant burden.

However, capillary and venous blood should not be assumed to be fully interchangeable for every metabolite[9,11–13]. Capillary blood reflects a mixture of arterial, venous and capillary contributions and may also be influenced by local tissue exchange. Differences between capillary and venous concentrations can therefore depend on the analyte, the metabolic condition and the timing of collection. In exercise studies, these differences may be especially relevant because metabolite gradients can emerge rapidly across tissues during and immediately after physical effort.

The objective of the present thesis was not to establish direct equivalence between capillary and venous concentrations. Instead, capillary blood was treated as a practical and biologically informative matrix for repeated within-subject measurements. The repeated-baseline design further strengthened this approach because each participant's post-exercise response could be interpreted relative to their own capillary resting measurements rather than only relative to an external reference population.

This within-subject perspective is important. Even when absolute capillary concentrations differ from venous values for some metabolites, capillary sampling may still be highly informative for longitudinal and repeated-measure designs, provided that collection and processing procedures are sufficiently consistent and that short-term baseline variability is explicitly considered[9–11].

From a data analysis point of view, metabolomics data are high-dimensional and require careful preprocessing. Each participant is described by multiple metabolite concentrations, and these values may differ in scale, distribution and variability. Some metabolites may also contain missing or unreliable values due to analytical limits or quality-control procedures. For this reason, steps such as metabolite filtering, imputation, transformation, scaling and outlier detection are necessary before statistical analysis and modelling [8,33].

Another important aspect is that metabolomics data contain both biological information and variability. Part of the variation reflects meaningful differences between participants or between timepoints, while another may reflect technical variability, sampling conditions or normal biological fluctuation. This is especially important in acute exercise studies, where the objective is to identify

which metabolites truly respond to exercise. A metabolite may appear to change after exercise, but this change must be interpreted in relation to how stable that metabolite is under resting conditions.

For this reason, the metabolomics dataset used in this thesis is not analysed only through standard group comparisons. The repeated-baseline design makes it possible to estimate resting variability and to use this information in the interpretation of the post-exercise response. This is a central point of the thesis: metabolomics is not treated only as a list of metabolites that increase or decrease after exercise, but as a structured dataset in which reliability, variability and individual response patterns must be considered together.

In summary, metabolomics provides a useful bridge between exercise physiology and data science. It captures molecular information related to the acute exercise response, but it also produces complex data that require appropriate statistical and computational methods. In this thesis, metabolomics is used both to describe the biological response to maximal exercise and to build data-driven analyses aimed at understanding variability and predicting cardiorespiratory fitness.

2.4 Capillary Blood Sampling and Repeated Baseline Designs

Blood sampling is one of the most common approaches for studying circulating metabolites. In exercise studies, blood samples can be collected before and after the exercise stimulus in order to observe how the metabolome changes in response to physical effort. Traditionally, many studies use venous blood sampling, which provides sufficient sample volume and is widely used in clinical and research settings. However, venous sampling can be less practical when multiple samples need to be collected within a short time window, especially during exercise protocols.

Capillary blood sampling offers a more practical alternative [9,10,13]. It usually requires only a small blood volume and can be collected through a finger-prick procedure. This makes it suitable for repeated measurements before and after exercise, which is particularly useful because metabolic changes can occur rapidly, and the sampling procedure should be fast, minimally invasive and compatible with the experimental protocol [9,11–13].

However, the use of capillary blood also requires careful evaluation of data reliability. A small sample volume and repeated sampling conditions may introduce variability, and metabolite concentrations can also fluctuate naturally at rest. Therefore, before interpreting post-exercise changes, it is important to verify whether the resting metabolomic profile is sufficiently stable. If baseline measurements are not reproducible, then it becomes difficult to decide whether a post-exercise difference reflects a true physiological response or simply normal baseline fluctuation.

This is where repeated baseline sampling becomes important. Instead of collecting only one pre-exercise sample, two resting baseline samples are collected before the exercise test. The comparison between these two samples provides information about short-term within-subject variability [14,15]. In other words, the repeated baseline gives an estimate of how much each metabolite can vary before the exercise stimulus is applied. This estimate can then be used as a reference when interpreting the post-exercise response. The sampling workflow is conceptually summarized in Figure 3.1.

The advantage of this design is that it improves the interpretation of individual metabolite changes. With a single baseline, a metabolite can be classified as changed after exercise, but it is not possible to know whether this change is large compared with its normal resting variability. With two baseline samples, the post-exercise response can be compared with the baseline fluctuation. This allows the analysis to move beyond the simple question of whether a metabolite changes after exercise, toward a more informative question: whether the change is large enough to exceed resting variability.

Repeated baseline designs are also useful from a data science perspective. They provide an internal estimate of noise or variability directly from the dataset, rather than relying only on external assumptions. This makes it possible to construct noise-aware metrics, such as signal-to-noise ratios,

and to rank metabolites according to the strength of their exercise response relative to baseline fluctuation [14–18]. In this thesis, this concept is central because the repeated baseline is not used only as a control condition, but as an analytical reference for interpreting post-exercise metabolomic responses.

Overall, capillary blood sampling makes repeated measurements feasible in an acute exercise setting, while the repeated baseline design improves the reliability of interpretation. Together, these two elements provide the experimental foundation of this thesis: a practical sampling method combined with a framework that accounts for resting variability before studying the metabolic response to maximal exercise.

2.5 Biological Variability and Noise-Aware Response Analysis

A central challenge in metabolomics is that metabolite concentrations are variable [14,15]. Part of this variability is biologically meaningful, because individuals can differ in fitness level, body composition, nutritional state or metabolic regulation. Another part can depend on technical and experimental factors, such as sampling conditions, analytical measurement error or preprocessing steps. In exercise studies, this variability is particularly important because the objective is to identify which metabolites truly respond to the exercise stimulus [1–6].

2.5.1 Sources of Metabolomic Variability

Metabolomic variability can be conceptualized as the combination of several sources operating at different levels. Analytical variability originates from the measurement process itself and may include instrument drift, calibration differences, batch effects, ionization variability and uncertainty close to the analytical limit of detection. Pre-analytical variability occurs before the analytical measurement and may depend on blood collection, sample handling, processing delay, storage conditions, temperature and other procedural factors[31,32].

A further component is short-term within-subject biological variability. Even in the absence of exercise, metabolite concentrations may change because of recent food intake, hydration status posture, circadian rhythms, stress, sleep, low-intensity physical activity or normal metabolic regulation. This variability is biologically meaningful, but it may complicate the interpretation of acute exercise studies when post-exercise changes are compared with one pre-exercise sample[14,15].

Finally, between-subject variability reflects stable or semi-stable differences among participants. These differences may be associated with age, sex, body composition, habitual physical activity, nutritional patterns, genetic background or other physiological characteristics[15,34]. Between-subject variability is important because it contributes to metabolic individuality, but it should not be confused with the short-term variation that occurs within the same participant before exercise.

In repeated-baseline studies, these sources are not fully separable using only two pre-exercise measurements. The difference between Baseline 1 and Baseline 2 may contain analytical, pre-analytical and short-term biological components at the same time. For this reason, the present thesis uses the term baseline-derived methodological noise as an operational description of the combined short-term variability observed before the exercise stimulus. this terminology avoids interpreting the baseline difference as pure analytical error while still recognizing its relevance for evaluating the reliability of post-exercise responses.

2.5.2 Repeated Baselines as an Operational Uncertainty Estimate

The purpose of the repeated-baseline design is not to eliminate variability, but to quantify part of the uncertainty that would otherwise remain unobserved in a single-baseline design. If only one resting measurement is available, a post-exercise difference can be estimated, but it is not possible to determine whether the magnitude of that difference is clearly larger than the variability expected for the same metabolite under resting conditions[14,15].

Two baseline samples provide a within-visit estimate of short-term fluctuation. This estimate can then be compared with the average magnitude of the post-exercise response. In a conceptual signal-to-noise framework, the signal represents exercise reactivity and the noise represents baseline-derived variability. A large signal relative to noise indicates that the average response is greater than the short-term fluctuation observed before exercise[14,15].

This approach does not prove that a response is physiologically causal, nor does it guarantee that every individual participant responds in the same direction. Instead, it provides an additional level of interpretation. A metabolite can be statistically significant at the group level but show only a modest response relative to baseline fluctuation. Conversely, another metabolite may show a response that is clearly larger than its observed resting variability and may therefore be more suitable for individual-level interpretation.

The repeated-baseline approach is therefore complementary to conventional significance testing. Statistical tests address whether a group-level difference is unlikely to be explained by random sampling variation under a specified model. The signal-to-noise perspective addresses a different question: whether the observed response is large relative to the short-term variability measured within the same experimental design. Both perspectives are useful, but they should not be interpreted as interchangeable[16–18].

2.5.3 Individual Reference Ranges and Responder Interpretation

Repeated baseline information can also support individual-level response classification. A single two-point baseline envelope can be defined as the observed interval between Baseline 1 and Baseline 2 for a given participant and metabolite. If the post-exercise value falls outside this interval, the participant can be described as showing a deviation beyond the directly observed within-visit baseline range.

However, an interval defined only by two observations should not be interpreted as a formal 95% individual reference interval. A statistically defined 95% interval requires an explicit estimation procedure, for example based on cohort-level variance, residual distributions, bootstrap resampling or a hierarchical model. This distinction is important because the observed interval between two repeated baselines is descriptive, whereas a 95% envelope requires a model-based or distribution-based justification.

In this thesis, the baseline-envelope analysis was used as an exploratory individual-level interpretation tool. It was not used to establish a clinical threshold or to define a confirmed responder phenotype. The analysis was instead intended to identify whether post-exercise values frequently fell outside the repeated resting range and whether selected directional patterns were associated with physiological or experimental characteristics.

This distinction is particularly important for responder analysis. When responder groups are defined after observing the data, they may be influenced by regression to the mean, sampling variability and the choice of threshold. Therefore, responder findings should be treated as hypothesis-generating unless they are pre-specified, replicated in an independent cohort and evaluated using an appropriately designed longitudinal protocol.

The most common approach to study an acute exercise response is to compare a pre-exercise sample with a post-exercise sample. If the post-exercise values are statistically different from baseline, the metabolite is considered exercise-responsive. This type of analysis is useful because it identifies group-level changes. For example, a volcano plot can show which metabolites have both a larger fold-change and a statistically significant difference after correction for multiple testing. However, this approach has an important limitation: it does not indicate whether the post-exercise change is large compared with the normal variability of that metabolite at rest.

This distinction is important because statistical significance and individual reliability are not the same concept. A metabolite can be statistically significant at the group level because the average change is consistent across participants, but the magnitude of that change may still be close to the normal baseline fluctuation. On the other hand, another metabolite may show a response that is clearly larger than its resting variability and may therefore be more reliable as an individual-level exercise response marker. For this reason, group-level significance alone is not sufficient to fully interpret acute metabolomic responses [16–18].

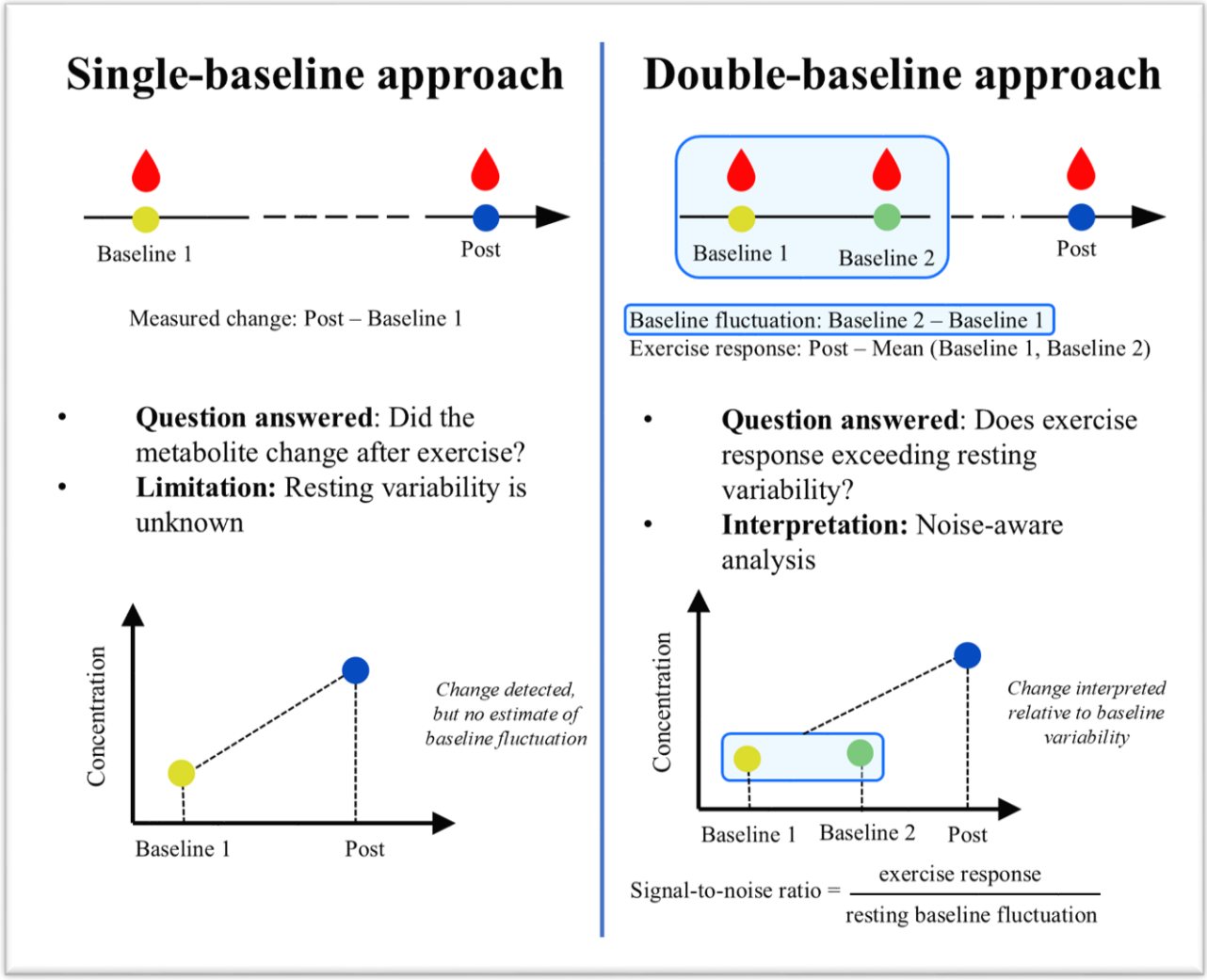


Figure 2.1: Conceptual difference between single-baseline and repeated-baseline response analysis. *A single-baseline comparison can identify post-exercise changes but does not quantify resting variability. A repeated-baseline design estimates baseline fluctuation before exercise, allowing the post-exercise response to be interpreted through a signal-to-noise framework.*

Repeated baseline sampling provides a way to include resting variability directly in the analysis. By comparing Baseline 1 and Baseline 2, it is possible to estimate how much each metabolite fluctuates before exercise. This baseline fluctuation can then be used as a reference for interpreting the post-exercise change [14,15]. In this way, the analysis becomes noise-aware: the response is not evaluated only as a difference between pre- and post-exercise values, but as the difference between exercise reactivity and resting variability.

In the context of this thesis, the term methodological noise is used to describe the baseline-derived variability estimated from repeated resting samples. This estimate includes analytical and pre-analytical variability, but also short-term resting biological fluctuation.

One useful metric for this purpose is the signal-to-noise ratio. In this context, the signal represents the magnitude of the exercise-induced response, while the noise represents the variability observed between the two baseline samples. A signal-to-noise ratio close to 1 means that the average post-exercise change is similar to the resting fluctuations. Higher values indicate that the exercise response is larger than baseline variability. Therefore, SNR can be used to rank metabolites according to how clearly their response exceeds resting noise.

This type of analysis changes the interpretation of the results. A conventional single-baseline analysis can identify a broad set of metabolites that change significantly after exercise, while a noise-aware analysis can show which of these responses are more robust relative to baseline variability. In the study used for this thesis, the single-baseline analysis identified many significant post-exercise changes, but the repeated-baseline analysis allowed these responses to be ranked according to their magnitude relative to resting fluctuation. This is one of the main methodological motivations of the thesis.

Noise-aware analysis is also useful because it connects group-level and individual-level interpretation. The same principle can also support exploratory responder analyses, in which participants are classified according to whether their post-exercise values exceed a baseline-derived reference range. Such analyses can help identify whether stronger individual deviations are associated with physiological or experimental characteristics [16–18]. At the group level, it helps prioritize metabolites whose responses are stronger than resting variability. At the individual level, it provides a more cautious framework for interpreting whether a participant's post-exercise value is meaningfully different from their own baseline state. This is particularly relevant when studying heterogeneous responses, where some participants may show stronger or weaker metabolic changes than others.

In this thesis, noise-aware response analysis is used to move beyond a simple list of metabolites that increase or decrease after exercise. The aim is to identify which metabolites show responses that are statistically detectable, biologically plausible and large enough to be interpreted relative to baseline variability. This approach is important for a data science perspective because it treats the repeated baseline as information that can improve the reliability of downstream analyses, including participant-level modelling and prediction.

2.6 Multivariate and Predictive Modelling in Metabolomic Data

Metabolomics datasets are naturally multivariate. Each participant is described by many metabolites, and each metabolite may contain only a small part of the overall biological information. For this reason, analysing one metabolite at a time is useful but not sufficient. A complete interpretation requires methods that can describe the structure of the dataset as a whole, identify patterns across metabolites and relate these patterns to participant-level characteristics.

Exploratory multivariate methods are commonly used as a first step. Principal component analysis reduces the dimensionality of the data by creating new variables, called principal components, that summarize the main sources of variation. In a metabolomics study, PCA can show whether samples collected at different timepoints have different global metabolic profiles. It can also help identify which metabolite classes contribute most to the separation between groups or conditions. For example, if post-exercise samples separate from baseline samples along the first principal component, this suggests that exercise produces a structured shift in the overall metabolome.

Other visualization methods, such as t-SNE, can be useful to explore sample similarity in a lower-dimensional space [35]. While PCA focuses on linear sources of variation, t-SNE is often used to visualize local relationships between samples. In repeated-measure designs, this can help assess whether samples from the same participant remain close to each other, suggesting the presence of stable individual metabolic profiles. These visualizations are exploratory and should not be interpreted

as formal statistical proof, but they are useful for understanding the structure of complex metabolomic data.

Clustering methods can also be applied to identify groups of participants with similar response patterns. In the context of exercise metabolomics, clustering can be performed on post-exercise fold-changes or noise-adjusted responses to explore whether participants show different metabolic response profiles. However, clustering results must be interpreted carefully. The presence of two or more clusters does not automatically mean that there are clearly separated biological subtypes. Cluster robustness, effect size and association with participant characteristics should be evaluated before assigning biological meaning to the groups.

In addition to exploratory methods, statistical models can be used to test associations between metabolite responses and participant-level variables. For example, linear models can evaluate whether VO_2peak, body composition, age, sex, fasting status or exhaustion status are associated with the magnitude of the metabolite response after exercise. This approach is useful because it connects the metabolomic data with physiological and experimental characteristics. However, because many metabolites and predictors may be tested, multiple-testing correction is necessary to reduce the risk of false positive findings [36].

A further data science objective is predictive modelling. Instead of only asking whether a variable is associated with a metabolite response, predictive modelling asks whether a set of features can estimate an outcome of interest. In this thesis, the outcome is VO_2peak, which represents cardiorespiratory fitness. The predictors may include baseline metabolite concentrations, anthropometric variables and physiological characteristics. By comparing different feature sets, it is possible to evaluate whether metabolomic information adds predictive value beyond standard participant-level variables.

Predictive modelling in metabolomics requires particular caution because the number of available participants is often limited compared with the number of potential features. This increases the risk of overfitting, meaning that a model may perform well on the training data but poorly on new data. To reduce this risk, model performance should be evaluated using appropriate validation strategies, such as cross-validation. Performance metrics such as R^2, mean absolute error (MAE) and root mean squared error can be used to quantify how accurately the model predicts the target variable.

Interpretability is also important. A predictive model is more useful if it not only produces accurate predictions, but also helps identify which variables contribute most to the prediction. Depending on the model, this can be assessed through regression coefficients, feature importance scores or other interpretability methods. In biomedical applications, this step is essential because the goal is not only to build a model but also to understand whether the selected features have plausible physiological meaning.

2.6.1 Association, Explanation and Prediction

Association analysis and predictive modelling answer related but distinct question. Association analysis examines whether a variable is statistically related to another variable within the observed dataset. For example, a linear model can test whether fasting status is associated with the magnitude of a metabolite response. The main objective is interpretability: identifying variables that may help explain observed variation[37].

Prediction asks a different question. Instead of testing whether a relationship exists, it evaluates whether a set of input variables can estimate an outcome for new observations that were not used during model training. A variable may show a statistically significant association with an outcome but still provide limited predictive value when evaluated in unseen participants. Conversely, a variable may contribute to prediction even when its isolated association is modest, particularly when it provides complementary information in combination with other features[37].

This distinction is important in metabolomics because many metabolites may be biologically relevant without being useful predictors of a specific target. In the present thesis, the determinant analyses examined associations between participant characteristics and noise-adjusted exercise responses, whereas the prediction analysis evaluated whether baseline metabolomic and participant-level features could estimate VO_2peak. The two analyses were therefore complementary but should not be interpreted as equivalent.

2.6.2 Small-Sample Biomedical Machine Learning

Biomedical metabolomics often involves a limited number of participants relative to the number of potential predictors. This setting increases the risk of overfitting, in which a model learns idiosyncratic patterns of the available dataset rather than information that generalizes to new participants. Overfitting can lead to apparently strong training performance but weak performance when the model is evaluated on independent data[38].

Regularized regression methods are useful in this context because they constrain model complexity. L1 regularization encourages sparse models by shrinking some coefficients toward zero, whereas L2 regularization stabilizes estimates when predictors are correlated. ElasticNet combines these two penalties and is therefore suitable for metabolomics datasets in which several metabolites may be correlated and the available sample size is limited[39,40].

However, regularization alone is not sufficient to guarantee valid predictive performance. The evaluation procedure must also ensure that information from the validation data does not influence preprocessing, feature selection, hyperparameter tuning or model selection. Otherwise, performance estimates may become optimistically biased[38].

2.6.3 Leakage, Nested Validation and Stability Selection

Data leakage occurs when information from the validation data influences any step of model development. In biomedical prediction studies, leakage can arise if missing-value imputation, scaling, feature selection or hyperparameter tuning are performed using the full dataset before cross-validation. Even when this influence is indirect, it can lead to performance estimates that are better than the model would achieve on genuinely unseen participants[38,41].

Nested cross-validation addresses this issue by separating model development from performance evaluation. The outer loop provides held-out data for estimating out-of-sample performance. The inner loop uses only the outer training data for hyperparameter tuning, feature selection and model comparison. In this way, the validation observations remain independent from all model-development decisions[38,41].

Stability selection provides an additional interpretability layer. Rather than selecting features only once from the full dataset, the method examines how frequently a feature is retained across repeated resampling procedures. A metabolite selected consistently across training resamples may represent a more stable model component than one selected only occasionally[42]. Nevertheless, high selection frequency should not automatically be interpreted as proof of biological importance or as evidence of a validated biomarker. Feature stability indicates consistency within the modelling framework, whereas external validation is required to establish generalizability[38].

For these reasons, predictive performance should be interpreted using multiple metrics. R^2 describes the proportion of variance explained by the predictions, MAE quantifies the average absolute prediction error and RMSE gives greater weight to larger errors. Reporting all three metrics provides a more complete view of model performance, particularly in small datasets where individual prediction errors and fold-to-fold variation can be substantial.

Overall, the data-science component of the present thesis combines exploratory multivariate analysis, association modelling and predictive modelling while maintaining a distinction between interpretation and generalization. This distinction is central to the interpretation of the VO_2peak prediction results, where selected metabolomic features may reveal internal model structure without necessarily providing sufficient incremental predictive value beyond standard participant-level variables[37].

In this thesis, multivariate and predictive methods are used together. Exploratory analyses are used to describe the structure of the metabolomic dataset, clustering is used to investigate response-pattern heterogeneity, statistical models are used to test participant-level determinants of noise-adjusted responses, and predictive models are used to estimate VO_2peak from metabolomic and physiological features. This combination reflects the data science nature of the work: the aim is not only to describe metabolite changes after exercise, but also to extract reliable and interpretable information from complex biomedical data.

3 Materials and Methods

3.1 Study Design and Cohort Overview

This thesis is based on a cross-sectional exercise metabolomics study designed to investigate the relationship between the capillary blood metabolome, acute exercise response and cardiorespiratory fitness. The study combined physiological measurements obtained during cardiopulmonary exercise testing with targeted metabolomics data collected from capillary blood samples. The main objective of the experimental design was to characterize post-exercise metabolite changes while accounting for short-term variability under resting conditions.

The study cohort consisted of 100 healthy adults. The participants included 68 males and 32 females, with a mean age of 26.89 ± 8.06 years and a mean body mass index of 24.30 ± 3.54 $kg \cdot m^{-2}$. During the visit, demographic, anthropometric and physiological information was collected. Body composition was assessed using bioelectrical impedance analysis, allowing relative fat mass and relative fat-free mass to be included among the participant-level variables.

Most participants were tested in a non-fasted state, while a smaller group arrived after an overnight fast. Specifically, 79 participants were non-fasted and 21 participants were fasted at the time of sampling. Fasting status was therefore retained as a participant-level characteristic in the downstream analyses, because nutritional state may influence circulating metabolite concentrations and the metabolic response to exercise.

Each participant underwent a maximal cardiopulmonary exercise test on a cycle ergometer. The protocol was designed to progressively increase exercise intensity until volitional exhaustion. During the test, respiratory gas exchange was measured continuously, allowing the assessment of peak oxygen uptake, ventilatory thresholds, respiratory exchange ratio and other exercise-related variables. These physiological measures were used both to describe the cohort and to investigate whether participant-level characteristics were associated with metabolite responses.

The sampling design included repeated capillary blood measurements before and after the exercise test. Two resting baseline samples were collected before exercise, followed by one immediate post-exercise sample. An additional sample was collected after a verification bout, but this timepoint was not included in the primary metabolomics analyses of this thesis. The main analytical dataset therefore included three timepoints for each participant: Baseline 1, Baseline 2 and Post-exercise.

The repeated-baseline design represents a central element of the study. Instead of using only one pre-exercise measurement as the reference condition, the two baseline samples allowed the estimation of short-term within-subject variability before the exercise stimulus. This information was then used to interpret post-exercise changes relative to resting fluctuation. In this way, the study design supported both conventional single-baseline analyses and noise-aware analyses based on repeated baseline information.

3.1.1 Cohort Characteristics and Analytical Subsets

The full study cohort included 100 participants and was used for the baseline reliability, acute exercise-response and participant-level determinant analyses. The prediction analysis as conducted on a restricted subset of participants classified as exhausted according to the predefined CPET and verification-bout criteria. This restriction was applied because VO_2peak was used as the prediction target, and potentially underestimated values could introduce additional noise into model development and evaluation.

The distinction between the full cohort and the prediction subset is relevant because the analytical objectives differed across chapters. The repeated-baseline and acute response analyses aimed to characterise metabolic variability and exercise-induced perturbation in the complete available cohort. In contrast, the prediction analysis required a target variable that was as physiologically reliable as possible. Cohort characteristics for the full dataset and the prediction subset are summarised in Table 3.1.

Table 1 *Participant characteristics and cardiopulmonary exercise testing outcomes. Values are reported as mean ± standard deviation [minimum, maximum] unless otherwise indicated. Sex and fasting status are reported as counts.*

	Mean ± Standard Deviation [Min, Max]
Demographic and anthropometric characteristics	
Sex, M/F	68/32
Age, yr	26.89 ± 8.06 [18, 63]
Height, cm	176 ± 8 [156, 193]
Weight, kg	75.98 ± 14.17 [44.5, 119.8]
BMI, kg m^{-2}	24.30 ± 3.54 [18.26, 36.18]
Fasting status, N/Y	79/21
Relative fat mass, %	21.92 ± 8.52 [5.73, 49.39]
Relative fat-free mass, %	78.08 ± 8.52 [50.62, 94.27]
Performance characteristics	

VO_2peak, mL min^{-1} kg^{-1}	41.91 ± 8.42 [18, 65]
VO_2 at VT1, mL min^{-1} kg^{-1}	23.07 ± 6.85 [7, 40]
VO_2 at VT2, mL min^{-1} kg^{-1}	36.56 ± 8.27 [15, 62]
Duration, s	561.24 ± 123.10 [275, 829]
RER at VO_2peak	1.15 ± 0.05 [1.01, 1.33]

Abbreviations: BMI, body mass index; F, female; M, male; N, no; RER, respiratory exchange ratio; SD, standard deviation; VO_2peak, peak oxygen uptake; VT1, first ventilatory threshold; VT2, second ventilatory threshold; Y, yes

Overall, the dataset combined three main types of information: participant-level characteristics, physiological variables obtained from CPET and targeted metabolomics measurements from capillary blood. This structure made it possible to study the data from different perspectives: baseline reliability, acute exercise-induced metabolic changes, participant-level determinants of metabolite responses and prediction of VO_2peak.

3.2 Cardiopulmonary Exercise Testing Protocol

All participants performed a maximal cardiopulmonary exercise test on a cycle ergometer. The test was conducted using an individualized starting workload of either 20 W or 50 W, followed by a fixed ramp increase of 25 W min^{-1} until volitional exhaustion. This incremental protocol was used to progressively increase the physiological demand and to assess the participant's maximal aerobic response.

Respiratory gas exchange was measured continuously during the test using breath-by-breath spirometry. This allowed the assessment of oxygen uptake, carbon dioxide production and ventilation throughout the exercise protocol. From these measurements, peak oxygen uptake was obtained and used as the main indicator of cardiorespiratory fitness. VO_2peak was expressed relative to body mass as mL min^{-1} kg^{-1}.

In addition to VO_2peak, other CPET-derived variables were retained for analysis. These included the first and second ventilatory thresholds, peak respiratory exchange ratio and total test duration. Ventilatory thresholds were used as submaximal indicators of the physiological response to increasing exercise intensity, while peak RER provided information about the intensity of the effort reached during the test. Total test duration was included as an additional descriptor of exercise performance.

The verification bout was included in the study protocol as a physiological quality-control step. Its purpose was to evaluate whether the VO_2peak achieved during the incremental CPET was consistent with the highest oxygen uptake reached during a subsequent verification phase. Agreement between the incremental CPET and the verification bout was defined using a difference within ±120 mL·min^{-1}. The post-verification blood sample was not included in the primary metabolomics analyses, because the present thesis focuses on resting variability and the immediate post-exercise response to the incremental CPET.

Information from the verification bout was instead used to define exhaustion status. Participants were classified as potentially not exhausted if the VO_2peak measured during the verification bout exceeded the VO_2peak obtained during the incremental CPET by more than 120 mL min^{-1} or if peak RER during the incremental CPET was below 1.10. This variable was later used in participant-level determinant analyses, because incomplete attainment of maximal effort may attenuate the physiological and metabolomic response to exercise.

Overall, the CPET protocol provided both the standardized exercise stimulus and the physiological variables required for downstream analyses. VO_2peak was used as the main marker of cardiorespiratory fitness, while RER, ventilatory thresholds, test duration and exhaustion status were used to interpret inter-individual variability in the metabolomic response.

3.3 Capillary Blood Sampling and Measurement Timepoints

Capillary blood samples were collected at four timepoints during the experimental session. The first two samples were collected under resting conditions before the exercise test and were defined as Baseline 1 and Baseline 2. These repeated baseline samples were used to estimate short-term within-subject variability before the exercise stimulus. A third sample was collected immediately after the maximal CPET and was defined as the Post-exercise sample. A fourth sample was collected after the verification bout.

The primary metabolomics analyses in this thesis focused on the two resting baseline samples and the immediate post-exercise sample. Therefore, the main analytical dataset included three timepoints for each participant: Baseline 1, Baseline 2 and Post-exercise. Since the cohort included 100 participants, this resulted in 300 samples for the primary analyses.

The post-verification sample was not included in the main metabolomics analyses. This choice was made because the main objective of the thesis was to evaluate resting variability and the immediate metabolic response to the maximal incremental CPET. Including the post-verification sample would have introduced a second exercise-related timepoint with a different physiological meaning. However, information from the verification bout was still used to evaluate exhaustion status, as described in the previous section.

Capillary blood was collected using a minimally invasive sampling procedure suitable for repeated measurements within the same experimental session. This sampling approach allowed the collection of multiple blood samples with limited participant burden and made it possible to obtain repeated baseline measurements before exercise.

The repeated-baseline structure was essential for the downstream analyses. Baseline 1 was used as the reference timepoint for conventional single-baseline analyses, such as post-exercise fold-change and volcano plot analysis. Baseline 1 and Baseline 2 together were used to estimate baseline fluctuation and to calculate noise-aware response metrics. The Post-exercise sample was then interpreted both as a conventional post-exercise measurement and as a response relative to resting variability.

Overall, the sampling design provided the basis for two complementary analytical perspectives. The first perspective compared the immediate post-exercise sample with a single baseline, reproducing the standard approach used in many acute exercise metabolomics studies. The second perspective used the repeated baseline to evaluate whether post-exercise metabolite changes exceeded the variability observed before exercise.

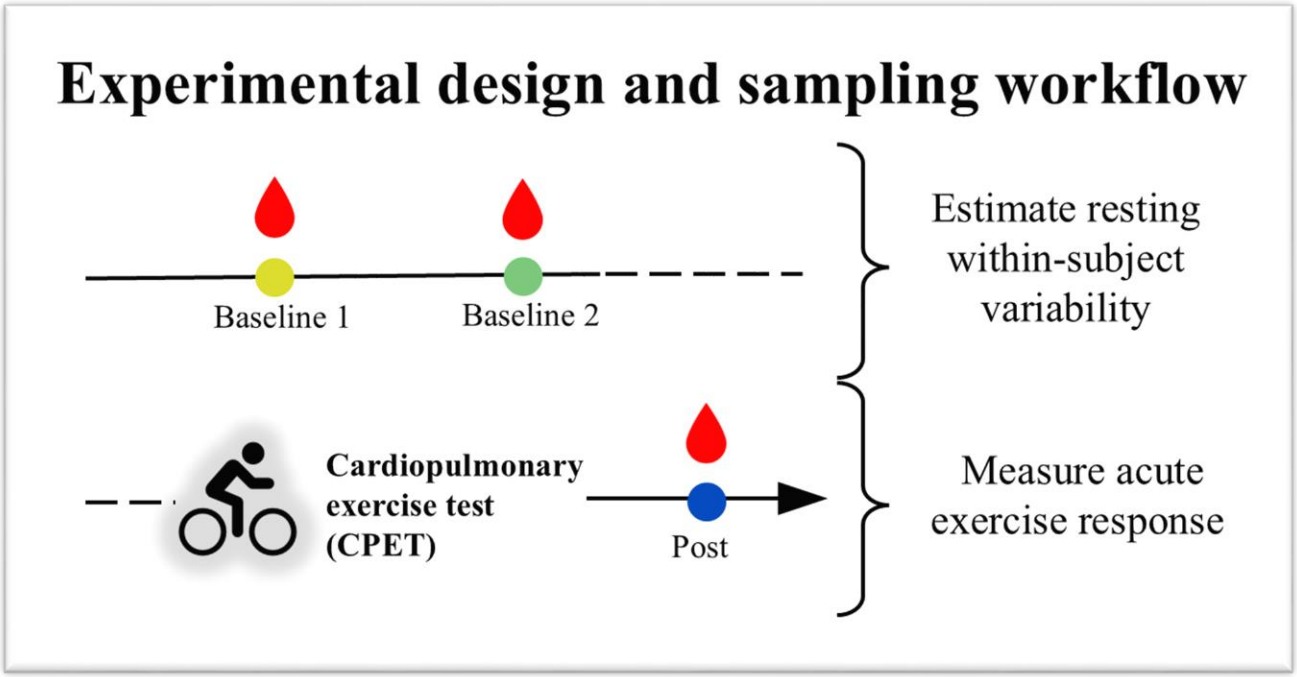


***Figure 3.1: Repeated-baseline capillary blood sampling design**. Two resting capillary blood samples are collected before the maximal cardiopulmonary exercise test to estimate short-term within-subject variability. The immediate post-exercise sample is then interpreted in relation to both the exercise stimulus and the resting baseline fluctuation.*

3.4 Targeted Metabolomics Panel and Preprocessing

Targeted metabolomics analysis was performed on capillary blood samples to quantify a predefined panel of metabolites related to exercise metabolism [7,8]. The analytical platform used for the study was the MxP Quant 500 assay, which allows the measurement of multiple metabolite classes in small-volume blood samples. The analysis focused on metabolites with direct relevance to acute exercise physiology, including lactate, amino acids, amino-acid-related compounds and fatty acids.

The initial targeted metabolomics panel included 62 metabolites. Before downstream statistical analysis, metabolite-level filtering was applied to retain only analytes with sufficient analytical quality and data availability. This step was necessary because not all measured metabolites were consistently detected across samples or above the analytical limit of detection.

A project-specific consensus limit of detection was used to evaluate metabolite reliability. This consensus threshold was estimated from the limits of detection obtained across the analytical plates. Metabolites were retained if at least 80% of the values were valid and above the limit of detection within the relevant experimental groups, following a modified 80% rule. This criterion was applied to reduce the influence of poorly detected metabolites and to ensure that downstream analyses were based on a reliable metabolite panel.

After filtering, 10 metabolites were excluded from the dataset. These excluded metabolites included seven amino-acid-related compounds and three fatty acids. The final retained panel consisted of 52 metabolites. These included 20 amino acids, 23 amino-acid-related metabolites, 8 fatty acids and lactate.

The retained metabolite panel provided the basis for all subsequent analyses. Amino acids and amino-acid-related metabolites were used to investigate systemic changes related to acute exercise and participant-level variability. Fatty acids were included as markers of lipid-related metabolic variation, while lactate was used as an expected marker of the glycolytic response to intense exercise. Therefore, the final panel was suitable for studying both the global post-exercise metabolic shift and the reliability of individual metabolite responses relative to baseline variability.

Overall, metabolite filtering reduced the initial panel to a smaller but more reliable set of analytes. This step was important because the objective of the thesis was not only to detect post-exercise differences, but also to compare these responses with resting variability. For this reason, the quality and consistency of each retained metabolite were essential for the validity of the repeated-baseline and signal-to-noise analyses.

3.5 Data Quality Control and Transformation

Before statistical analysis, the metabolomics dataset was processed to improve data quality, reduce the influence of unreliable observations and make metabolite values comparable across samples. The preprocessing pipeline included handling of values below the limit of detection, integration with participant metadata, transformation of metabolite concentrations, outlier detection and imputation for selected exploratory analyses.

For retained metabolites, values below the limit of detection were replaced using the logspline-based imputation procedure available in the preprocessed dataset. This step allowed low-abundance values to be handled without removing entire metabolites or samples from the analysis. After this step, metabolite concentrations were merged with participant-level metadata, including participant ID, sampling timepoint, sex, age, body mass index, fasting status and plate information.

Metabolite concentration values were $\log_2$-transformed before downstream statistical analyses. This transformation was applied to reduce right-skewness and to make the distribution of metabolite values more suitable for statistical modelling. Log-transformation is commonly used in metabolomics because concentration values can span different orders of magnitude and may show heteroscedasticity. After transformation, changes between timepoints could be interpreted as $\log_2$ differences, which are closely related to fold-change values.

For multivariate analyses, including principal component analysis and t-SNE, $\log_2$-transformed metabolite values were additionally standardized. Standardization was necessary because metabolites can differ substantially in absolute concentration and variance. Without scaling, metabolites with larger numerical ranges could dominate the multivariate structure, even if they are not biologically more informative. Therefore, z-standardization was used to place metabolites on a comparable scale before exploratory dimensionality reduction.

Additional quality control was performed at the metabolite-value level. Extreme observations were identified using standardized z-scores, and values with an absolute z-score greater than 3.5 were flagged as potential outliers. In total, 37 metabolite-level observations were removed from the non-imputed dataset when appropriate. This outlier filtering step was applied to reduce the influence of extreme values that could distort statistical comparisons or response estimates.

The main univariate and repeated-measures analyses were performed on the outlier-filtered dataset without applying missForest imputation. This choice was made to limit the influence of model-based imputation on the primary statistical results. In contrast, some exploratory multivariate analyses required a complete data matrix. For these analyses, missing values introduced by outlier removal were imputed separately using missForest [33]. This additional imputation step was therefore used only for exploratory analyses such as PCA and t-SNE, and not for the main statistical testing.

This distinction between primary and exploratory analyses was important for the reliability of the workflow. Primary analyses were based as much as possible on observed and quality-controlled data, while imputation was used only when required by methods that cannot handle missing values. Overall, the preprocessing procedure aimed to preserve the biological structure of the dataset while reducing technical noise and ensuring that the data were suitable for both statistical analysis and machine learning methods.

3.6 Baseline Reliability Analysis

Baseline reliability was assessed by comparing the two resting samples collected before exercise, Baseline 1 and Baseline 2. The aim of this analysis was to determine whether the capillary blood metabolome was sufficiently stable under resting conditions to be used as a reference for the interpretation of post-exercise responses. Since the main objective of the thesis was to compare exercise-induced changes with resting variability, evaluating baseline reproducibility was a necessary first step.

Intra-individual variability was calculated to quantify how much metabolite concentrations changed within the same participant between the two baseline samples. For each participant and metabolite, the difference between Baseline 1 and Baseline 2 was computed on $\log_2$-transformed values. Absolute baseline differences were then used to summarize short-term within-subject variability across the metabolite panel. This analysis provided an estimate of the resting fluctuation observed before the exercise stimulus.

Inter-individual variability was also evaluated to describe how much metabolite concentrations differed between participants at each baseline timepoint. For this purpose, coefficients of variation were calculated separately for Baseline 1 and Baseline 2. Comparing these values across the two baseline samples helped assess whether the overall distribution of metabolite concentrations remained consistent under resting conditions.

Baseline agreement was further assessed using Pearson correlation. First, an overall correlation was calculated by comparing $\log_2$-transformed metabolite values between Baseline 1 and Baseline 2 across all participant-metabolite pairs. This provided a global measure of agreement between the two repeated baseline measurements. In addition, metabolite-level Pearson correlations were calculated across participants to evaluate whether individual metabolites showed consistent ranking of participants between the two baseline samples.

To test whether any metabolite showed a systematic difference between Baseline 1 and Baseline 2, paired statistical tests were performed for each metabolite. The resulting p-values were adjusted for multiple comparisons using the Benjamini-Hochberg false discovery rate procedure [36]. This analysis was used to identify potential baseline drift or order effects between the two resting samples.

Individual metabolic profiles were also explored using dimensionality reduction. In particular, t-SNE was applied to the standardized $\log_2$-transformed metabolite matrix to visualize whether repeated samples from the same participant remained close in the reduced-dimensional space [35]. This analysis was used as an exploratory assessment of metabolic individuality, or the tendency of participants to preserve a recognizable metabolic profile across repeated measurements.

Additional reliability analyses were used to complement the main baseline comparison. Bland-Altman plots were used for representative metabolites to visualize agreement between the two baseline measurements and to detect possible systematic bias. Variance decomposition was used to describe the relative contribution of between-subject variability and within-subject baseline variability for selected metabolites. These analyses provided further information on whether the repeated baseline design captured stable individual profiles or substantial short-term fluctuation.

Overall, the baseline reliability analysis was designed to evaluate the quality and stability of the resting metabolomic reference state. The results of this step provided the methodological basis for the later signal-to-noise analysis, where post-exercise responses were interpreted relative to the variability estimated from the two baseline samples.

3.7 Single-Baseline and Noise-Adjusted Exercise Response Analysis

The post-exercise metabolomic response was analysed using two complementary approaches. First, a conventional single-baseline analysis was performed by comparing the immediate post-exercise

sample with Baseline 1. Second, a noise-adjusted analysis was performed by incorporating both baseline samples to evaluate whether the post-exercise response exceeded resting within-subject variability.

For the single-baseline analysis, $\log_2$ fold-change values were calculated for each participant and metabolite as the difference between the Post-exercise value and Baseline 1. These values were used to describe the conventional exercise-induced response. This approach is similar to the standard design used in many acute exercise metabolomics studies, where one pre-exercise sample is compared with one post-exercise sample.

The single-baseline fold-change matrix was used for two main analyses. First, hierarchical clustering was applied to explore whether participants showed different post-exercise response patterns across the metabolite panel. Clustering was performed on the participant-by-metabolite matrix of $\log_2$ fold-changes. This allowed the identification of groups of participants with similar global response profiles. Because clustering is exploratory, cluster robustness was later evaluated using complementary approaches, including bootstrap consensus clustering, permutation testing of silhouette scores and comparison with alternative clustering methods.

Second, metabolite-level post-exercise changes were evaluated using statistical testing and volcano plot analysis. For each metabolite, post-exercise values were compared with Baseline 1 to identify metabolites showing significant changes after exercise. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg false discovery rate procedure [36]. The volcano plot was used to visualize both the magnitude of the change and its statistical significance, providing a group-level overview of the acute exercise response.

Although the single-baseline analysis is useful for identifying metabolites that change after exercise, it does not quantify whether these changes are large compared with resting variability. For this reason, a noise-adjusted analysis was performed using the repeated-baseline information. Exercise reactivity was defined as the mean absolute change between the immediate post-exercise sample and the mean of Baseline 1 and Baseline 2. Methodological noise was estimated from the variability between Baseline 1 and Baseline 2.

In this thesis, methodological noise refers to the short-term variability estimated from the difference between Baseline 1 and Baseline 2. This baseline-derived estimate may include contributions from sample collection, sample preparation, analytical measurement and short-term resting biological fluctuation, which cannot be fully separated in the present design. It is therefore used as an operational estimate of the noise against which post-exercise metabolite responses are interpreted.

To make the signal-to-noise procedure explicit, let ($x_{i,m}^{B1}$, $x_{i,m}^{B2}$) and $x_{i,m}^{P}$ denote the $\log_2$-transformed concentration of metabolite (m) for participant (i) at Baseline 1, Baseline 2 and Post-exercise, respectively. The participant-specific resting reference value was defined as:

$$\bar{x}_{i,m}^{B} = \frac{x_{i,m}^{B1} + x_{i,m}^{B2}}{2}$$

Exercise reactivity for metabolite (m) was then calculated as the mean absolute deviation of the post-exercise value from the participant-specific baseline reference:

$$R_m = \frac{1}{n_m} \sum_{i=1}^{n_m} \left| x_{i,m}^{P} - \bar{x}_{i,m}^{B} \right|$$

where n_m represents the number of valid observations available for metabolite (m) after quality-control procedures.

Baseline-derived methodological noise was estimated as the mean absolute difference between the two repeated resting measurements:

$$N_m = \frac{1}{n_m} \sum_{i=1}^{n_m} \left| x_{i,m}^{B2} - x_{i,m}^{B1} \right|$$

The signal-to-noise ratio was finally defined as:

$$SNR_m = \frac{R_m}{N_m}$$

Because all values were analysed on the $\log_2$ scale, both exercise reactivity and baseline-derived methodological noise reflect average absolute $\log_2$ differences. An SNR equal to 1 indicates that the mean post-exercise deviation was similar in magnitude to the short-term fluctuation observed between the two resting samples. Values greater than 1 indicate progressively stronger exercise reactivity relative to baseline-derived methodological noise.

For each metabolite, a signal-to-noise ratio was calculated as the ratio between exercise reactivity and resting fluctuation. In this framework, the signal represents the magnitude of the post-exercise response, while the noise represents the short-term baseline variability observed before exercise. An SNR close to 1 indicates that the average exercise response is similar in magnitude to resting fluctuation. Higher SNR values indicate that the response is larger than baseline variability and therefore more reliable from an individual-level interpretation perspective.

Metabolites were ranked according to their SNR values. This ranking allowed exercise-responsive metabolites to be stratified into different response tiers, from responses close to resting variability to responses clearly exceeding baseline fluctuation. In this thesis, this analysis was used to compare the information obtained from the conventional single-baseline approach with the additional interpretation provided by the repeated-baseline design.

Overall, the combination of single-baseline and noise-adjusted analyses allowed the post-exercise metabolomic response to be evaluated from two perspectives. The single-baseline analysis identified group-level changes after exercise, while the signal-to-noise analysis assessed whether these changes were large relative to resting within-subject variability. This distinction is central to the methodological contribution of the thesis.

3.8 Participant-Level Determinant Analysis

Participant-level determinant analysis was performed to investigate whether inter-individual variability in the post-exercise metabolomic response could be explained by physiological, anthropometric or experimental characteristics. This analysis focused on metabolites whose exercise response exceeded resting variability, because these metabolites were more informative for studying individual differences in response magnitude.

For this purpose, only metabolites with a signal-to-noise ratio greater than 1 were included. An SNR greater than 1 indicates that the average post-exercise response was larger than the estimated resting baseline fluctuation. This filtering step reduced the analysis to metabolites with responses that were more likely to reflect exercise-related changes rather than baseline variability.

For each selected metabolite, a continuous noise-adjusted response value was used as the outcome variable. This response represented the magnitude of the post-exercise change relative to the baseline variability estimated from the two repeated baseline samples. Using a continuous response variable allowed the analysis to evaluate not only whether a metabolite changed after exercise, but also whether the strength of this change differed between participants.

Six participant-level characteristics were tested as predictors: VO_2peak, relative fat mass, age, sex, fasting status and exhaustion status. VO_2peak was included as the main marker of cardiorespiratory fitness. Relative fat mass was included as an indicator of body composition. Age and sex were

included as demographic variables. Fasting status was included because nutritional state may influence circulating metabolites and substrate availability before exercise. Exhaustion status was included because incomplete attainment of maximal effort may attenuate the physiological and metabolic response to CPET.

Exhaustion status was defined using two criteria based on the CPET and verification-bout data. Participants were classified as potentially not exhausted if the VO_2peak measured during the verification bout exceeded the VO_2peak obtained during the incremental CPET by more than 120 $mL \cdot min^{-1}$, or if peak respiratory exchange ratio during the incremental CPET was below 1.10. Participants not meeting these criteria were classified as exhausted. This variable was used to account for differences in the effective intensity of the exercise stimulus.

Univariable linear models were fitted separately for each metabolite-predictor combination. Continuous predictors were standardized before modelling, while categorical predictors were encoded as binary variables. This approach allowed the association between each participant characteristic and each metabolite response to be evaluated independently. The use of univariable models was chosen to provide an interpretable overview of the relationship between participant characteristics and noise-adjusted metabolite responses.

Because multiple metabolites and multiple predictors were tested, p-values were adjusted using the Benjamini-Hochberg false discovery rate procedure [36]. Correction was applied to reduce the risk of false positive associations due to multiple testing. Associations passing the correction threshold were considered as candidate determinants of inter-individual variability in the metabolomic response.

For participant characteristics that showed relevant associations, follow-up group comparisons were performed. In particular, metabolites associated with exhaustion status or fasting status were compared between the corresponding groups using appropriate statistical tests. These comparisons were used to visualize and interpret how the noise-adjusted response differed between exhausted and not-exhausted participants, or between fasted and non-fasted participants.

Overall, participant-level determinant analysis was used to connect the metabolomic response with physiological and experimental variables. This step helped evaluate whether inter-individual differences in metabolite responses were related to measurable participant characteristics or whether substantial variability remained unexplained.

3.9 Directional Response, Baseline-Envelope and Exploratory Responder Analysis

Directional response analysis was performed as a complementary approach to investigate whether individual participants showed increases, decreases or unchanged metabolite values after exercise. While the signal-to-noise analysis quantified the magnitude of the response at the metabolite level, the baseline-envelope analysis focused on the direction of the response at the individual level.

For each participant and metabolite, the individual baseline envelope was defined as the interval between Baseline 1 and Baseline 2. The post-exercise value was then compared with this interval. If the post-exercise value was above the baseline envelope, the participant was classified as showing an increase for that metabolite. If the post-exercise value was below the baseline envelope, the participant was classified as showing a decrease. If the post-exercise value remained within the interval defined by the two baseline samples, the response was classified as unchanged.

This approach allowed post-exercise responses to be interpreted relative to each participant's own resting range, rather than only relative to the group average. It was particularly useful for identifying metabolites with consistent directional responses and metabolites with mixed response directions. A consistent directional response indicates that most participants changed in the same direction after exercise, while a mixed response suggests that some participants increased and others decreased relative to their own baseline envelope.

The directional analysis was first applied to metabolites highlighted by the participant-level determinant analysis, especially those associated with exhaustion status or fasting status. This step was used to evaluate whether these metabolites showed mostly unidirectional patterns or whether the associated differences reflected more heterogeneous individual response directions.

As a complementary exploratory analysis, the baseline-envelope approach was extended to all metabolites with signal-to-noise ratio greater than 1. Among these metabolites, bidirectional response patterns were defined as cases where at least 25 participants showed an increase and at least 25 participants showed a decrease beyond the individual baseline envelope. This criterion was used to identify metabolites with sufficiently represented response directions on both sides of the baseline range.

For metabolites showing bidirectional response patterns, associations between response direction and participant characteristics were evaluated. Continuous participant characteristics were compared between directional groups using non-parametric tests, while categorical characteristics were analysed using Fisher's exact test. Effect estimates were reported with 95% bootstrap confidence intervals when appropriate. These analyses were used to explore whether mixed response directions could be partly explained by variables such as fasting status, sex, body composition or cardiorespiratory fitness [16–18].

The baseline-envelope analysis was therefore not used as the primary method to define exercise-responsive metabolites, but as an additional individual-level interpretation tool. Its purpose was to complement the continuous noise-adjusted response analysis by showing whether metabolite changes occurred in a consistent or heterogeneous direction across participants.

As an additional exploratory analysis, responder patterns were investigated for selected metabolites using post-exercise deviations from baseline-derived reference ranges. This analysis focused on metabolites with robust exercise responses, defined by SNR values greater than 1.5, and with intermediate responder proportions, so that responder and non-responder groups were both sufficiently represented. For each metabolite, participants were classified as responders when the post-exercise value fell outside the baseline-derived envelope. Responder and non-responder groups were then compared with respect to selected physiological and anthropometric characteristics, including peak RER, ventilatory thresholds, exercise duration, VO_2peak, relative fat-free mass and relative fat mass. This analysis was considered exploratory and hypothesis-generating

3.10 Predictive Modelling of VO_2peak

Predictive modelling was performed as an additional data science analysis to investigate whether baseline metabolomic and participant-level features could be used to estimate cardiorespiratory fitness. In this thesis, VO_2peak was selected as the prediction target because it represents the main physiological marker of aerobic capacity obtained from the cardiopulmonary exercise test.

The prediction task was formulated as a supervised regression problem. Each participant was represented by a set of input features, while the corresponding relative VO_2peak value, expressed in mL min^{-1} kg^{-1}, was used as the continuous outcome. The aim of this analysis was not to develop a final clinical prediction tool, but to explore whether the resting capillary metabolome contained predictive information about cardiorespiratory fitness and whether this information improved prediction beyond standard participant-level variables.

For the metabolomics branch of the analysis, baseline metabolite concentrations from the first resting timepoint were used. This choice allowed the model to evaluate the association between the resting metabolomic profile and VO_2peak, independently of the acute post-exercise response. Participants classified as potentially not exhausted were excluded from the prediction dataset, in order to reduce the influence of unreliable VO_2peak values on model training and evaluation. After this exclusion, the effective prediction cohort included 77 participants, consisting of 51 males and 26 females. The sex distribution therefore remained unbalanced, although the proportion was similar to that observed in the

original cohort. This imbalance was considered when interpreting model performance, because VO_2peak is strongly influenced by sex-related physiological differences.

Participant-level predictors included demographic, anthropometric and physical-activity-related variables. The clinical feature set included age, sex, body mass index, relative fat mass, relative fat-free mass and endurance sport frequency. Categorical variables were encoded numerically, while continuous variables were handled as numerical predictors. Missing values in the clinical matrix were imputed within the cross-validation procedure using training-set medians, in order to avoid information leakage from the test fold.

ElasticNet regression was used as the main predictive model. This model combines L1 and L2 regularization and is suitable for datasets with correlated predictors and a limited number of observations. Regularization is particularly important in metabolomics because the number of potential predictors can be high relative to the number of participants. Model hyperparameters were selected using ElasticNetCV within the training data.

Model performance was evaluated using a nested cross-validation framework. The outer validation loop consisted of five folds and was used to estimate out-of-sample predictive performance. Within each outer training set, metabolomic feature selection was performed using repeated cross-validation. A stability selection approach was applied to identify metabolites that were repeatedly selected across internal resampling runs. Only features exceeding a predefined selection-frequency threshold were retained for the metabolomics branch of the model. This procedure was used to reduce overfitting and to improve the interpretability of the metabolomic predictor set.

Three predictive branches were compared. The first model used only participant-level clinical features. The second model used only baseline metabolomic features selected through the nested stability-selection procedure. The third model combined the clinical and metabolomic predictions using an ensemble approach. The ensemble was constructed by weighting the clinical-only and omics-only predictions, allowing the analysis to evaluate whether metabolomic information added predictive value beyond standard participant-level features.

Prediction performance was assessed using out-of-fold predictions from the outer cross-validation loop. The coefficient of determination, R^2, was used to quantify the proportion of variance in VO_2peak explained by the model. Mean absolute error was used to measure the average absolute difference between observed and predicted VO_2peak values, while root mean squared error was used to give greater weight to larger prediction errors. These metrics were calculated for the clinical-only, omics-only and combined models.

Feature contribution was also evaluated. For the metabolomics branch, selected metabolites were summarized across outer folds to identify features that were repeatedly retained by the stability-selection procedure. For the clinical branch, a leave-one-feature-out analysis was performed. In this analysis, the clinical-only model was refitted several times, each time removing one clinical feature. The reduction in predictive performance after removing each feature was used as an estimate of its contribution to VO_2peak prediction.

The predictive modelling analysis was interpreted with caution. Although the original cohort included 100 participants, the effective sample size for prediction was reduced after excluding participants classified as potentially not exhausted. This exclusion was applied to reduce the influence of unreliable VO_2peak values, because incomplete maximal effort may weaken the relationship between physiological fitness and the measured predictors. However, it also further limited the number of observations available for model training and validation. This aspect is particularly important in a metabolomics-based prediction task, where the number of potential predictors can be high compared with the number of participants.

Additional cohort characteristics were also considered when interpreting model performance. The cohort was not perfectly balanced by sex, with a larger number of male than female participants, and VO_2peak is known to be strongly influenced by sex-related physiological differences. Moreover, the participants were mostly young adults, with a mean age of approximately 27 years and a minimum age of 18 years, limiting the variability of age-related effects in the dataset. For this reason, the predictive results should be interpreted as exploratory and cohort-specific rather than as a generalizable model of cardiorespiratory fitness.

The clinical feature set also included self-reported weekly endurance physical activity. This variable was retained because it represents information that can be easily collected before a CPET through a simple questionnaire. Although self-reported physical activity may be subjective and unevenly distributed across participants, it provides a practical estimate of habitual endurance training exposure and was therefore considered potentially informative for VO_2peak prediction.

Overall, this analysis extended the thesis beyond descriptive and inferential metabolomics. While the previous analyses focused on baseline reliability, acute exercise response and participant-level determinants of metabolite changes, the predictive modelling analysis evaluated whether the same dataset could also be used to estimate an important physiological outcome. This provided a data-driven link between the resting capillary metabolome, participant characteristics and cardiorespiratory fitness.

3.11 Computational Reproducibility and Software Environment

All data processing, statistical analyses and visualisations were performed in `Python 3.10.11` using `Jupyter Notebook` within a dedicated virtual environment (`.venv`). Data handling and tabular data processing were conducted using `pandas 2.3.3` and `NumPy 2.2.6`. Statistical analyses were performed using `SciPy 1.15.3` and `statsmodels 0.14.6`.

Multivariate analyses, clustering procedures and predictive modelling were implemented using `scikit-learn 1.7.2`. This included data standardisation, principal component analysis, t-SNE, clustering methods, repeated cross-validation and ElasticNet regression. Data visualisations were generated using `matplotlib 3.10.8` and `seaborn 0.13.2`.

For reproducibility, all stochastic procedures were executed using predefined random seeds. These included resampling-based analyses, dimensionality-reduction procedures, clustering robustness analyses and repeated cross-validation. The analysis workflow was organised into Jupyter notebooks covering preprocessing, baseline reliability, acute exercise-response analysis, participant-level determinant modelling and VO_2peak prediction.

In the prediction workflow, preprocessing, clinical imputation, metabolite selection, hyperparameter tuning and ensemble-weight optimisation were performed only within the training portion of each outer validation split. This procedure preserved the independence of held-out observations and reduced the risk of information leakage during model evaluation.

3.12 Statistical Decision Rules and Sensitivity Analyses

To improve transparency across the different analytical stages, predefined decision rules were applied consistently throughout the study. Metabolites were retained after quality control when at least 80% of values were valid and above the project-specific consensus limit of detection. Metabolite-level observations with an absolute z-score greater than 3.5 were treated as potential outliers and removed from the non-imputed primary dataset when appropriate.

For univariate baseline and post-exercise comparisons, p-values were adjusted using the Benjamini-Hochberg false discovery rate procedure. Associations were interpreted as statistically significant when the adjusted q-value was below 005. The same false discovery false discovery rate criterion was applied to the participant-level determinant analyses, in which each metabolite-predictor association was tested separately.

For the noise-adjusted analysis, metabolites with SNR greater than 1 were retained for the primary determinant analysis because their average post-exercise response exceeded estimated baseline-derived methodological noise. The responder analysis was restricted to metabolites with SNR greater than 1.5 and intermediate responder proportions, ensuring that responder and non-responder groups were sufficiently represented for exploratory comparison. Bidirectional response patterns were defined when at least 25 participants showed an increase and at least 25 participants showed a decrease relative to their individual baseline envelope.

The predictive modelling analysis used a primary metabolite set of 18 metabolites selected from the participant-level determinant analyses. A sensitivity analysis repeated the same modelling framework using the full panel of 52 retained metabolites. This comparison was used to assess whether the limited incremental predictive value of baseline metabolomics was related to the restricted primary metabolite subset.

All analyses were interpreted according to their role in the study design. Baseline reproducibility, principal component analysis and t-SNE were treated primarily as descriptive or exploratory analyses. The single-baseline comparison, noise-adjusted determinant analysis and predictive modelling analysis were interpreted more formally, while responder analyses and metabolite-selection frequencies were considered hypothesis-generating in the absence of external validation.

4 Baseline Reliability and Individual Metabolic Fingerprints

4.1 Baseline Reproducibility Between Repeated Resting Samples

Before analysing the metabolic response to exercise, the reproducibility of the capillary blood metabolome under resting conditions was evaluated in the study cohort of 100 healthy adults. This step was essential because the repeated-baseline design was used as the reference for the later noise-adjusted response analysis. If the baseline samples had shown large or inconsistent differences, the interpretation of post-exercise changes relative to resting variability would have been less reliable [14,15].

The comparison between Baseline 1 and Baseline 2 showed that the capillary metabolome was generally stable over the repeated resting measurements. Across the 52 retained metabolites, intra-individual variability was low, indicating that metabolite concentrations changed only modestly within the same participant before exercise. This supported the idea that the two baseline samples captured a relatively stable resting metabolic state rather than two unrelated measurements (Figure 4.1A).

Inter-individual variability was also consistent between the two baseline timepoints. The coefficients of variation calculated at Baseline 1 and Baseline 2 showed comparable distributions, suggesting that the overall between-participant structure of the metabolomic data was preserved across repeated resting samples. In other words, the repeated baseline did not introduce a major shift in the distribution of metabolite concentrations across the cohort (Figure 4.1B).

Baseline agreement was further supported by the correlation between the two repeated measurements. When $\log_2$-transformed metabolite values were compared across all participant-metabolite pairs, Baseline 1 and Baseline 2 showed a strong global correlation ($r = 0.996$, $p < 0.001$). This indicates that the repeated baseline samples provided highly concordant measurements of the resting capillary metabolome (Figure 4.1C).

To assess whether the second baseline introduced systematic changes, metabolite-level statistical comparisons were also performed between Baseline 1 and Baseline 2. Only three metabolites showed significant differences after correction for multiple testing (Serine, Ornithine, and Taurine), while

most metabolites remained centred around no change. This suggests that there was no broad systematic drift between the two baseline collections (Figure 4.1D).

Overall, these results indicate that repeated capillary blood sampling provided a reproducible resting reference state. The low intra-individual variability, consistent inter-individual variability and strong baseline agreement support the use of the repeated baseline as an estimate of resting fluctuation. This finding is important for the central aim of the thesis, because it justifies the later comparison between exercise-induced changes and baseline variability.

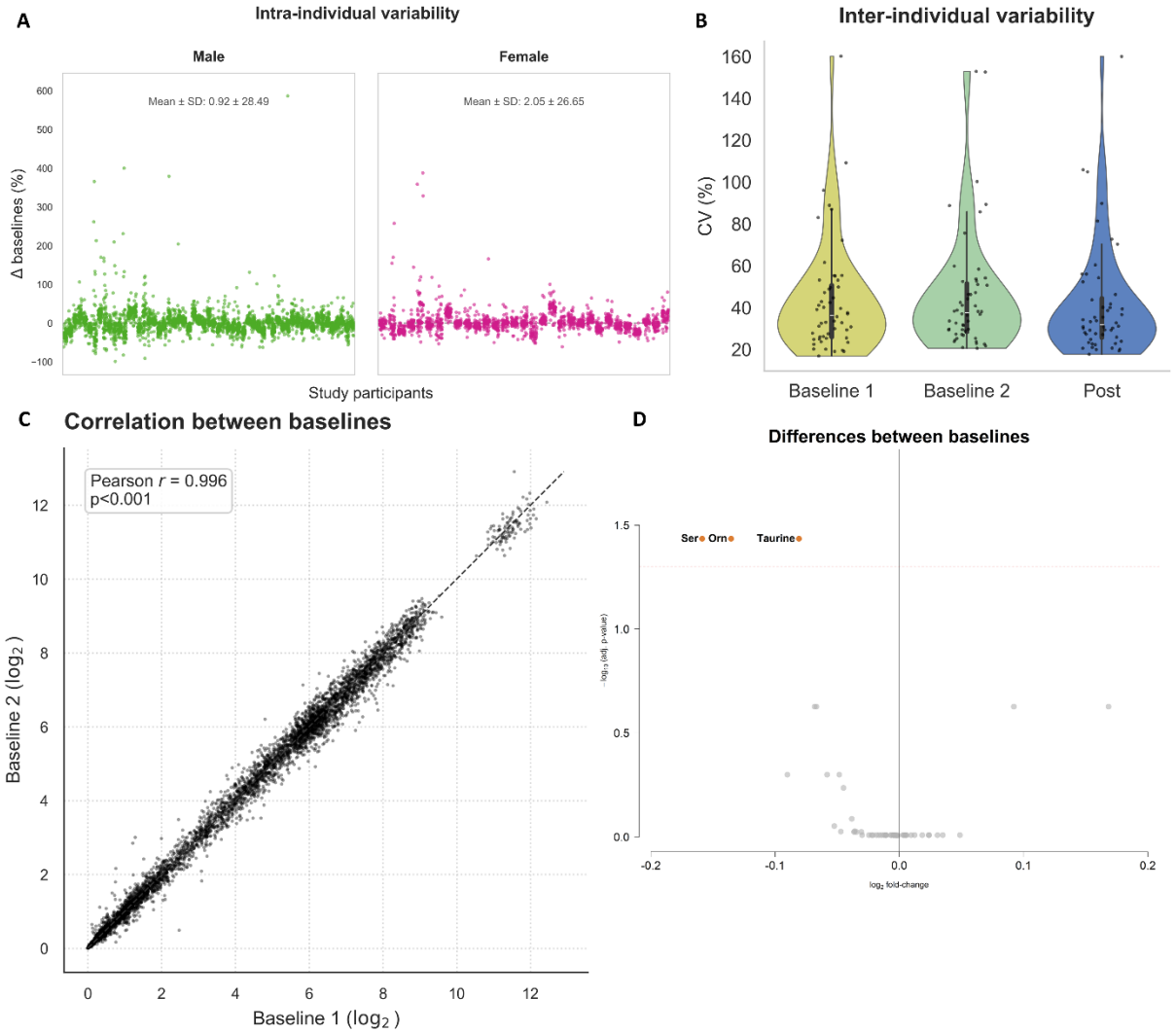


***Figure 4.1 Baseline reproducibility of capillary blood metabolome**. A: Intra-individual variability between Baseline 1 and Baseline 2, expressed as percentage difference between repeated resting samples and stratified by sex. Mean variability was low in both males and females, supporting the short-term stability of metabolite concentrations before exercise. **B**: Inter-individual variability, expressed as coefficient of variation, was comparable across Baseline 1, Baseline 2 and Post-exercise samples, indicating a consistent between-participant distribution across timepoints. **C**: Correlation between log_2-transformed metabolite values at Baseline 1 and Baseline 2 across all participant-metabolite pairs, showing strong agreement between repeated baseline measurements. **D**: Statistical comparison between Baseline 1 and Baseline 2 for each metabolite. Only 3 metabolites showed significant baseline differences after correction for multiple testing, while most metabolites remained close to zero log_2 fold-change. Overall, these analyses indicate that repeated capillary blood sampling provides a reliable resting reference for estimating baseline variability before exercise. Ser, serine; Orn, ornithine.*

4.2 Metabolite-Level Baseline Agreement

After evaluating the overall agreement between Baseline 1 and Baseline 2, baseline reproducibility was further assessed at the individual metabolite level. This analysis was important because the global correlation between repeated baseline samples could be influenced by the full range of metabolite concentrations, while individual metabolites may still differ in their short-term stability.

Metabolite-level Pearson correlations were calculated between Baseline 1 and Baseline 2 across participants. Overall, these correlations were generally high, with a median correlation of r = 0.830 and an interquartile range of 0.701 – 0.898. This indicates that, for most metabolites, participants maintained a similar relative ranking between the two repeated resting samples. In other words, individuals with higher or lower concentrations of a given metabolite at Baseline 1 tended to preserve a similar position at Baseline 2 (Figure 4.2).

However, baseline agreement was not identical across the full metabolite panel. Some metabolites showed stronger reproducibility, while others displayed greater short-term fluctuation. This heterogeneity is expected in metabolomics, because different metabolites may be affected by different levels of analytical variability, biological regulation, nutritional status or short-term physiological changes [14,15]. Therefore, the repeated-baseline design was useful not only to confirm global reproducibility, but also to identify metabolites with different degrees of resting stability.

The statistical comparison between Baseline 1 and Baseline 2 showed that only serine, ornithine and taurine differed significantly after correction for multiple testing. Although these differences indicate that a small subset of metabolites may have shown systematic changes supports the stability of the overall metabolite panel. Importantly, the presence of a few baseline differences does not invalidate the repeated-baseline design; rather, it highlights why estimating baseline variability is necessary before interpreting exercise-induced changes.

The metabolite-level analysis therefore provided a more detailed view of baseline reliability. Most metabolites showed good agreement between repeated resting samples, while a smaller subset displayed higher variability or modest systematic differences (Figure 4.1D). This finding supports the use of metabolite-specific baseline fluctuation in the subsequent signal-to-noise analysis, where each post-exercise response was interpreted relative to the resting variability of the same metabolite.

Overall, the metabolite-level baseline agreement analysis confirmed that the retained metabolite panel was suitable for repeated measures exercise response analysis. At the same time, it showed that baseline variability was not uniform across metabolites, reinforcing the need for a noise-aware framework rather than a single global estimate of variability.

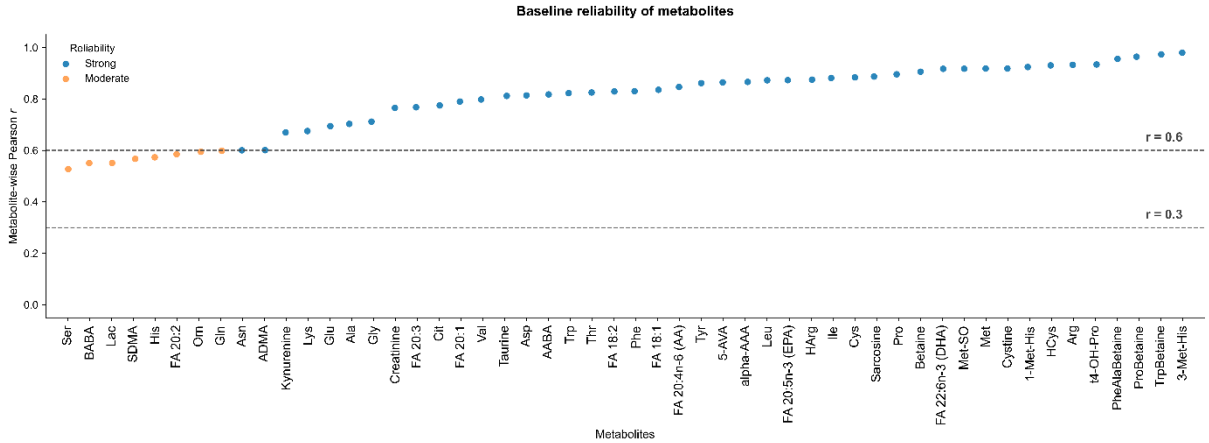


Figure 4.2 Metabolite-level baseline reliability. *Pearson correlations between Baseline 1 and Baseline 2 were calculated separately for each retained metabolite across participants. Metabolites are ordered by increasing reliability, with dashed lines indicating reference thresholds at r = 0.3 and r = 0.6. Most metabolites showed strong baseline agreement, while a*

smaller subset showed moderate reliability, supporting the use of metabolite-specific baseline variability in the noise-aware response analysis.

4.3 Individual Metabolic Fingerprints

After evaluating baseline reproducibility at the global and metabolite-specific levels, the structure of individual metabolic profiles was explored using dimensionality reduction. The aim of this analysis was to assess whether repeated samples from the same participant preserved a recognizable metabolic profile across timepoints.

The t-SNE visualization showed that samples from the same participant tended to cluster close to each other in the reduced-dimensional space. This pattern suggests that capillary blood metabolome retained a degree of individual specificity, or metabolic individuality, across repeated measurements [3,34,43,44]. In other words, each participant appeared to preserve a characteristic metabolomic profile that was partially maintained from the resting baseline samples to the post-exercise sample.

This finding is important for the interpretation of the repeated-baseline design. If samples from the same participant had been randomly distributed in the t-SNE space, this would have suggested that measurement variability or uncontrolled fluctuations dominated the data structure. Instead, the proximity of repeated samples from the same participant supports the idea that the metabolomic data contain stable subject-specific information [34,44].

The presence of individual metabolic fingerprints also has implications for the analysis of exercise responses. Since participants preserve partially stable metabolic profiles, post-exercise changes can be interpreted relative to a subject-specific resting state rather than only at the group level. This supports the use of repeated baseline measurements to estimate individual resting variability and strengthens the rationale for later noise-adjusted response analysis.

Overall, the t-SNE analysis provided exploratory evidence that the capillary metabolome captured stable individual-level information. Together with the baseline reproducibility and metabolite-level reliability results, this finding supports the use of the repeated-baseline framework as a reliable basis for interpreting acute exercise -induced metabolic changes.

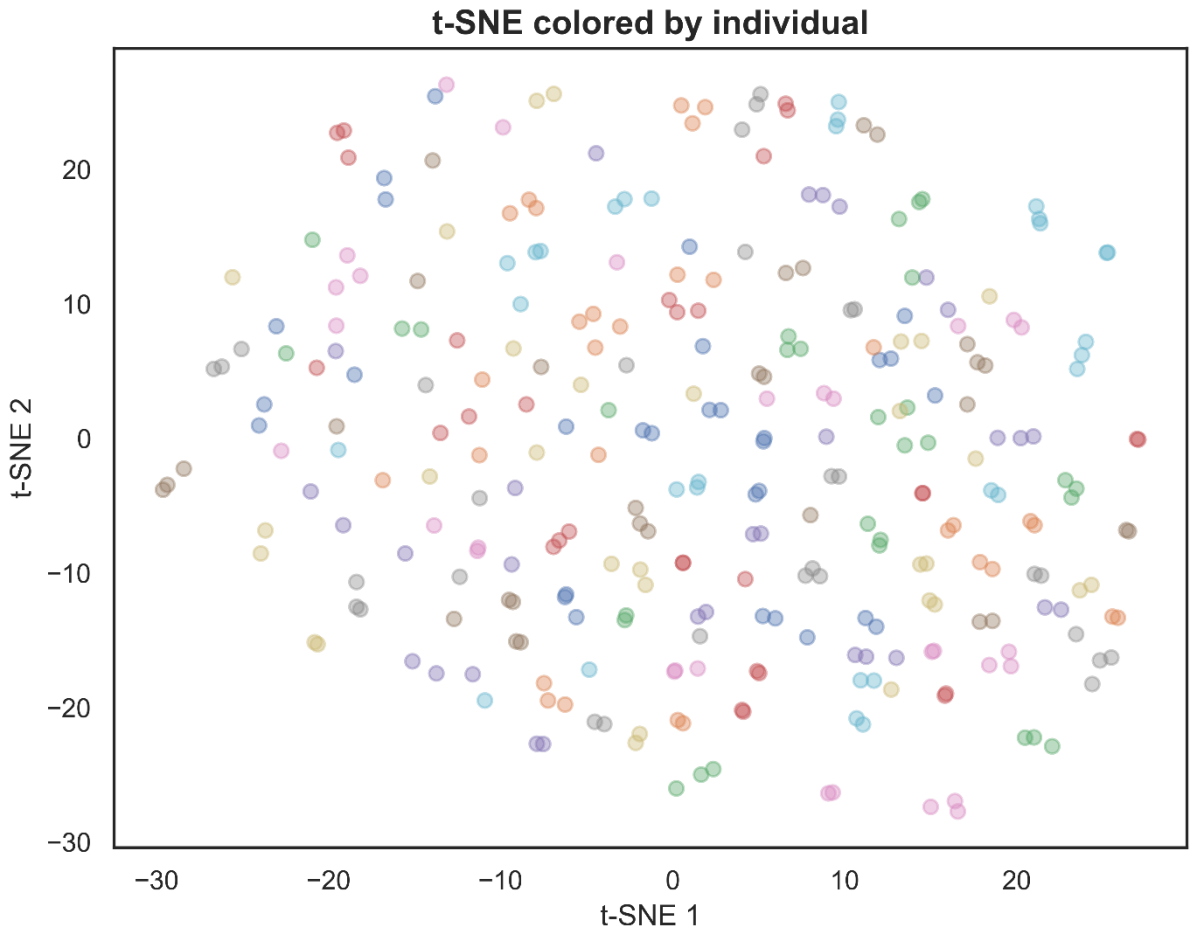


Figure 4.3: Individual metabolic fingerprints. *t-SNE visualization of log_2-transformed and standardized metabolite profiles, with samples coloured by participant. Samples from the same individual tend to appear close in the reduced-dimensional space, supporting the presence of stable subject-specific metabolic profiles across repeated measurements.*

4.4 Additional Reliability Checks

Additional reliability analyses were performed to complement the baseline reproducibility and metabolite-level agreement results. These analyses were used to evaluate whether the repeated baseline measurements showed systematic bias, whether variability differed across metabolites and whether the main source of variation was related to differences between participants rather than short-term fluctuation within the same participant [14,15,17].

First, baseline differences between Baseline 1 and Baseline 2 were inspected across the retained metabolite panel. As shown in Figure 4.4A, most baseline differences were centred around zero, confirming that short-term fluctuation between the two resting samples was generally limited. However, the width of the distributions varied across metabolites, indicating that resting variability was not identical for all analytes. This supported the need to estimate baseline fluctuation separately for each metabolite rather than assuming a single common noise level across the whole panel.

Shrinkage-adjusted baseline estimates were then compared with naïve baseline difference estimates for representative metabolites (Figure 4.4B). This analysis provided an additional view of how baseline variability behaved for metabolites with different response characteristics. The comparison helped identify metabolites with stable baseline profiles and metabolites for which baseline variation was more pronounced.

Variance decomposition provided a complementary perspective on baseline reliability. For selected metabolites, the total variability was separated into components related to between-subject differences, within-subject baseline variability and residual variation. As shown in Figure 4.4C, this analysis helped evaluate whether the observed variability was mainly driven by stable differences between participants or by short-term fluctuation within the same participant. A stronger between-subject component supports the presence of distinguishable individual metabolic profiles, whereas a larger within-subject component indicates greater baseline fluctuation.

Bland-Altman analyses were also used for representative metabolites to visualize agreement between Baseline 1 and Baseline 2 (Figure 4.4D). These plots show the difference between the two baseline measurements as a function of their average value, allowing possible systematic bias or concentration-dependent disagreement to be inspected. The comparison between stable and noisier metabolites further confirmed that baseline agreement was metabolite-specific.

Together, these additional checks supported the conclusions of the main baseline reliability analysis. The capillary metabolome showed reproducible resting measurements, but the amount of baseline variability differed across metabolites. This reinforces the central methodological rationale of the thesis: post-exercise responses should be interpreted relative to metabolite-specific resting variability rather than only through conventional group-level comparisons.

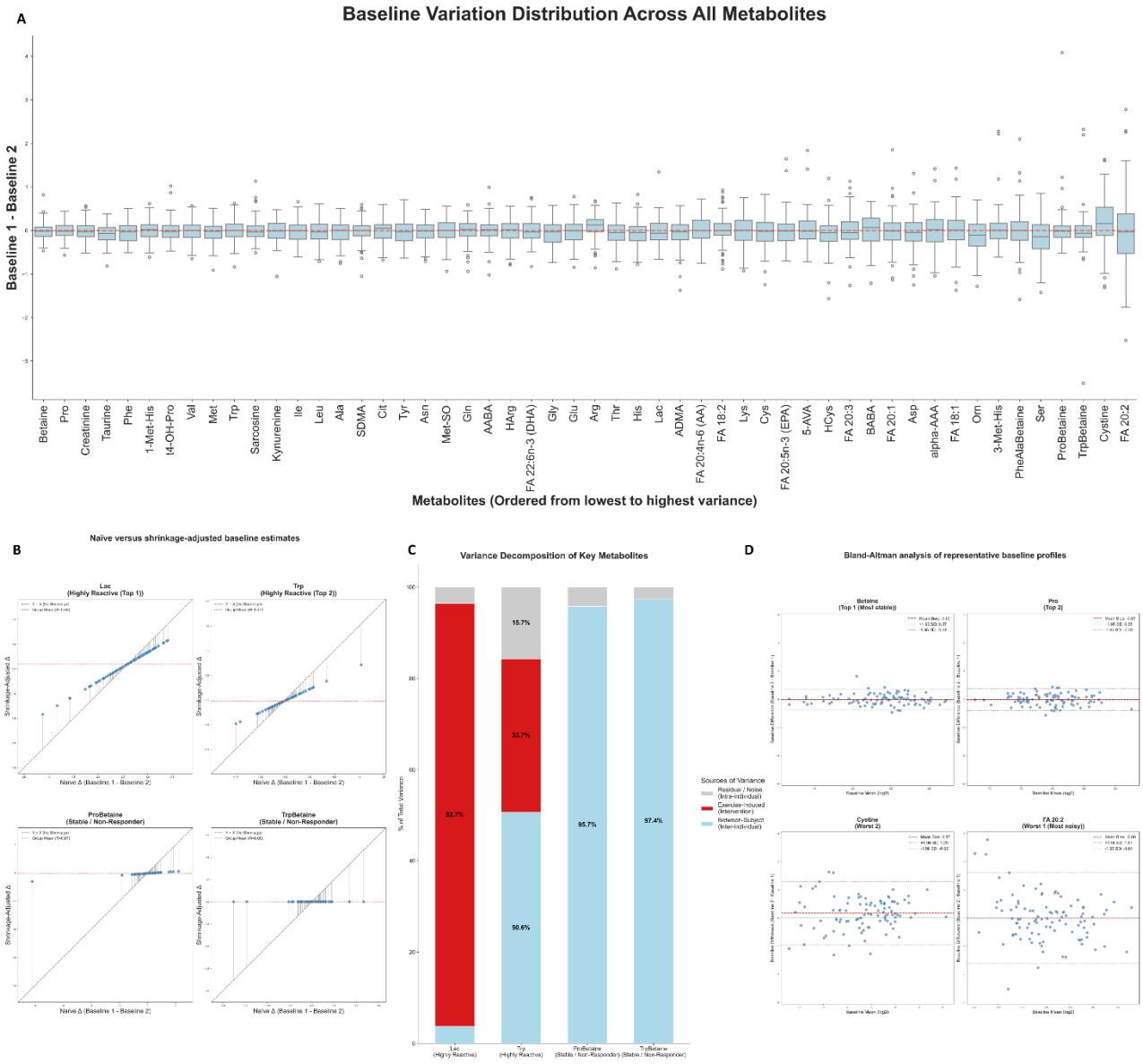


Figure 4.4. Additional reliability checks for repeated baseline measurements. ***A***: *distribution of baseline differences between Baseline 1 and Baseline 2 across all retained metabolites, ordered from the lowest to highest baseline variation.* ***B***: *comparison between naïve and shrinkage-adjusted baseline estimates for representative metabolites.* ***C***: *Variance decomposition for selected metabolites, showing the relative contribution of between-subject, within-subject and residual sources of variation.* ***D***: *Bland-Altman plots for representative stable and noisy metabolites. Together, these analyses show that baseline variability was generally limited but metabolite-specific, supporting the use of repeated baseline measurements to estimate resting fluctuation before analysing exercise responses. FA 20:2, eicosadienoic acid; Lac, lactate; Pro, proline; ProBetaine, proline betaine; TrpBetaine, tryptophan betaine; Trp, tryptophan*

4.5 Chapter Summary

This chapter evaluated whether the repeated baseline samples provided a reliable resting reference for the analysis of exercise-induced metabolite changes. The comparison between Baseline 1 and Baseline 2 showed low intra-individual variability, consistent inter-individual variability and strong overall agreement between repeated measurements, at the metabolite level, most analytes showed moderate-to-strong baseline reliability, although the degree of reproducibility was not identical across the full panel.

The t-SNE analysis further suggested the presence of individual metabolic fingerprints, with repeated samples from the same participant tending to remain close in the reduced-dimensional space. This

indicates that the capillary blood metabolome preserved subject-specific information across repeated measurements, rather than being dominated by random fluctuation.

Additional reliability checks confirmed that baseline differences were generally centred around zero, while also showing that resting variability differed across metabolites. This finding is important because it supports the use of metabolite-specific estimates of baseline fluctuation as an operational measure of methodological noise instead of a single global noise estimate.

Overall, the results of this chapter show that repeated capillary blood sampling provided a stable and informative resting reference state. This provides the methodological foundation for the next chapter, where post-exercise metabolite changes are analysed both through conventional single-baseline comparisons and through noise-adjusted metrics based on the repeated-baseline design.

5 Acute Exercise Response and Noise-Adjusted Metabolite Ranking

5.1 Global Metabolomic Shift After Maximal Exercise

After establishing the reproducibility of the repeated baseline measurements, the next step was to evaluate how the capillary blood metabolome changed after maximal exercise. The immediate post-

exercise sample was compared with the two resting baseline samples to determine whether the exercise stimulus induced a structured shift in the overall metabolomic profile.

Principal component analysis was used as an exploratory method to visualize the global structure of the dataset across timepoints. Before interpreting the PCA score distribution, the amount of variance explained by the first principal components was evaluated using a scree plot. The first principal component explained 29.7% of the total variance, while the second principal component explained 12.2%. Together, PC1 and PC2 therefore captured 41.9% of the total variance in the standardized metabolomic dataset. This supported the use of the first two principal components to visualize the main structure of the data while also indicating that additional components contributed to the remaining variability (Figure 5.1).

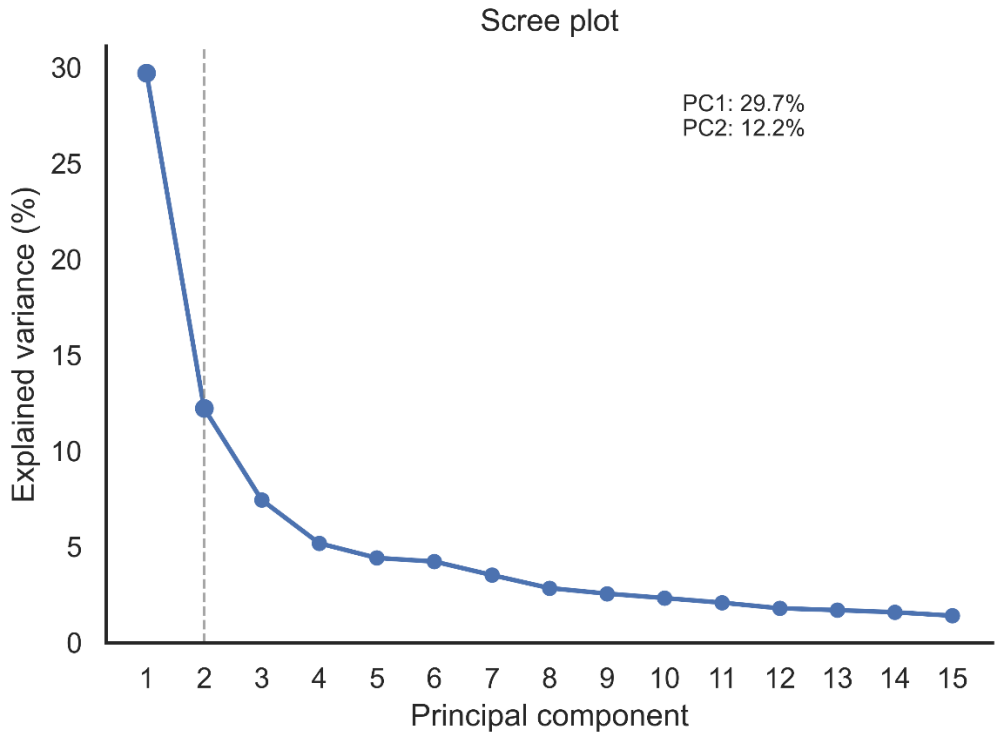


Figure 5.1. Scree plot of principal component analysis*. Explained variance across the first principal components of the log₂-transformed and standardized metabolomics dataset. PC1 explained 29.7% of the total variance, while PC2 explained 12.2%, together capturing 41.9% of the total variance. The first two components were therefore used to visualize the main metabolomic structure across timepoints.*

The two baseline samples showed a similar distribution in the PCA space, confirming the stability of the resting metabolomic state observed in the previous chapter. In contrast, post-exercise samples were clearly shifted away from the baseline samples, indicating that maximal CPET induced a broad and coordinated change in the capillary blood metabolome[1–6,44] (Figure 5.2).

This global shift is important because it shows that the metabolomic data captured a physiological response to exercise at the whole-panel level[1–5]. However, PCA is an exploratory analysis and does not identify which individual metabolites changed significantly or whether their responses exceeded resting variability. For this reason, the following sections examine the exercise response in more detail, first by evaluating metabolite class contributions and then by comparing conventional single-baseline analysis with noise-adjusted signal-to-noise stratification.

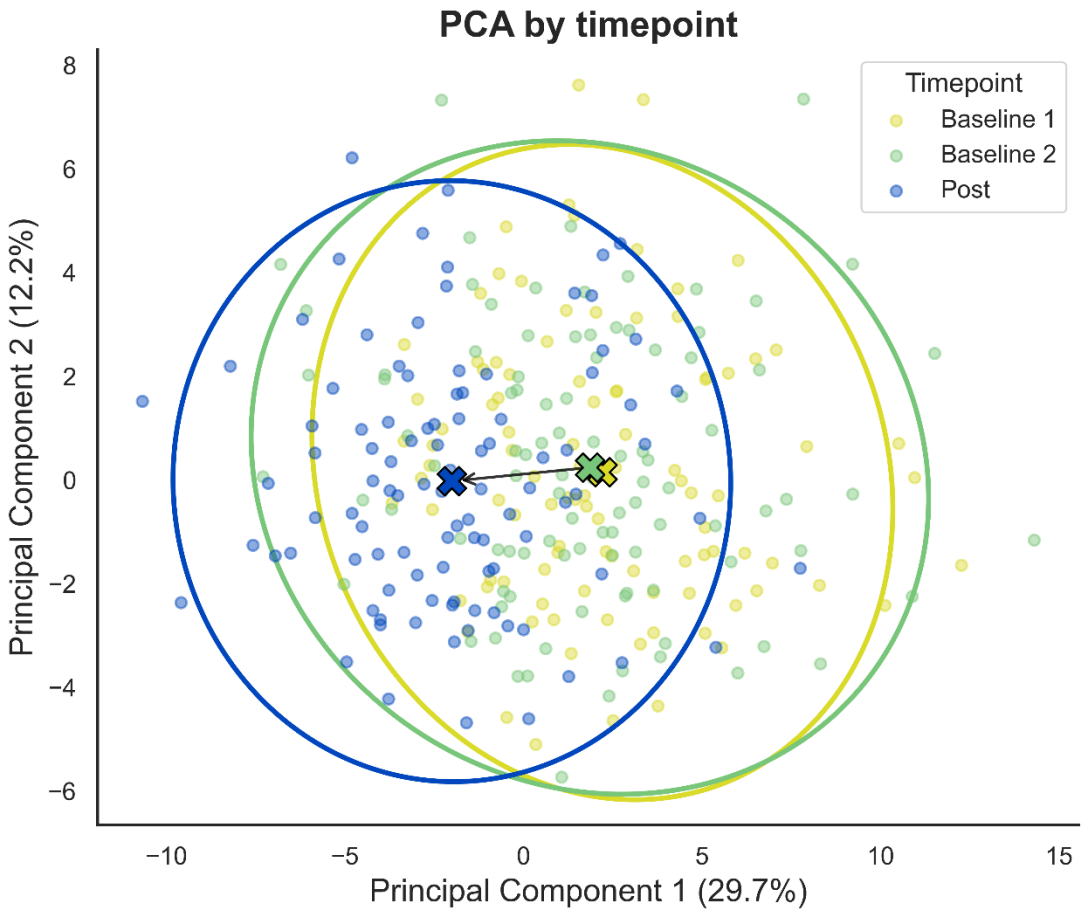


***Figure 5.2. Global metabolomic shift after maximal exercise**. Principal component analysis of log₂-transformed and standardized metabolite profiles across Baseline 1, Baseline 2 and Post-exercise samples. The two resting baseline timepoints show similar distributions, whereas post-exercise samples are separated along the first principal component, indicating a structured metabolomic response to maximal exercise.*

5.2 Metabolite Class Contribution to the Exercise Response

To better understand which metabolite classes contributed to the post-exercise shift observed in the PCA, the principal component loadings were examined. This analysis allowed the global separation between resting and post-exercise samples to be interpreted in terms of the metabolites and metabolite classes that contribute most strongly to the main axes of variation.

The first principal component was mainly driven by amino acids and amino-acid-related metabolites (Figure 5.3A,B). This indicates that the dominant post-exercise shift in the capillary blood metabolome was associated with changes in amino-acid-related metabolism. This finding is biologically plausible, because acute intense exercise can affect amino acid availability, tissue exchange, transamination reactions and systemic metabolic regulation[23].

In contrast, fatty acids contributed more strongly to the second principal component (Figure 5.3A,C). This suggests that fatty-acid related variation represented a secondary axis of variability in the dataset. Rather than defining the main separation between resting and post-exercise sample, fatty acids appeared to contribute more to inter-individual differences in the metabolic profile. This is consistent with the idea that lipid-related metabolites may be influenced by factors such as nutritional state, body composition and substrate availability[24–29].

Lactate was also an important metabolite in the interpretation of the acute exercise response. Its expected post-exercise increase reflects the strong glycolytic activation induced by maximal exercise[22] and provides a physiological reference for the validity of the experimental stimulus. However, the

global PCA structure indicated that the exercise response was not limited to lactate alone, but involved coordinated changes across broader metabolite classes.

Overall, the PCA loading analysis suggested that acute maximal exercise induced a structured metabolic response dominated by amino-acid-related changes, with fatty acids contributing to a secondary dimension of variability. This class-level interpretation provides a bridge between the global multivariate shift observed in the PCA and the metabolite-level analyses presented in the following sections.

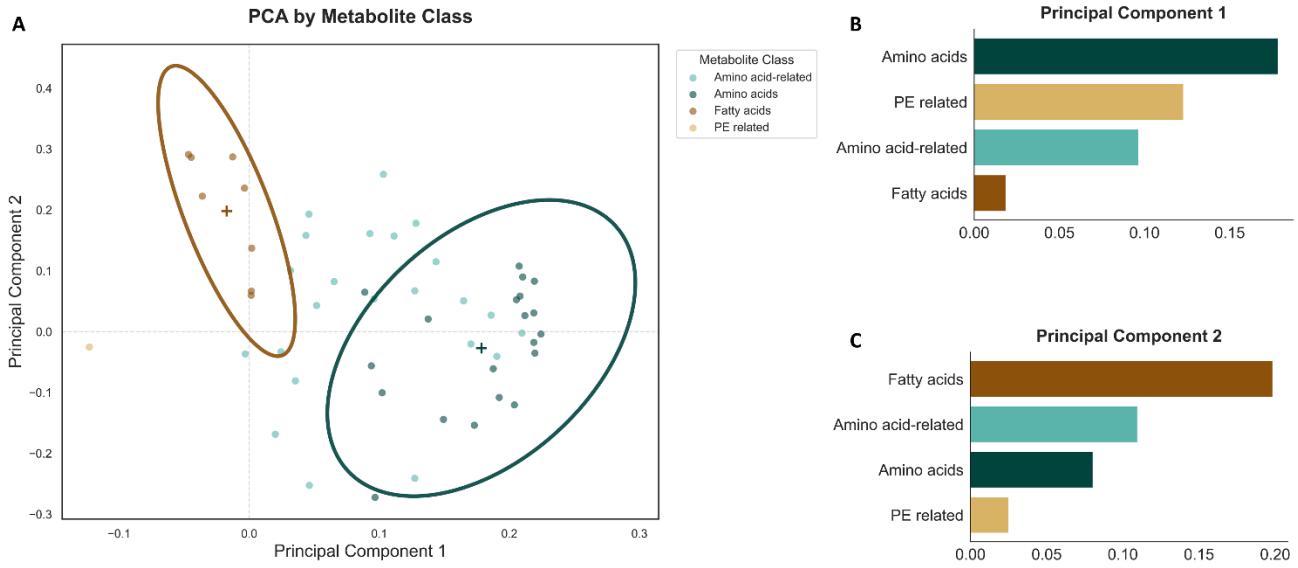


***Figure 5.3. Metabolite class contribution to the PCA structure. A**: PCA loading plot showing the distribution of metabolites according to their contribution to the first two principal components, coloured by metabolite class. **B**: Mean absolute loading contribution of each metabolite class to Principal Component 1, showing the dominant contribution of amino acids and amino-acid-related metabolites. **C**: Mean absolute loading contribution of each metabolite class to Principal Component 2, showing a stronger contribution of fatty acids to the secondary axis of variation.*

5.3 Participant Characteristics and PCA Structure

To further explore whether the global metabolomic structure was influenced by participant-level characteristics, the PCA space was stratified by sex, fasting status and BMI category. This analysis was exploratory and was used to evaluate whether major physiological or anthropometric variables contribute to the distribution of samples in the reduced-dimensional space.

When samples were coloured by sex, male and female participants showed largely overlapping distributions in the PCA space. This suggests that sex-related differences did not dominate the main axes of variation in the global metabolomic profile, although some degree of group-specific dispersion was present (Figure 5.4A).

BMI category was also inspected in the PCA space. Normal-weight, overweight and obese participants showed partial overlap, with some differences in group dispersion. This indicates that body composition and anthropometric status may contribute to inter-individual metabolic variability[26,27,45], but they do not appear to be the main driver of the global post-exercise shift (Figure 5.4B).

A similar exploratory analysis was performed for fasting status. Fasted and non-fasted participants also showed overlapping distributions, although the spread of the non-fasted group appeared wider. This is plausible because nutritional state can influence circulating metabolites, especially lipid-related and amino-acid-related compounds[25–27,46,47]. However, fasting status did not produce a clear separation comparable to the separation observed between resting and post-exercise samples (Figure 5.4C).

Overall, these PCA stratifications suggests that participant characteristics such as sex, fasting status and BMI may contribute to secondary variation in the metabolomic dataset, but they do not explain the dominant separation between resting and post-exercise samples. This supports the interpretation that the main PCA structure is primarily related to the acute exercise response, while participant-level variables may modulate specific dimensions of inter-individual variability.

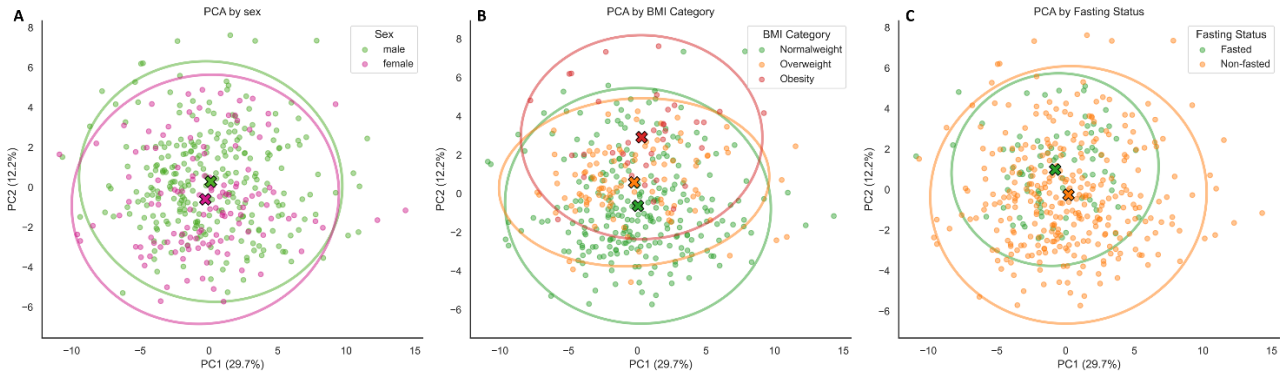


Figure 5.4. PCA stratified by participant characteristics. ***A****: PCA scores coloured by sex, showing substantial overlap between male and female participants.* ***B****: PCA scores stratified by BMI category, showing partial differences in group dispersion but no clear separation between categories.* ***C****: PCA scores coloured by fasting status, showing overlapping distributions between fasted and non-fasted participants. Overall, these stratifications suggests that participant characteristics may contribute to secondary inter-individual variability but do not dominate the main PCA structure.*

5.4 Conventional Single-Baseline Post-Exercise Response

After exploring the global metabolomic structure using PCA, the post-exercise response was analysed at the individual metabolite level using a conventional single-baseline approach. In this analysis, the immediate post-exercise sample was compared with Baseline 1 for each metabolite. This approach reflects the standard design used in many acute exercise metabolomics studies, where one pre-exercise measurement is used as the reference condition for identifying post-exercise changes[1–6].

The single-baseline $\log_2$ fold-change matrix provided an overview of the individual post-exercise response landscape across participants and metabolites. Hierarchical clustering of these fold-change values showed that the exercise response was broad and heterogeneous, with participants displaying different response patterns across the metabolite panel. Although this clustering suggested the presence of two main response-pattern groups, these groups should be interpreted cautiously, as exploratory patterns rather than clearly separated biological subtypes (Figure 5.5A).

5.4.1 Robustness of Exploratory Response-Pattern Clustering

Because the response-pattern clusters were exploratory, additional analyses were performed to evaluate whether the observed partition reflected a non-random structure and whether it was stable across resampling procedures and alternative clustering methods. Hierarchical Ward clustering of the participant-by-metabolite $\log_2$ fold-change matrix identified two groups containing 35 and 65 participants, respectively. The corresponding silhouette score was 0.1588, indicating limited separation between the two clusters.

A permutation test was used to determine whether this level of clustering structure exceeded that expected under a random response-pattern arrangement. After permuting metabolite values across participant 1000 times, the mean silhouette score under the null distribution was 0.052 ± 0.013. The observed silhouette score was higher than almost all permuted values, resulting in an empirical p-value of 0.001. This indicates that the clustering structure was unlikely to arise solely from random variation in the fold-change matrix.

Bootstrap consensus clustering provided an additional measure of stability. Across 1000 subsampling iterations, the mean within-cluster consensus was 0.664, whereas the mean between-cluster consensus

was 0.393. the difference between within-cluster and between-cluster consensus was therefore 0.270. This result suggests that participants assigned to the same hierarchical cluster tended to co-cluster more frequently than participants assigned to different clusters, although the separation was not strong enough to support the interpretation of sharply distinct biological subtypes.

Alternative clustering methods produced partially different assignments. K-means and Gaussian mixture models showed higher silhouette scores than hierarchical clustering, but their agreement with the original hierarchical solution was moderate with adjusted Rand indices of 0.485 and 0.457, respectively. Taken together, these analyses indicate that the response data contained non-random heterogeneity, but that the precise assignment of participants to discrete groups depended on the clustering method. Therefore, the identified groups were interpreted as exploratory response-pattern clusters rather than definitive responder phenotypes.

At the metabolite level, the single-baseline analysis showed a broad post-exercise response. After correction for multiple testing, 40 out of 52 retained metabolites were significantly altered following maximal CPET. This indicates that acute exercise induced widespread changes across the capillary blood metabolome, involving several metabolite classes rather than a small number of isolated markers[1,3–5,44] (Figure 5.5B).

Among all metabolites, lactate showed by far the strongest post-exercise change and the highest statistical significance ($p = 1.26 \times 10^{-222}$). This result was expected, because lactate is a well-established marker of high-intensity exercise and glycolytic activation[22]. Its strong increase supports the physiological validity of the exercise stimulus and confirms that the metabolomics data captured a clear acute response to maximal exercise.

The significant metabolites included amino acids, amino-acid-related metabolites and fatty acids, confirming that the response to maximal exercise was not limited to lactate alone. This is consistent with the PCA results, which suggested a broad and structured shift in the metabolomic profile after exercise. Therefore, the single-baseline analysis provided both an individual-level overview of response heterogeneity and a metabolite-level confirmation of the global separation observed in the multivariate analysis.

However, the single-baseline approach has an important limitation. Although it identifies metabolites that are significantly different after exercise, it does not indicate whether the magnitude of each change is large compared with the normal resting variability of that metabolite. In other words, a metabolite may reach statistical significance at the group level, but its response may still be close to the fluctuation observed between repeated baseline samples.

For this reason, the single-baseline analysis should be interpreted as a first descriptive step rather than as a complete measure of individual response reliability. It demonstrates that maximal exercise produces a broad metabolomic shift, but it does not distinguish between responses that clearly exceed resting variability and responses that remain close to baseline fluctuation. This limitation motivates the signal-to-noise analysis presented in the next section.

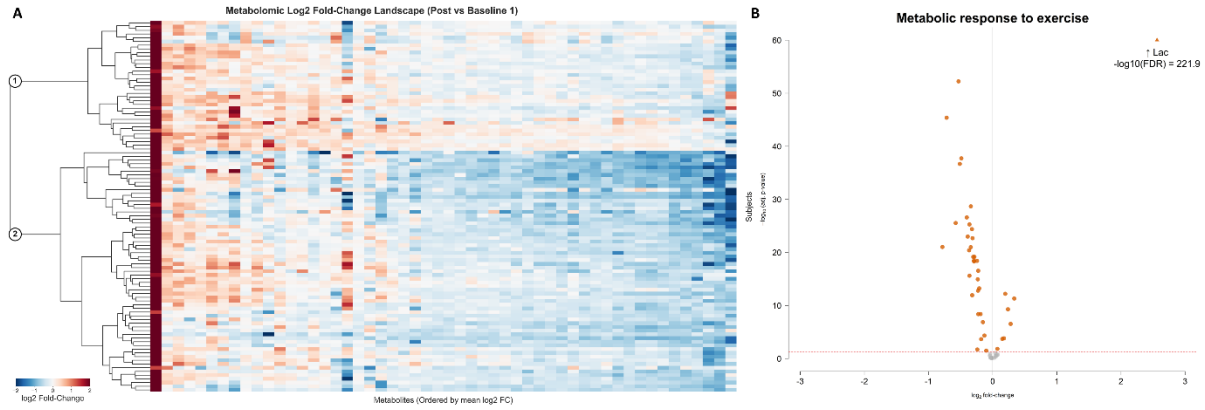


***Figure 5.5. Conventional single-baseline post-exercise response. A**: Hierarchical clustering heatmap of post-exercise log_2 fold-change values calculated relative to Baseline 1 across participants and metabolites. The heatmap shows a broad and heterogeneous metabolic response, with two exploratory participant response-pattern clusters. **B**: Volcano plot of metabolite-level changes between Post-exercise and Baseline 1. A total of 40 out of 52 metabolites were significantly altered after maximal CPET, with lactate showing the strongest response and higher statistical significance ($p = 1.26 \times 10^{-22}$). Together, these analyses describe the conventional single-baseline post-exercise response but do not account for resting baseline variability*

5.5 Signal-to-Noise Stratification of Exercise Responses

The conventional single-baseline analysis showed that maximal exercise induced widespread metabolite changes. However, this approach did not indicate whether each post-exercise response was large compared with the resting fluctuation of the same metabolite. For this reason, the repeated-baseline information was incorporated into the analysis through a signal-to-noise framework.

For each metabolite, exercise reactivity was compared with the variability observed between Baseline 1 and Baseline 2. In this framework, the exercise-induced change represents the signal, while the inter-baseline fluctuation represents the noise. Therefore, the signal-to-noise ratio provides a metabolite-specific measure of how large the post-exercise response is relative to resting variability.

The SNR analysis revealed a graded structure of exercise responsiveness across the retained metabolite panel. Eight metabolites showed very strong responses, with SNR values greater than or equal to 2.0. These metabolites had post-exercise changes at least twice as large as their estimated methodological noise Fifteen metabolites showed strong responses, with SNR values between 1.5 and 2.0, while 23 metabolites showed moderate responses, with SNR values between 1.0 and 1.5. The remaining 6 metabolites had SNR values below 1.0, indicating that their average post-exercise response was smaller than the estimated resting baseline variability (Figure 5.6).

This stratification refined the interpretation of the single-baseline results. While the conventional analysis identified 40 out of 52 metabolites as significantly altered after exercise, the SNR analysis showed that not all significant responses had the same reliability relative to methodological noise. Some metabolites showed responses that clearly exceeded resting variability, whereas others were closer to the range of normal baseline fluctuation.

This distinction is important because statistical significance and response robustness are not equivalent[16–18]. A metabolite can be significantly altered at the group level but still show a response magnitude that is modest relative to its resting variability. Conversely, metabolites with high SNR values are more likely to represent robust exercise-responsive signals because their post-exercise changes are large compared with the variability observed before exercise.

Overall, the signal-to-noise stratification provided a more conservative and informative ranking of exercise-responsive metabolites. Instead of treating all statistically significant metabolites as equally interpretable, the repeated-baseline framework allowed responses to be prioritized according to their magnitude relative to methodological noise[14,15,17,18,29]. This result represents a central methodological

contribution of the thesis and provides the basis for the participant-level determinant analyses presented in the next chapter.

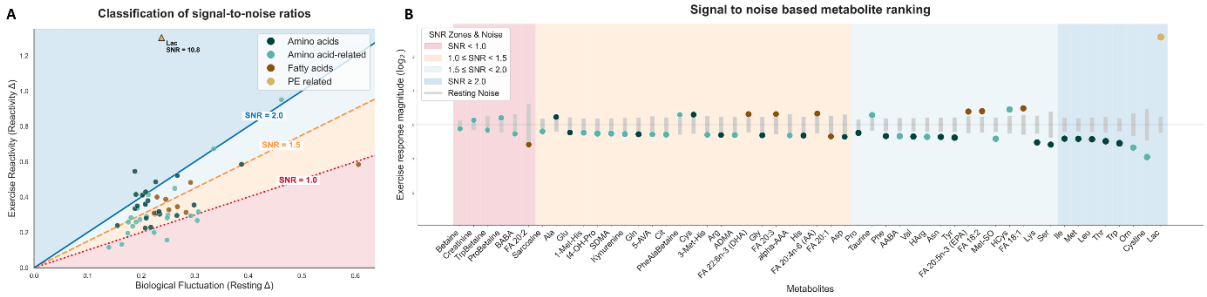


Figure 5.6. Signal-to-noise stratification of post-exercise metabolite responses. A*: Classification of metabolites according to exercise reactivity and resting baseline fluctuation. Dashed lines indicate signal-to-noise ratio thresholds at SNR = 1.0, 1.5 and 2.0, separating responses comparable to baseline variability from stronger exercise-induced responses.* ***B****: Metabolite ranking based on signal-to-noise ratio, with background shading indicating response zones. This analysis shows that repeated-baseline information refines the interpretation of post-exercise changes by identifying which metabolites exceed resting variability most clearly.*

5.6 Chapter Summary

This chapter characterized the acute metabolomic response to maximal exercise. The PCA analysis showed that post-exercise samples were clearly shifted from the repeated resting baseline samples, indicating that maximal CPET induced a structured change in the capillary blood metabolome. The first two principal components captured a substantial proportion of the overall variance, and the loading analysis suggested that the main exercise-related shift was primarily driven by amino acids and amino-acid-related metabolites, while fatty acids contributed more strongly to a secondary axis of variation.

The conventional single-baseline analysis confirmed this broad post-exercise response at the metabolite level. Comparing Post-exercise with Baseline 1 identified 40 out of 52 retained metabolites as significantly altered after exercise, with lactate showing the strongest and most statistically significant response. The single-baseline fold-change heatmap also showed that the response was heterogeneous across participants, although the resulting clusters were interpreted only as exploratory response-pattern groups rather than clearly defined biological subtypes.

The signal-to-noise analysis provided a more refined interpretation of these results. By incorporating Baseline 2, post-exercise changes could be compared with resting baseline fluctuation. This showed that not all statistically significant metabolites had the same response magnitude relative to resting variability. Some metabolites clearly exceeded baseline noise, whereas others showed responses closer to the range of normal baseline fluctuation.

Overall, this chapter shows that maximal exercise produced a broad and structured metabolic response, but also that conventional statistical significance alone is not sufficient to determine the reliability of individual metabolite responses. The repeated-baseline signal-to-noise framework therefore provides a more informative ranking of exercise-responsive metabolites. This result motivates the next chapter, where participant-level characteristics are analysed to investigate whether inter-individual variability in the noise-adjusted response can be partly explained by physiological, anthropometric or experimental factors. In addition the next chapter further explores whether selected individual responder patterns can be related to physiological performance variables

6 Participant-Level Determinants and Directional Response Patterns

6.1 Rationale for Participant-Level Determinant Analysis

The previous chapter showed that maximal exercise induced a broad and structured metabolomic response, and that repeated-baseline information allowed this response to be ranked according to its magnitude relative to resting variability. However, the signal-to-noise analysis was performed at the metabolite level and did not explain why the magnitude of the response differed across participants.

Inter-individual variability is expected in acute exercise metabolomics[4,5,16–18]. Even when a metabolite shows a clear group-level post-exercise change, individual participants may show stronger, weaker or even opposite responses. These differences may reflect physiological characteristics, such as cardiorespiratory fitness, age, sex or body composition, but they may also depend on experimental factors, such as fasting status or the degree to which the participant reached maximal effort during the cardiopulmonary exercise test[5,16,17].

For this reason, participant-level determinant analysis was performed to investigate whether selected physiological, anthropometric and experimental variables were associated with the noise-adjusted metabolite response. The analysis focused on metabolites with signal-to-noise ratio greater than 1, because these metabolites showed an average post-exercise response larger than their estimated resting baseline fluctuation. This filtering step allowed the determinant analysis to focus on metabolites for which the exercise response was more likely to reflect a true exercise-related signal rather than baseline noise.

Six participant-level characteristics were considered: VO_2peak, relative fat mass, age, sex, fasting status and exhaustion status. VO_2peak was included as the main marker of cardiorespiratory fitness, while relative fat mass, age and sex were included as physiological and anthropometric descriptors of

the participants. Fasting status was included because nutritional state can influence circulating metabolites and substrate availability before exercise. Exhaustion status was included because incomplete attainment of maximal effort during CPET may attenuate the physiological and metabolic response to the exercise stimulus.

The global determinant map showed that exhaustion status and fasting status were the most prominent participant-level variables associated with noise-adjusted metabolite responses. Exhaustion status was associated with 16 metabolites, mainly belonging to the amino acid and amino-acid-related classes, while fasting status was associated with a smaller subset of 8 metabolites. In contrast, VO_2peak, relative fat mass, age and sex showed fewer significant associations, suggesting a more limited or metabolite-specific contribution to response variability. These results are summarized in Figure 6.1 and provide the basis for the more detailed analyses presented in the following sections.

This chapter therefore moves from metabolite-level response ranking to participant-level interpretation. The aim is not only to identify which metabolites respond to exercise, but also to evaluate whether part of the observed response heterogeneity can be explained by measurable characteristics of the participants or by experimental conditions. The following sections focus first on exhaustion status and fasting status, which emerged as the strongest determinants of noise-adjusted metabolite responses. The final part then examines individual response directions using the baseline-envelope approach and includes an exploratory responder analysis to evaluate whether selected post-exercise deviation patterns are associated with physiological performance variables.

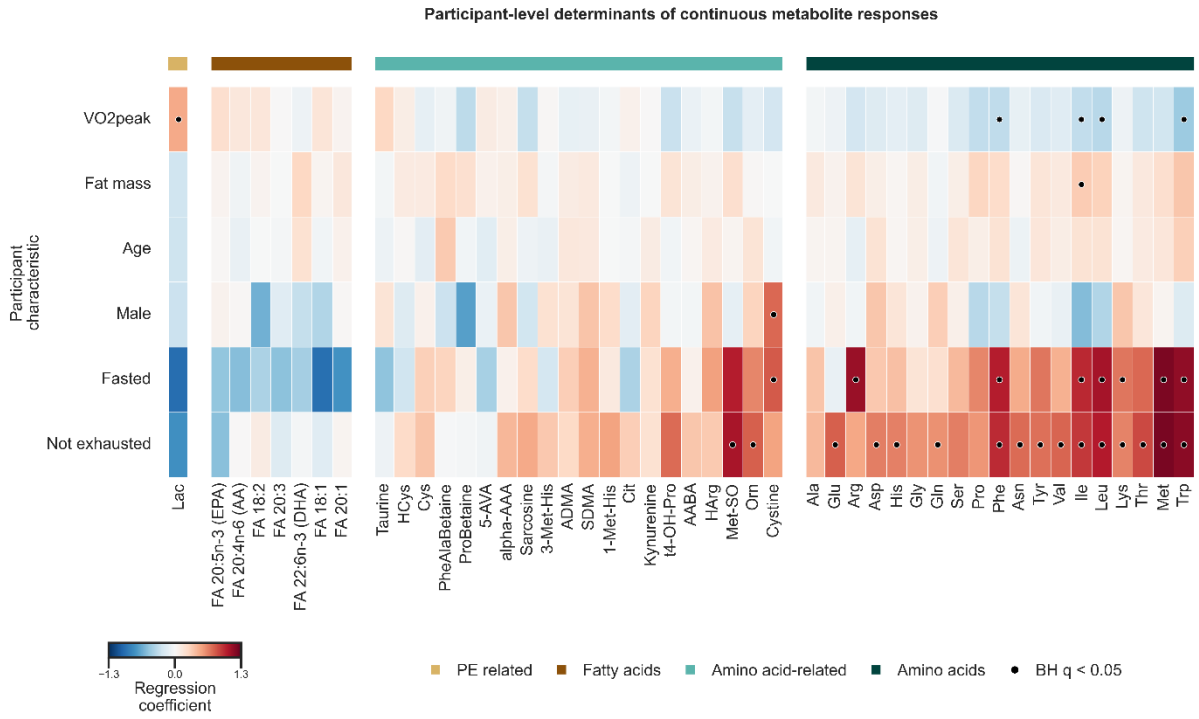


Figure 6.1. Participant-level determinants of noise-adjusted metabolite responses. *Heatmap of regression coefficients from univariable linear models testing the association between participant-level characteristics and noise-adjusted post-exercise metabolite responses. Rows represent participant characteristics, while columns represent metabolites grouped by metabolite class. Colours indicate the direction and magnitude of the regression coefficient, and black dots mark associations significant after Benjamini-Hochberg correction. Exhaustion status and fasting status showed the strongest association patterns, mainly involving amino acids and amino-acid-related metabolites, whereas VO_2peak, relative fat mass, age and sex showed fewer significant associations.*

6.2 Exhaustion Status and Attenuated Amino-Acid Responses

Among the participant-level characteristics evaluated, exhaustion status showed the strongest association with the noise-adjusted metabolite response. This variable was derived from the CPET and verification-bout data and was used to identify participants who may not have reached a fully maximal effort during the incremental test. Since the metabolomic response was measured immediately after incremental CPET, exhaustion status was considered an important experimental factor for interpreting inter-individual variability.

The determinant analysis showed that exhaustion status was associated with the response of 16 metabolites after correction for multiple testing (Figure 6.1). Most of these metabolites belonged to the amino acid and amino-acid-related classes, indicating that incomplete attainment of maximal effort was mainly reflected in this part of the metabolome[5]. This result suggests that the amino-acid-related response to exercise is sensitive to the effective intensity of the physiological stimulus.

This pattern is shown in Figure 6.2, where the noise-adjusted responses of the 16 exhaustion-associated metabolites are compared between exhausted and potentially not-exhausted participants. Across most metabolites, the not-exhausted group showed responses closer to zero or less pronounced changes, consistent with an attenuated metabolic response to the exercise stimulus.

Participants classified as potentially not exhausted generally showed an attenuated post-exercise metabolite response compared with participants classified as exhausted. In other words, for several amino acids and amino-acid-related metabolites, the magnitude of the noise-adjusted response was smaller in participants who did not meet the exhaustion criteria. This pattern is physiologically plausible, because a lower effective exercise stimulus may reduce the systemic metabolic perturbation induced by CPET[5,23].

This finding is important for two reasons. First, it shows that exhaustion status is not only a technical quality-control variable for interpreting VO_2peak, but also a relevant determinant of the metabolomic response itself participants who do not reach a sufficiently intense effort may show a weaker metabolic response, even if they completed the same experimental protocol. Second, it indicates that part of the inter-individual variability observed in the exercise response may be explained by differences in effort attainment rather than only by stable biological differences between participants.

From an analytical perspective, this result supports the inclusion of exhaustion status in participant-level modelling. Without accounting for this variable, attenuated responses in potentially not-exhausted participants could be misinterpreted as biological heterogeneity unrelated to test performance. Therefore, exhaustion status provides an important link between the physiological validity of the CPET and the interpretation of downstream metabolomics results.

Overall, the association between exhaustion status and multiple amino-acid-related responses suggests that maximal effort is a key factor for obtaining a robust exercise-induced metabolomic signal[16–18]. This reinforces the need to interpret acute exercise metabolomics data in relation not only to sampling design and baseline variability, but also to the quality and intensity of the exercise stimulus.

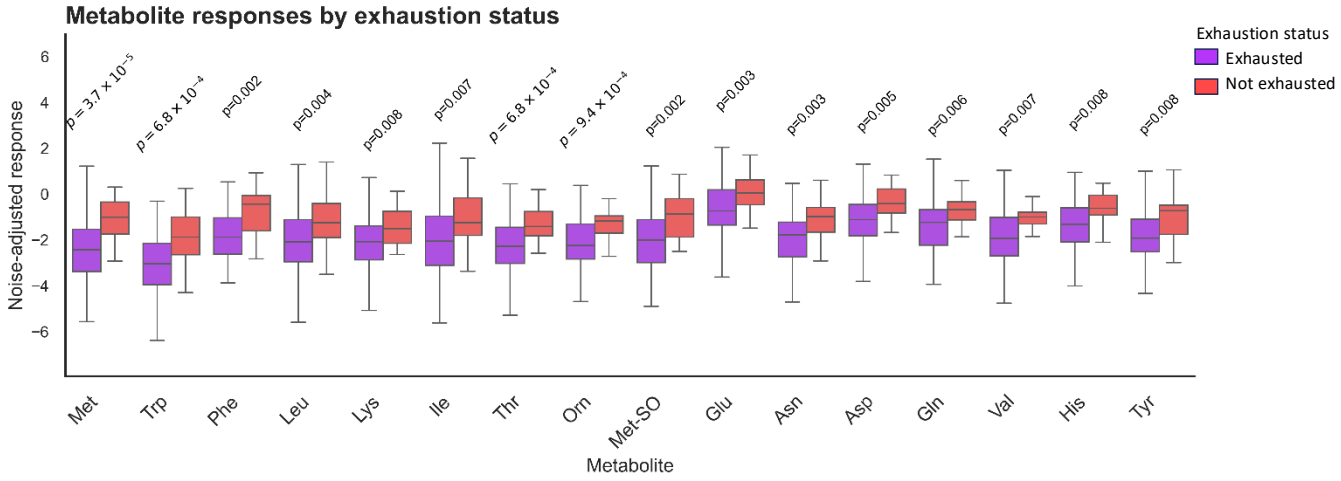


Figure 6.2. Exhaustion status and attenuated amino-acid-related responses. *Noise-adjusted post-exercise responses for metabolites significantly associated with exhaustion status. Boxplots compare participants classified as exhausted and potentially not exhausted according to CPET and verification-bout criteria, with p-values reported above each metabolite. Most significant metabolites belonged to the amino acid and amino-acid-related classes. Participants classified as potentially not exhausted generally showed attenuated responses, indicating that incomplete attainment of maximal effort was associated with a weaker exercise-induced metabolomic response. Asn, asparagine; Asp, aspartic acid; FDR, false discovery rate; Gln, glutamine; Glu, glutamic acid; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Met-SO, methionine-sulfoxide; Orn, ornithine; Phe, phenylalanine; SNR, signal-to-noise ratio; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Val, valine.*

6.3 Fasting Status and Amino-Acid Response Variability

After exhaustion status, fasting status emerged as the second most relevant participant-level determinant of the noise-adjusted metabolite response. This variable distinguished participants who performed the exercise protocol after an overnight fast from those tested in a non-fasted state. Since circulating metabolites can be influenced by recent food intake, fasting status was considered an important experimental and physiological factor for interpreting inter-individual variability.

The determinant analysis showed that fasting status was associated with the response of 8 metabolites after correction for multiple testing (Figure 6.1). These associations mainly involved amino acids and amino-acid-related metabolites, indicating that pre-exercise nutritional state influenced the metabolic response to maximal CPET. This result is plausible because fasting can affect substrate availability, amino acid concentrations and the balance between nutrient use and mobilization during exercise[25,46–48].

Interestingly, 6 of the 8 metabolites associated with fasting status were also among the metabolites associated with exhaustion status. This overlap suggests that a subset of amino-acid-related responses was sensitive to both the effective intensity of the exercise stimulus and the pre-exercise nutritional state[46–48]. Therefore, these metabolites may represent a response region in which experimental conditions strongly influence the magnitude of the observed post-exercise change.

Compared with exhaustion status, the association pattern for fasting status was more limited. While exhaustion status was associated with a broader set of metabolites, fasting status affected a smaller subset of response. This suggests that the intensity and completeness of the exercise stimulus had a stronger overall influence on the noise-adjusted response, whereas nutritional state contributed more selectively to specific metabolite changes (Figure 6.3).

The presence of fasting-related associations is nevertheless important for the interpretation of acute exercise metabolomics. Participants tested in a fasted or non-fasted state may differ in baseline substrate availability and in this way metabolites are mobilized or consumed during exercise. As a

result, part of the inter-individual variability in post-exercise metabolite responses may reflect differences in nutritional state rather than only differences in fitness, body composition or sex.

This finding also highlights the importance of considering experimental conditions when analysing metabolomics data. In studies where participants are not tested under identical nutritional conditions, fasting status can act as a source of biological and experimental variability. Including fasting status in participant-level models therefore helps reduce the risk of attributing nutrition-related response differences to other physiological characteristics.

Overall, fasting status contributed to the variability of selected amino-acid-related responses after maximal exercise. Although its effect was less widespread than that of exhaustion status, it represented an important determinant of response heterogeneity and should be considered when interpreting acute exercise metabolomics data based on mixed fasting and non-fasting cohorts.

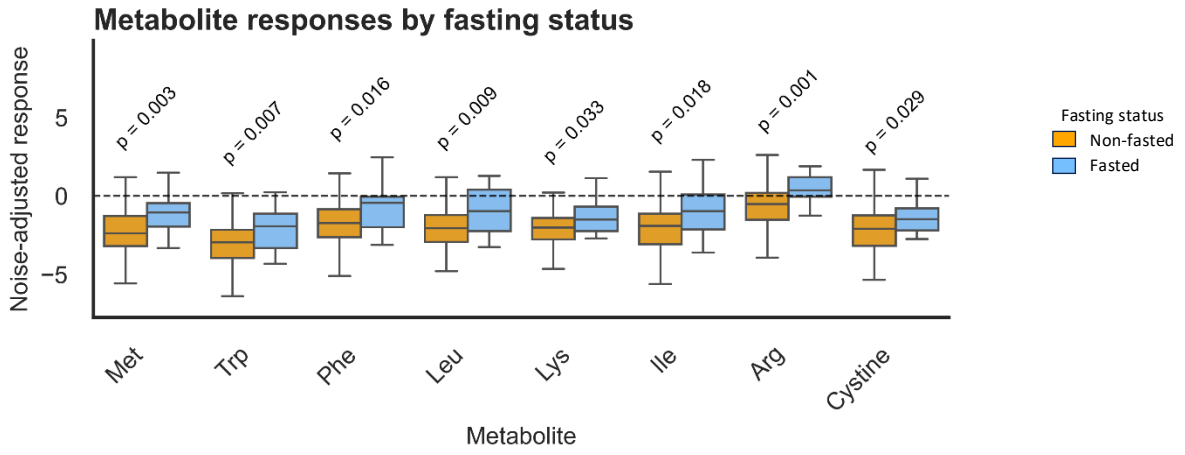


Figure 6.3. Fasting status and noise-adjusted metabolite responses. *Noise-adjusted post-exercise responses for metabolites significantly associated with fasting status. Boxplots compare participants tested in fasted and non-fasted conditions, with p-values reported above each metabolite. Fasting status was associated with 8 metabolites, mainly amino acids and amino-acid-related metabolites. Across these metabolites, fasted and non-fasted participants showed different response distributions, indicating that pre-exercise nutritional state contributed to inter-individual variability in the acute metabolomic response.*

6.4 Limited Role of Other Participant Characteristics

In addition to exhaustion status and fasting status, the determinant analysis also evaluated the association between noise-adjusted metabolite responses and VO_2peak, relative fat mass, age and sex. These variables were included because they represent relevant physiological, anthropometric and demographic characteristics that may influence exercise metabolism and inter-individual response variability.

Compared with exhaustion status and fasting status, these participant characteristics showed fewer significant associations after correction for multiple testing, as shown in Figure 6.1. VO_2peak was included as the main marker of cardiorespiratory fitness, but it did not show a broad association pattern across the metabolite panel. This suggests that, within this cohort, cardiorespiratory fitness was not the main determinant of the acute noise-adjusted metabolite response. Instead, the effective intensity of the exercise stimulus and the nutritional state before exercise appeared to explain a larger part of the observed response variability.

Relative fat mass also showed a limited association pattern. Although body composition can influence substrate availability and metabolic regulation, its relationship with the acute post-exercise metabolite response appears more selective than global[24,26,27,45]. This indicates that body composition may contribute to specific aspects of response variability, but it did not dominate the overall determinant structure.

Age and sex also showed fewer significant associations compared with exhaustion status and fasting status. This does not imply that age and sex are irrelevant for exercise physiology or metabolism[26,27,49]. Rather, in the present dataset, their effects on the selected noise-adjusted metabolite responses were less evident after multiple-testing correction. This may also reflect the characteristics of the cohort, which consisted mainly of young adults and was not perfectly balanced by sex.

Overall, these results suggest that the main sources of inter-individual variability in the acute metabolomic response were related more strongly to experimental and physiological state variables, particularly exhaustion status and fasting status, than to general demographic or anthropometric characteristics. Therefore, VO_2peak, relative fat mass, age and sex were interpreted as having a more limited or metabolite-specific role in this analysis.

6.5 Directional Response and Baseline-Envelope Analysis

In addition to analysing continuous noise-adjusted metabolite responses, an individual-level baseline-envelope analysis was performed to characterize the direction of post-exercise changes. While the signal-to-noise analysis quantified the magnitude of metabolite responses relative to resting variability, the baseline-envelope approach evaluated whether each participant showed an increase, a decrease or no relevant change relative to their own repeated-baseline range[16–18].

For each participant and metabolite, the baseline envelope was defined as the interval between Baseline 1 and Baseline 2. The post-exercise value was then compared with this interval. If the post-exercise value was above the envelope, the participant was classified as showing an increasing response; if it was below the envelope, the response was classified as decreasing; if it remained within the envelope, the response was classified as unchanged. This allowed post-exercise responses to be interpreted relative to each participant's own resting variability.

The first part of this analysis focused on the 18 metabolites previously highlighted by the participant-level determinant analysis. These included 6 metabolites associated with both exhaustion status and fasting status, 10 additional metabolites associated only with exhaustion status, and 2 additional metabolites associated only with fasting status. For these metabolites, the proportion of participants showing post-exercise values outside the individual baseline envelope was high, indicating that their responses frequently exceeded the subject-specific resting range. For example, tryptophan and threonine showed post-exercise values outside the baseline envelope in more than 90% of participants. Several amino acids showed particularly strong directional patterns, with very high percentages of participants outside the envelope.

To further investigate heterogeneity in response direction, all 46 metabolites with SNR > 1 were screened for bidirectional response patterns. Bidirectional metabolites were defined as those for which at least 25 participants showed an increasing response and at least 25 participants showed a decreasing response relative to their individual baseline envelope. Using this criterion, 8 bidirectional metabolites were identified: 4 amino-acid-related metabolites, including arginine, and 4 fatty acids.

A secondary determinant analysis was then performed on these 8 bidirectional metabolites to test whether the direction of response, increasing versus decreasing, was associated with participant-level characteristics. In most cases, no strong or consistent associations were observed. This suggests that, for the majority of bidirectional metabolites, the available physiological and experimental variables were not sufficient to explain why some participants increased while others decreased after exercise.

Arginine represented the main exception. Its response direction showed an association with fasting status, with participants in the increasing-response group more frequently belonging to the fasted condition. This is particularly notable because arginine was also among the metabolites previously associated with fasting status in the continuous determinant analysis. Therefore, arginine represents a

metabolite for which both the magnitude and the direction of the post-exercise response appear to be influenced, at least in part, by pre-exercise nutritional state.

Overall, the baseline-envelope analysis complemented the continuous signal-to-noise framework by providing an individual-level view of response direction. It confirmed that many of the previously identified determinant-associated metabolites frequently exceeded individual baseline variability, while also showing that truly bidirectional responses were limited to a smaller subset of metabolites. Among these, only arginine showed a clearly interpretable association with one of the available participant-level characteristics, namely fasting status.

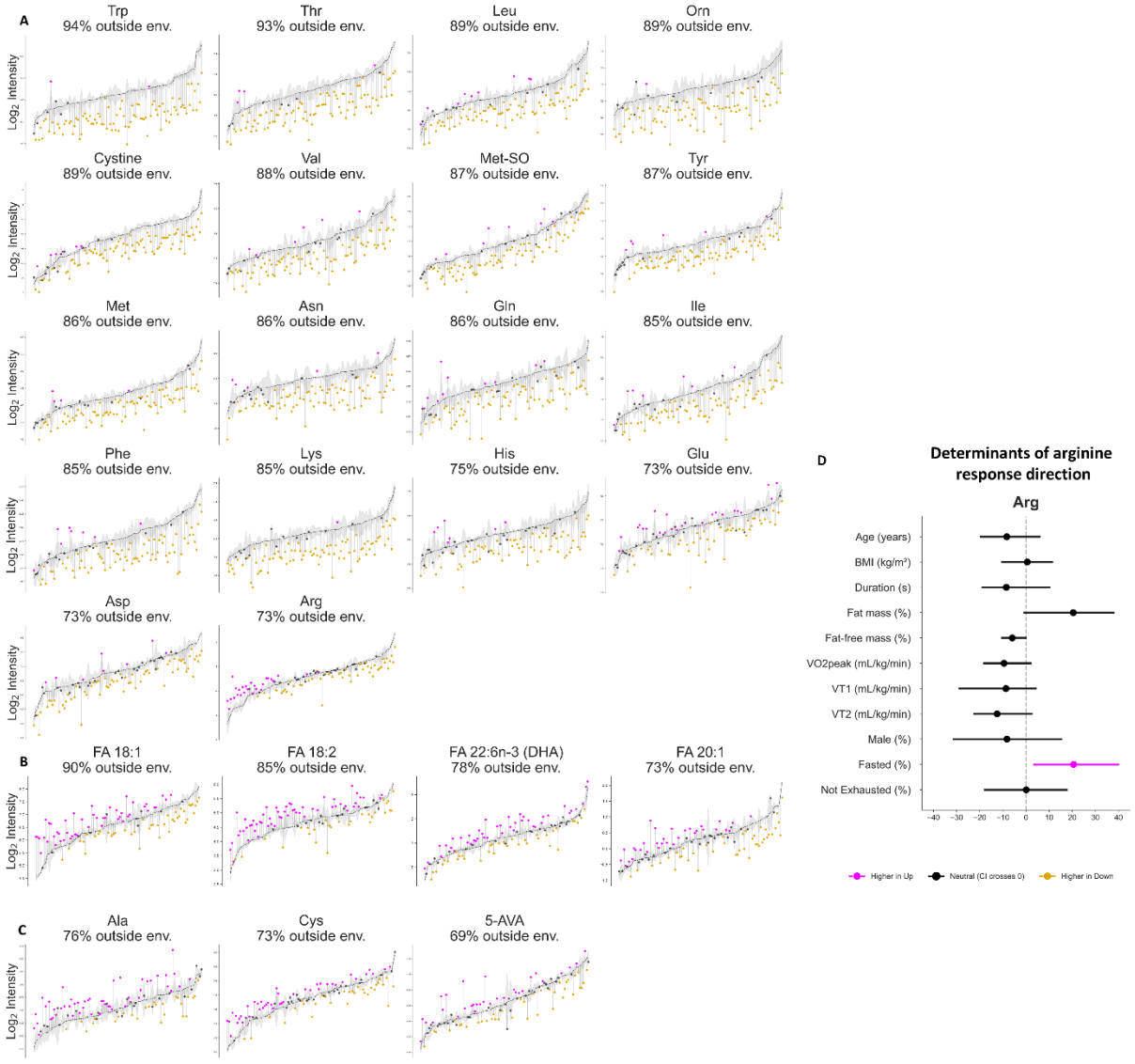


Figure 6.4. Directional response patterns relative to the individual baseline envelope. A: *individual response trajectories for the 18 metabolites highlighted by exhaustion status or fasting status in the participant-level determinant analysis. Percentages indicate the proportion of participants whose post-exercise value fell outside the individual baseline envelope. Most selected metabolites showed largely unidirectional responses, with post-exercise values predominantly below the baseline envelope, whereas arginine showed a more mixed directional pattern.* ***B****: fatty acids with bidirectional response patterns, defined as at least 25 participants showing an increase and at least 25 participants showing a decrease beyond the individual baseline envelope.* ***C****: amino-acid-related metabolites with bidirectional response patterns according to the same criterion.* ***D****: association analysis between response direction and participant characteristics for arginine, the bidirectional metabolite with the clearest participant-level association. Points indicate effect estimates and horizontal lines indicate 95% confidence intervals. Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; FA 18:1, octadecenoic acid; FA 18:2, octadecadienoic acid; FA 20:1, eicosenoic acid; FA 22:6n-3 (DHA), docosahexaenoic acid; Gln, glutamine; Glu, glutamic acid; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Met-SO, methionine sulfoxide; Orn, ornithine; Phe, phenylalanine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Val, valine; 5-AVA, 5-aminovaleric acid*

6.6 Exploratory Responder Analysis Based on Post-Exercise Deviations

As an additional exploratory analysis, responder patterns were investigated using post-exercise deviations from baseline-derived reference ranges[16–18]. While the previous baseline-envelope analysis classified responses according to whether post-exercise values were above or below the individual interval defined by Baseline 1 and Baseline 2, this analysis focused on metabolite for which post-exercise deviations showed an intermediate distribution across participants.

The analysis was restricted to metabolites with robust exercise response, particularly those with SNR values greater than 1.5. Among these, metabolites with an intermediate proportion of post-exercise deviations were considered suitable for responder/non-responder stratification, because they provided two groups of sufficient size of comparison. This selection highlighted several metabolites that overlapped with those identified in the previous determinant analyses, including tryptophan, threonine, methionine, leucine, phenylalanine, isoleucine, ornithine, valine and methionine sulfoxide (Figure 6.5).

Participants were then divided into responder and non-responder groups for selected metabolites, and these groups were compared with respect to physiological and anthropometric characteristics. The clearest exploratory association was observed for tryptophan. Tryptophan responders showed higher peak RER compared with non-responders (1.15 vs 1.10, $p = 0.020$), suggesting that responder status may be linked to a stronger effective exercise stimulus. Tryptophan responders also showed higher VT1 (23.4 vs 17.6 mL min^{-1} kg^{-1}, $p = 0.049$), indicating a possible relationship with submaximal aerobic capacity.

Additional suggestive differences were observed for other metabolites and participant characteristics. For isoleucine, ornithine, methionine and threonine, responders tended to show longer exercise duration than non-responders. For threonine and isoleucine, responders also tended to have higher VO_2peak. Tryptophan responder status showed suggestive differences in body composition, with responders tending to have higher relative fat-free mass and lower relative fat mass. However, these associations were weaker and should be interpreted cautiously.

Overall, this exploratory responder analysis suggests that, for selected metabolites, individual post-exercise deviation patterns may be partly related to physiological performance and body composition. The strongest signal was observed for tryptophan, which appeared to be associated with markers of exercise intensity and aerobic capacity. Nevertheless, this analysis should be considered hypothesis-generating, because responder groups were defined post hoc and the observed associations were not part of the primary analysis framework.

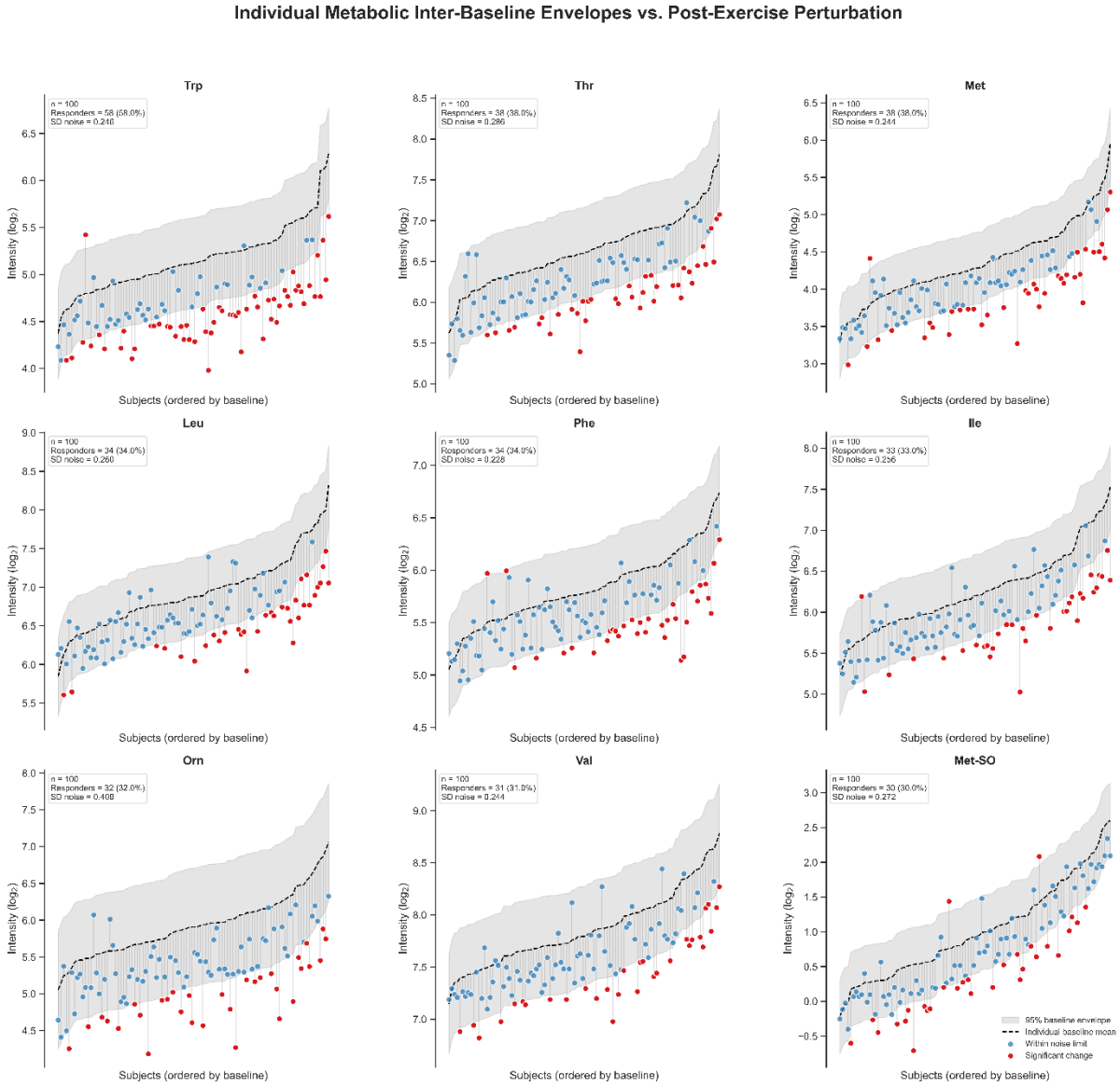


Figure 6.5. Exploratory responder patterns based on inter-baseline envelopes and post-exercise perturbation. *Individual baseline-envelope plots for selected metabolites used in the exploratory responder analysis. Subjects are ordered according to their individual baseline mean, shown by the dashed black line. The grey band represents the 95% inter-baseline envelope estimated from the repeated resting samples, while post-exercise values are shown as individual points. Red points indicate participants classified as responders, whose post-exercise value fell outside the baseline-derived envelope, whereas blue points indicate participants whose post-exercise value remained within the envelope. The selected metabolites showed intermediate responder proportions and SNR values greater than 1.5, allowing responder and non-responder groups of sufficient size to be compared in exploratory downstream analysis.*

6.7 Chapter Summary

This chapter investigated whether inter-individual variability in the noise-adjusted metabolite response could be partly explained by participant-level characteristics and individual response patterns. Building on the signal-to-noise analysis presented in the previous chapter, the analysis focused on metabolite whose post-exercise response exceeded resting baseline variability.

The participant-level determinant analysis showed that exhaustion status and fasting status were the most relevant variables associated with noise-adjusted metabolite responses. Exhaustion status showed the strongest association pattern, involving 16 metabolites, mainly amino acids and amino-acid-related metabolites. Participants classified as potentially not exhausted generally showed attenuated responses, indicating that incomplete attainment of maximal effort during CPET can reduce the magnitude of the acute metabolomic response.

Fasting status was associated with a smaller subset of 8 metabolites, again mainly involving amino-acid-related responses. The overlap between fasting-associated and exhaustion-associated metabolites suggested that part of the amino-acid-related response was sensitive to both the effective intensity of the exercise stimulus and the pre-exercise nutritional state. In contrast, VO_2peak, relative fat mass, age and sex showed fewer significant associations, indicating a more limited or metabolite-specific role in explaining response variability within this cohort.

The baseline-envelope analysis provided an additional individual-level interpretation of the post-exercise response. By comparing post-exercise values with each participant's repeated-baseline range, this analysis showed that several metabolites frequently exceeded individual resting variability. It also identified a smaller subset of bidirectional metabolites, for which some participants showed increasing response while others showed decreasing responses. Among these, arginine was the clearest interpretable case, with response direction associated with fasting status.

Finally, an exploratory responder analysis was performed for selected metabolites with robust responses and intermediate responder proportions. This analysis suggested that responder status for some metabolites, especially tryptophan, may be related to physiological performance variables such as peak RER and VT1. However, these findings were considered hypothesis-generating because responder groups were defined post hoc and the analysis was not part of the primary statistical framework[18].

Overall, this chapter showed that the acute metabolomic response to maximal exercise was not only metabolite-specific, but also participant-dependent. Experimental and physiological state variables, particularly exhaustion status and fasting status, explained part of the observed heterogeneity, while additional exploratory responder patterns suggested possible links with exercise intensity and aerobic capacity. These results provide a more detailed interpretation of individual metabolic response variability and prepare the transition to the next chapter, where baseline metabolomic and participant-level features are used to predict VO_2peak.

7 Prediction of VO_2peak from Baseline Metabolomic and Participant-Level Features

7.1 Prediction Task and Analytical Rationale

After characterizing baseline reliability, acute exercise responses and participant-level determinants of metabolite variability, the final analytical step of this thesis focused on prediction. While the previous chapters investigated how the capillary metabolome changed after exercise and which factors influenced this response, the predictive analysis asked a different question: whether baseline metabolomic and participant-level features could be used to estimate cardiorespiratory fitness.

VO_2peak was selected as the prediction target because it represents the main physiological marker of aerobic capacity obtained from the cardiopulmonary exercise test. In this analysis, VO_2peak was treated as a continuous outcome and the prediction task was formulated as a supervised regression problem[39–41]. The objective was not to develop a clinical diagnostic model, but to explore whether the resting capillary metabolome contained information related to individual fitness status and whether this information improved prediction beyond standard participant-level variables.

Only baseline metabolomic features were used for the metabolomics branch of the analysis. Specifically, Baseline 1 concentrations were used to evaluate whether the resting metabolomic profile contained information about VO_2peak before the exercise stimulus. This choice avoided using post-exercise values or exercise-induced changes, which would be directly influenced by the CPET protocol and therefore less suitable for pre-exercise prediction.

The metabolomic predictor set was restricted to a subset of 18 metabolites highlighted by the participant-level determinant analyses in Chapter 6. This subset included metabolites associated with exhaustion status, fasting status or both. The aim of this choice was to reduce the dimensionality of the metabolomic predictor space and to focus the prediction analysis on metabolites that had already shown relevance for exercise-response variability. Importantly, these metabolites were selected from the previous noise-adjusted response analyses, not from their predictive association with VO_2peak.

Participants classified as potentially not exhausted were excluded from the prediction dataset. This resulted in an effective prediction cohort of 77 participants. The exclusion was applied because incomplete attainment of maximal effort may lead to an unreliable estimate of VO_2peak. Since VO_2peak was the target variable of the predictive model, including participants with potentially underestimated values could weaken the relationship between predictors and outcome. At the same time, this choice reduced the sample size available for model training and validation, and therefore the predictive results were interpreted as exploratory[41].

The prediction analysis compared different sources of information. Participant-level clinical features were used to represent standard demographic, anthropometric and activity-related variables. The selected baseline metabolite subset was used to evaluate the predictive contribution of the resting capillary metabolome. Finally, combined modelling strategies were used to assess whether metabolomic information added value when integrated with participant-level features[40,41].

Overall, this chapter extends the thesis from descriptive and inferential metabolomics toward predictive modelling. By evaluating the prediction of VO_2peak, the analysis connects the capillary metabolome with a functional physiological outcome and assesses whether metabolomic data can contribute to the data-driven characterization of cardiorespiratory fitness.

7.2 Feature Sets and Modelling Strategy

To evaluate the predictive contribution of different types of information, several modelling branches were considered. The first branch included only participant-level clinical and physiological variables. These variables were age, sex, body mass index, relative fat mass, relative fat-free mass and weekly

endurance sport frequency. This model was used as the clinical reference model, representing information that can be collected without metabolomic profiling.

The second branch included only the selected baseline metabolomic variables. For this branch of the analysis, Baseline 1 concentrations of the 18 metabolites highlighted by the previous determinant analyses were used as candidate predictors. This allowed the model to evaluate whether a focused subset of response-relevant resting metabolites contained predictive information about VO_2peak before the exercise stimulus.

A third branch combined clinical and metabolomic information through an ensemble approach. Clinical-only and metabolomics-only predictions were first generated separately and were then combined using an optimized weighted average. The ensemble weight was tuned within the training data, so that the validation fold remained independent from the weight-selection procedure. This approach allowed the relative contribution of the clinical and metabolomic branches to be evaluated while reducing information leakage.

In addition to the ensemble model, a residual omics strategy was also tested. In this approach, a clinical model was first used to generate out-of-fold predictions within the training data. The residual component, defined as the difference between observed VO_2peak and the clinical prediction, represented the part of VO_2peak not explained by the clinical variables. A metabolomics model was then trained to predict this residual component from the selected baseline metabolite subset. This analysis tested whether baseline metabolites captured information about VO_2peak that was not already explained by participant-level clinical features.

ElasticNet regression was used as the main modelling approach. ElasticNet combines L1 and L2 regularization and is suitable for datasets with correlated predictors and a limited number of observations[39,40]. This was appropriate for the present analysis because metabolomic features can be correlated with each other, and the number of available participants was limited after excluding potentially not-exhausted individuals.

For the metabolomics branch, feature selection was performed within the cross-validation procedure. A stability-selection approach was used to identify metabolites that were repeatedly selected across internal resampling runs [42]. Selection was performed only within the current training data of each outer validation split, preventing information from the validation fold from influencing the metabolite selection step.

Model performance was evaluated using a repeated nested cross-validation framework. The outer validation loop was used to estimate out-of-sample predictive performance, while model tuning, metabolomic feature selection and ensemble-weight selection were performed only within the training data. RepeatedKFold was used because each participant appeared only once in the subject-level prediction matrix. This structure was used to reduce information leakage and to avoid optimistic performance estimates due to feature selection or hyperparameter tuning performed outside the validation loop[41].

Clinical missing values were imputed within the cross-validation procedure using training-set medians, so that information from the validation fold was not used during preprocessing. Continuous predictors were treated as numerical variables, while categorical variables were encoded before modelling. The same general preprocessing logic was applied consistently within each training fold to preserve the independence of the validation data.

Prediction performance was assessed using out-of-fold predictions. The main evaluation metrics were the coefficient of determination, mean absolute error and root mean squared error. These metrics were used together because they provide complementary information: R^2 describes the proportion of variance

explained by the model, MAE gives the average absolute prediction error, and RMSE gives greater weight to larger errors.

As a sensitivity analysis, the same prediction workflow was repeated using the full panel of 52 retained metabolites instead of the selected 18-metabolite subset. This analysis was used to evaluate whether the limited predictive contribution of the metabolomic branch in the main analysis was due to the restricted feature set. The full-panel analysis was interpreted as a secondary comparison, whereas the 18-metabolite model remained the main analysis because it was directly linked to the participant-level determinant results from Chapter 6

Overall, this modelling strategy was designed to balance prediction accuracy, interpretability and protection against overfitting. The clinical-only model provided a reference based on standard participant-level variables, the metabolomics-only model tested the predictive value of the selected resting metabolite subset, the ensemble model evaluated whether the two branches improved prediction when combined, and the residual omics model tested whether metabolites explained variance not already captured by clinical features.

7.3 Predictive Performance

Predictive performance was evaluated using out-of-fold predictions from the repeated cross-validation procedure. This allowed model performance to be estimated on held-out data and reduced the risk of overly optimistic results due to model tuning or feature selection performed on the full dataset.

The clinical-only model showed the best overall predictive performance. Using age, sex, body mass index, relative fat mass, relative fat-free mass and weekly endurance sport frequency, the model achieved an R^2 of 0.606, a mean absolute error of 4.322 $mL \cdot min^{-1} \cdot kg^{-1}$ and a root mean squared error of 5.293 $mL \cdot min^{-1} \cdot kg^{-1}$. This indicates that standard participant-level features captured a substantial proportion of the variance in VO_2peak within the exhausted subset of the cohort.

The clinical plus omics ensemble model showed nearly identical, but slightly lower, performance compared with the clinical-only model. The ensemble achieved an R^2 of 0.605, a mean absolute error of 4.330 $mL \cdot min^{-1} \cdot kg^{-1}$ and a root mean squared error of 5.300 $mL \cdot min^{-1} \cdot kg^{-1}$. The mean ensemble weights strongly favoured the clinical branch, with an average clinical weight of 0.91 and an average omics weight of 0.09. This indicates that the selected baseline metabolite subset contributed only marginally to the combined prediction.

The clinical plus residual omics model also did not improve performance beyond the clinical-only model. This model achieved an R^2 of 0.578, a mean absolute error of 4.421 $mL \cdot min^{-1} \cdot kg^{-1}$ and a root mean squared error of 5.480 $mL \cdot min^{-1} \cdot kg^{-1}$. The residual strategy was designed to test whether baseline metabolites could explain the part of VO_2peak that was not already captured by clinical variables. However, the lower performance suggests that the selected metabolite subset did not provide a strong independent residual signal in this cohort.

The omics-only model performed poorly. Using only the selected 18 baseline metabolites, the model achieved an R^2 of -0.044, a mean absolute error of 6.626 $mL \cdot min^{-1} \cdot kg^{-1}$ and a root mean squared error of 8.612 $mL \cdot min^{-1} \cdot kg^{-1}$. The negative R^2 indicates that the omics-only model performed worse than a simple prediction based on the mean VO_2peak value. Therefore, the selected resting metabolite subset was not sufficient on its own to estimate cardiorespiratory fitness.

Fold-wise results showed substantial variability across validation folds, especially for R^2. This variability is expected in a small dataset and confirms that the predictive results should be interpreted as exploratory rather than definitive. Despite this variability, the overall pattern was consistent: clinical features provided the dominant predictive information, while the selected baseline metabolite subset had limited standalone and incremental predictive value. The overall predictive performance of the four modelling strategies is summarized in Figure 7.1.

Overall, the prediction analysis showed that VO_2peak could be estimated with moderate accuracy from participant-level clinical features. In contrast, the selected baseline metabolite subset did not meaningfully improve prediction when added to clinical variables and performed poorly when used alone. These findings suggest that, within the present cohort and modelling framework, standard physiological and anthropometric features were more informative for VO_2peak prediction than the selected baseline capillary metabolite profile. Because the main metabolomic model was restricted to 18 metabolites selected from the previous response analyses, an additional sensitivity analysis was performed using the full retained panel of 52 metabolites. This analysis is presented in Section 7.5.

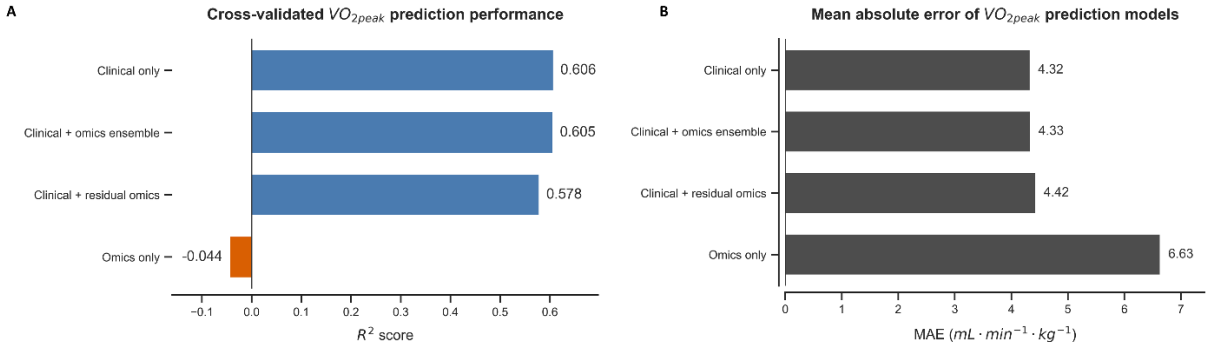


***Figure 7.1. Cross-validated predictive performance of VO_2peak models. A**: R^2 score for each model. The clinical-only model achieved the highest performance, closely followed by the clinical plus omics ensemble. The clinical plus residual omics model showed slightly lower performance, while the omics-only model had negative R^2, indicating poor standalone predictive ability. **B**: Mean absolute error for each model. The clinical-only and clinical plus omics ensemble models showed nearly identical prediction errors, whereas the omics-only model showed substantially higher error.*

7.4 Feature Contribution and Stable Predictors

After evaluating predictive performance, feature contribution was examined to understand which variables supported VO_2peak prediction. This step was important because the clinical-only model showed the strongest performance, whereas the metabolomic models provided limited additional predictive value.

Clinical feature contribution was evaluated using a leave-one-feature-out approach. In this analysis, the clinical model was refitted after removing one feature at a time, and the reduction in R^2 was used as an estimate of feature importance. Removing body mass index caused the largest decrease in performance, with an R^2 loss of 0.116. Removing sex produced the second largest decrease, with an R^2loss of 0.078, followed by weekly endurance sport frequency, with an R^2 loss of 0.060.

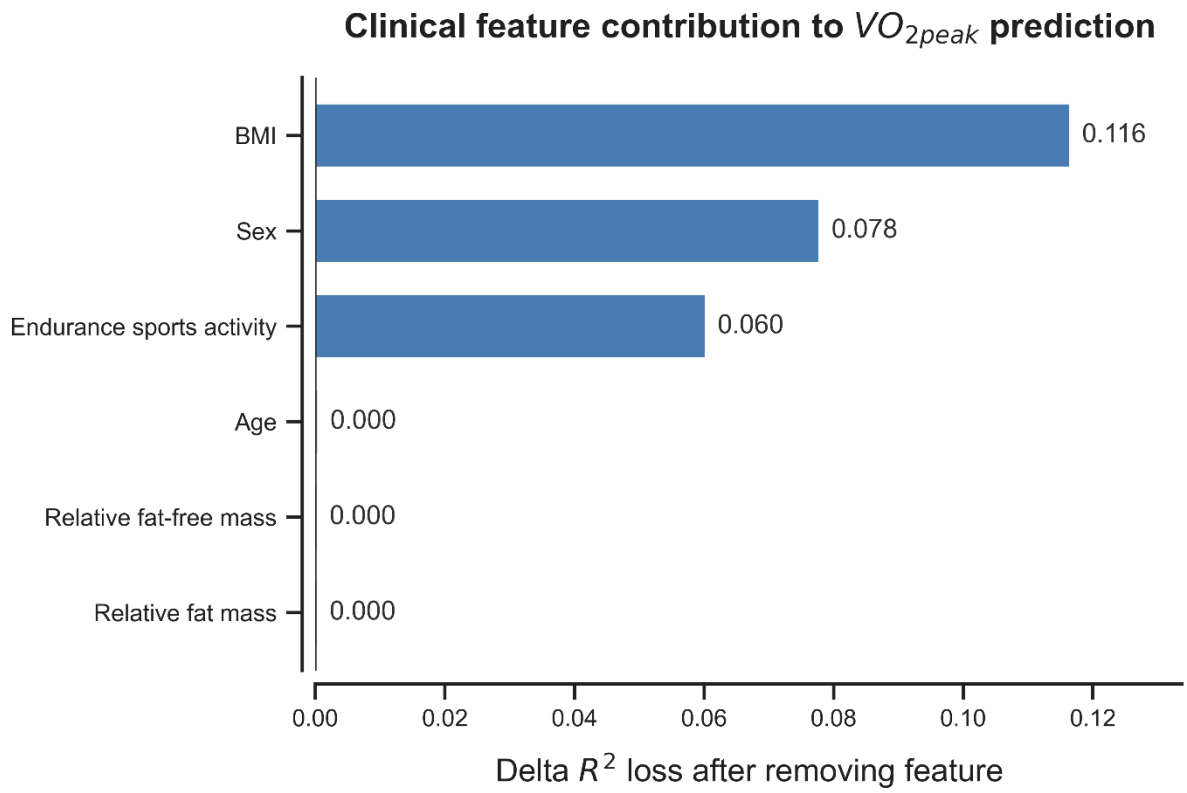


***Figure 7.2. Clinical feature contribution to VO_2peak prediction**. Leave-one-feature-out analysis for the clinical-only model. Bars represent the loss in R^2 after removing each participant-level feature from the model. Larger values indicate a stronger contribution to predictive performance. Body mass index showed the largest contribution, followed by sex and endurance sports activity. In contrast, age, relative-fat-free mass and relative fat mass produced almost no reduction in R^2 when removed, suggesting limited independent predictive contribution beyond the other clinical variables in this cohort.*

In contrast, removing age, relative fat-free mass or relative fat mass produced almost no loss in R^2. This does not necessarily mean that these variables are biologically irrelevant, but rather that they did not add independent predictive information beyond the other clinical features in this dataset. The limited contribution of age is plausible because the cohort consisted mainly of young adults, reducing the range of age-related variability. The limited independent contribution of relative fat mass and relative fat-free mass may also reflect overlap with body mass index and sex.

The metabolomic branch was interpreted by examining the frequency with which each metabolite was selected across outer folds. In the omics-only branch, asparagine was the most stable metabolite, selected in 98% of folds, followed by histidine, selected in 86% of folds. Arginine and methionine were each selected in 56% of folds, while lysine was selected in 50%. Other metabolites were selected less consistently, suggesting weaker or less stable contributions to the omics-only prediction.

The residual omics models showed a different stability pattern. Methionine sulfoxide, cystine, histidine, arginine, tyrosine, threonine and asparagine were selected frequently when the metabolomic model was trained to predict the clinical residual component. However, because the residual omics model did not improve prediction compared with the clinical-only model, these selection frequencies should not be interpreted as evidence of robust independent biomarkers. Rather, they indicate which metabolites were most often retained by the regularized model within an exploratory residual framework.

Overall, the feature contribution analysis confirmed that VO_2peak prediction was mainly driven by clinical and participant-level variables, especially body mass index, sex and weekly endurance sport frequency. The selected baseline metabolite subset showed some internal feature-selection structure, but this did not translate into improved predictive performance. Therefore, metabolomic predictors should

be interpreted cautiously and as exploratory in the absence of stronger incremental performance and external validation.



***Figure 7.3. Stability of selected metabolite predictors. A**: Omics-only model. Asparagine and histidine were the most consistently selected metabolites, followed by methionine, arginine and lysine. **B**: residual omics model. Methionine sulfoxide, histidine, cystine, tyrosine, arginine and threonine were frequently selected when the metabolomic model was trained to predict the residual component not explained by clinical variables. Because the metabolomic branches did not improve predictive performance beyond the clinical-only model, these metabolites should be interpreted as exploratory model-selected features rather than validated biomarkers of VO_2peak.*

7.5 Sensitivity Analysis Using the Full Metabolite Panel

To evaluate whether the limited predictive contribution of the metabolomic branch depended on the restricted 18-metabolite subset, the prediction workflow was repeated using the full panel of 52 retained metabolites. This sensitivity analysis used the same exhausted prediction cohort and the same modelling framework branch to select predictors from the complete targeted metabolite panel.

Using all 52 metabolites slightly improved the performance of the omics-only model compared with the main 18-metabolite analysis. The omics-only model achieved an R^2 of 0.147, a mean absolute error of 6.171 $mL \cdot min^{-1} \cdot kg^{-1}$ and a root mean squared error of 7.788 $mL \cdot min^{-1} \cdot kg^{-1}$. This was better than the omics-only performance obtained with the selected 18-metabolite subset, but remained substantially weaker than the clinical-only model. Therefore, the full metabolite panel contained more predictive information than the restricted subset, but this information was not sufficient to provide accurate standalone prediction of VO_2peak.

The clinical plus omics ensemble model using all 52 metabolites achieved an R^2 of 0.606, a mean absolute error of 4.367 $mL \cdot min^{-1} \cdot kg^{-1}$ and a root mean squared error of 5.292 $mL \cdot min^{-1} \cdot kg^{-1}$. This performance was almost identical to the clinical-only model, which achieved an R^2 of 0.606, a mean absolute error of 4.322 $mL \cdot min^{-1} \cdot kg^{-1}$ and a root mean squared error of 5.293 $mL \cdot min^{-1} \cdot kg^{-1}$. Although the ensemble model assigned a higher mean weight to the omics branch than the main 18-metabolite analysis, with an average clinical weight of 0.73 and an average omics weight of 0.27, this did not translate into a meaningful improvement in predictive performance.

The clinical plus residual omics model also did not improve prediction. Using the full 52-metabolite panel, the residual omics model achieved an R^2 of 0.554, a mean absolute error of 4.519 $mL \cdot min^{-1} \cdot kg^{-1}$ and a root mean squared error of 5.633 $mL \cdot min^{-1} \cdot kg^{-1}$. This was lower than the clinical-only model, suggesting that the full baseline metabolite panel did not capture a strong residual component of VO_2peak after accounting for clinical variables.

Model	Feature set	R^2	MAE	RMSE	Interpretation
Clinical only	**Clinical**	**0.606**	**4.322**	**5.293**	**Best reference model**
Clinical + omics ensemble	Clinical + 18 metabolites	0.605	4.330	5.300	No improvement
Clinical + residual omics	Clinical + 18 metabolites	0.578	4.421	5.480	Lower than clinical
Omics only	18 metabolites	-0.044	6.626	8.612	Poor standalone prediction
Clinical + omics ensemble	Clinical + 52 metabolites	0.606	4.367	5.292	Similar to clinical
Clinical + residual omics	Clinical + 52 metabolites	0.554	4.519	5.633	Lower than clinical
Omics only	52 metabolites	0.147	6.171	7.788	Improved vs 18, still weak

Table 7.1. Sensitivity analysis comparing selected and full metabolite panels for VO_2peak prediction. *Out-of-fold predictive performance of the main 18-metabolite analysis and the full 52-metabolite sensitivity analysis. The clinical-only model was used as the reference. Models combining clinical and metabolomic features showed performance similar to, but not meaningfully better than, the clinical-only model. The full metabolite panel improved the omics-only model compared with the selected 18-metabolite subset, but standalone metabolomic prediction remained substantially weaker than clinical prediction. MAE and RMSE are reported in $mL \cdot min^{-1} \cdot kg^{-1}$.*

The stability profile of selected metabolites differed from the main 18-metabolite analysis. In the full-panel omics-only model, arachidonic acid, sarcosine, taurine, eicosapentaenoic acid and asparagine were selected most consistently across folds. Arachidonic acid and sarcosine were selected in 100% of folds, followed by taurine, eicosapentaenoic acid and asparagine. In the residual omics model, arachidonic acid, alpha-aminoadipic acid, histidine, methionine sulfoxide, serine, glycine and taurine were among the most frequently selected metabolites. However, these selection frequencies should be interpreted cautiously, because the metabolomic models did not improve overall prediction beyond the clinical-only model.

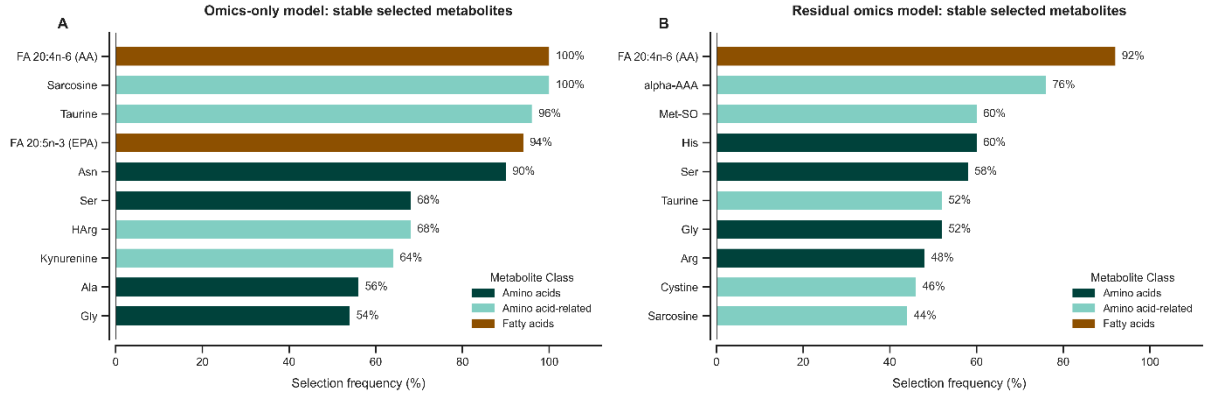


Figure 7.4. Stability of selected metabolite predictors in the full-panel sensitivity analysis. A*: Omics-only model. Arachidonic acid, sarcosine, taurine, eicosapentaenoic acid and asparagine were the most consistently selected predictors.* ***B****: Residual omics model. Arachidonic acid, alpha-aminoadipic acid, histidine, methionine sulfoxide, serine, glycine and taurine were among the most frequently selected metabolites when the metabolomic model was trained to predict the residual component not explained by clinical variables.*

7.5.1 Repeated Cross-Validation Variability

The full-panel sensitivity analysis was evaluated across 50 outer validation splits generated through repeated five-fold cross-validation. Reporting the variability across repeated outer splits provides an additional perspective on the uncertainty of model performance in a limited-size cohort.

For the clinical-only model, the mean fold-level R^2 was 0.518 with a standard deviation of 0.291. The clinical plus omics ensemble showed a similar mean fold-level R^2 of 0.535 ± 0.233. In contrast, the omics-only model showed a substantially lower and more variable fold-level performance, with a mean R^2 of 0.039 ± 0.415. The residual omics model also displayed substantial variability across validation splits.

These results should be interpreted carefully. Fold-level R^2 values are inherently variable when the held-out test partitions are small, because a limited number of prediction errors can strongly affect the score in an individual split. Therefore, the pooled out-of-fold metrics reported in Table 7.1 remain the primary estimate of overall performance. Nevertheless, the repeated-split analysis confirms the same

general pattern observed in the pooled results: clinical features provided the most stable predictive information, whereas metabolomic models showed greater uncertainty and did not consistently improve prediction beyond the clinical reference model.

Model	Mean fold R^2	SD	Minimum	Maximum
Clinical only	0.518	0.291	- 0.465	0.878
Clinical + omics ensemble	0.535	0.233	- 0.336	0.886
Clinical + residual omics	0.422	0.504	-2.460	0.884
Omics only	0.039	0.415	-1.956	0.586

Overall, the full-panel sensitivity analysis confirmed the main conclusion of the prediction chapter. Expanding the metabolomic feature set from 18 to 52 metabolites improved the standalone omics model but did not provide meaningful incremental value beyond standard participant-level features. Therefore, the limited contribution of metabolomics to VO_2peak prediction was not only a consequence of using the restricted 18-metabolite subset. Rather, within this cohort and modelling framework, baseline metabolomic information remained secondary to clinical, anthropometric and activity-related variables.

7.6 Interpretation and Limitations of the Prediction Analysis

The prediction analysis showed that VO_2peak could be estimated with moderate accuracy using participant-level clinical features, whereas the selected baseline metabolite subset had limited standalone and incremental predictive value. This result indicates that, in the present cohort, standard demographic, anthropometric and activity-related variables captured most of the information relevant to VO_2peak prediction.

The strong performance of the clinical-only model is physiologically plausible. VO_2peak is influenced by factors such as sex, body size, body composition and habitual endurance training, these variables were directly represented in the clinical feature set, which explains why the clinical model performed well despite the limited sample size. The leave-one-feature-out analysis further supported this interpretation, showing that body mass index, sex and endurance sports activity were the main contributors to predictive performance.

In contrast, the selected baseline metabolite subset did not provide strong independent predictive information. the omics-only model performed poorly, indicating that the 18 selected resting metabolites were not sufficient on their own to estimate VO_2peak. Moreover, adding the metabolomic branch to the clinical model did not improve performance, as shown by the nearly identical results of the clinical-only and clinical plus omics ensemble models. The ensemble weights also confirmed this pattern, with most of the weight assigned to the clinical branch and only a small contribution assigned to the omics branch.

The residual omics analysis provided an additional test of whether baseline metabolites captured information not already explained by clinical variables. This model did not improve prediction compared with the clinical-only model. Therefore, within this dataset, the selected metabolite subset did not appear to explain a substantial residual component of VO_2peak after accounting for standard participant-level features.

Several factors may explain the limited predictive value of the metabolomic features. First, the metabolomic branch was restricted to 18 metabolites selected from previous response analyses, rather than using the full retained metabolite panel. This choice reduced dimensionality and avoided an unrestricted search across many predictors, but it may also have excluded metabolites with stronger baseline associations with VO_2peak. Second, the selected metabolites were originally identified

because of their relevance to exercise-response variability, exhaustion status and fasting status, not because of their direct relationship with cardiorespiratory fitness.

Third, only baseline metabolite values were used for the main prediction task. This choice was conceptually important because it allowed pre-exercise prediction of VO_2peak, but it may have limited the amount of physiological information available to the model. Exercise-response features, such as post-exercise values or delta changes, may be more directly-connected to CPET performance because they reflect the dynamic metabolic response to the exercise stimulus. however, such features are less useful for true pre-exercise prediction, since they require completion of the exercise test.

The prediction analysis was also limited by sample size. After excluding participants classified as potentially not exhausted, the effective prediction cohort included 77 participants. This exclusion was methodologically justified because unreliable VO_2peak estimates could distort model training, but it reduced the number of observations available for cross-validation. In small biomedical datasets, model performance estimates can vary substantially across folds, and feature selection can be unstable. This was reflected by the fold-wise variability observed in the prediction results[41,42].

Sex imbalance represents an additional limitation of the prediction analysis. The original cohort included 68 males and 32 females, and after excluding potentially not-exhausted participants the effective prediction cohort still remained unbalanced, with 51 males and 26 females. This is relevant because VO_2peak is strongly influenced by sex-related physiological differences, and the leave-one-feature-out analysis confirmed that sex contributed substantially to predictive performance. From a modelling perspective, a potentially more appropriate strategy would have been to train sex-specific prediction models. However, the female subgroup was too small to support a reliable independent model, and preliminary sex-stratified analyses were unstable, particularly in females. The male-only model showed more stable performance, but the sample size was still limited. For this reason, the main analysis used a single combined cohort and included sex as a predictor, while acknowledging that the resulting model may partly reflect the sex distribution of the dataset.

Finally, the interpretation of metabolomic predictors should remain cautious. Some metabolites were selected frequently across folds, particularly asparagine and histidine in the omics-only model and methionine sulfoxide, cystine, histidine, arginine and tyrosine in the residual omics model. However, because the metabolomic branches did not improve predictive performance, these metabolites should not be interpreted as validated biomarkers of VO_2peak. They are better understood as exploratory model-selected features within the present dataset.

Overall, the prediction analysis suggests that participant-level clinical variables are more informative than the selected baseline metabolite subset for estimating VO_2peak in this cohort. The resting capillary metabolome may still contain information related to cardiorespiratory fitness, but this information was not strong enough to improve prediction beyond standard clinical features in the current modelling framework. These findings highlight both the potential and the limitations of metabolomics-based prediction in small exercise physiology datasets.

7.7 Chapter Summary

This chapter evaluated whether VO_2peak could be predicted from baseline metabolomic and participant-level features. The prediction analysis was performed after excluding participants classified as potentially not exhausted, resulting in an effective prediction cohort of 77 participants. This exclusion was applied to reduce the influence of potentially unreliable VO_2peak values on model training and evaluation.

The main analysis used a selected subset of 18 baseline metabolites identifies in the previous participant-level determinant analyses, together with clinical and physiological variables. Different modelling strategies were compared, including clinical-only, omics-only, clinical plus omics ensemble and clinical

plus residual omics models. The aim was to evaluate not only predictive performance, but also whether baseline metabolomic information added value beyond standard participant-level features.

The clinical-only model showed the best overall predictive performance, with an R^2 of approximately 0.606 and a mean absolute error of about 4.3 $mL \cdot min^{-1} \cdot kg^{-1}$. Body mass index, sex and endurance sports activity were the main contributors to this performance, as shown by the leave-one-feature-out analysis. In contrast, the omics-only model based on the selected 18 metabolites performed poorly, indicating that this restricted baseline metabolite subset was not sufficient on its own to estimate VO_2peak.

Adding metabolomic information to the clinical model did not meaningfully improve prediction. The clinical plus omics ensemble model performed almost identically to the clinical-only model and assigned most of the predictive weight to the clinical branch. Similarly, the residual omics strategy did not improve performance, suggesting that the selected baseline metabolites did not capture a strong component of VO_2peak variability beyond that already explained by clinical features.

A sensitivity analysis using the full panel of 52 retained metabolites was then performed to test whether the limited omics contribution was due to the restricted 18-metabolite subset. The full metabolite panel improved the omics-only model compared with the selected subset, but standalone metabolomic prediction remained substantially weaker than clinical prediction. Moreover, combining the full metabolite panel with clinical variables still did not improve performance beyond the clinical-only model.

Overall, the prediction results indicate that VO_2peak was mainly predictable from participant-level clinical and physiological variables in this cohort, whereas baseline metabolomic information provided limited incremental value. The resting capillary metabolome may still contain information related to cardiorespiratory fitness, but this information was not strong enough to improve prediction beyond standard variables in the present dataset and modelling framework.

These findings should be interpreted cautiously. The effective prediction sample size was limited, the cohort remained unbalanced by sex after excluding not-exhausted participants, and no independent external validation cohort was available. Therefore, the prediction analysis should be considered exploratory and cohort-specific. Nevertheless, it provides a useful data-science extension of the thesis by testing whether the capillary metabolome can contribute to the estimation of a functional physiological outcome.

8 Discussion

8.1 Main Findings

This thesis investigated the relationship between the capillary blood metabolome, acute exercise response and cardiorespiratory fitness using a repeated-baseline metabolomics design. The central objective was to evaluate whether two resting baseline samples could improve the interpretation of post-exercise metabolite changes by providing an estimate of baseline-derived methodological noise. In addition, the thesis examined whether participant-level characteristics could explain inter-individual variability in the metabolomic response and whether baseline metabolomic features could contribute to the prediction of VO_2peak.

The first main finding was that the capillary blood metabolome showed good short-term reproducibility under resting conditions. The two baseline samples were strongly correlated, most metabolites showed moderate-to-high baseline agreement, and dimensionality reduction suggested that individual metabolic profiles were partly preserved across repeated measurements. This supported the use of the repeated baseline as a reliable reference state for interpreting post-exercise changes.

The second main finding was that maximal exercise induced a broad and structured shift in the capillary blood metabolome. Principal component analysis showed a clear separation between resting and post-exercise samples, and the loading structure indicated that amino acids and amino acid-related metabolites contributed strongly to the main exercise-related variation. The conventional single-baseline analysis confirmed this broad response, identifying 40 out of 52 retained metabolites as significantly altered after exercise, with lactate showing the strongest post-exercise change.

However, the repeated-baseline analysis showed that statistical significance alone was not sufficient to evaluate response robustness. By comparing exercise reactivity with baseline-derived methodological noise, the signal-to-noise analysis stratified metabolites according to the magnitude of their response relative to resting variability. This showed that not all statistically significant metabolites had the same interpretability at the individual level. Some metabolites showed responses that clearly exceeded baseline variability, whereas others showed changes closer to the expected resting fluctuation.

The third main finding was that inter-individual variability in the noise-adjusted metabolite response was partly explained by experimental and physiological state variables. Exhaustion status emerged as the strongest determinant, with participants classified as potentially not exhausted showing attenuated amino acid-related response. Fasting status also influenced a smaller subset of metabolites, several of which overlapped with those associated with exhaustion status. In contrast, VO_2peak, relative fat mass, age and sex showed more limited associations with the acute noise-adjusted response.

The fourth main finding was that individual response patterns contained additional heterogeneity beyond group-level changes. The baseline-envelope analysis showed that several metabolites frequently exceeded individual resting ranges, while a smaller subset showed bidirectional response. Among bidirectional metabolites, arginine showed the clearest association with fasting status. A additional exploratory responder analysis suggested that tryptophan response status may be related to physiological performance variables such as peak RER and VT1, although these findings should be interpreted as hypothesis-generating.

Finally, the prediction analysis showed that VO_2peak could be estimated with moderate accuracy from participant-level clinical features, but that baseline metabolomic features provided limited incremental predictive value. The clinical-only model performed best, whereas the omics-only model based on the selected 18-metabolite subset performed poorly. Adding metabolomic information through ensemble or residual omics strategies did not meaningfully improve prediction. A sensitivity analysis using the full panel of 52 retained metabolites improved the standalone omics model but still did not outperform or improve upon the clinical-only model.

Overall, these findings support the value of repeated-baseline capillary blood metabolomics for interpreting acute exercise response. The main contribution of the thesis is methodological rather than purely predictive: the repeated-baseline design allowed post-exercise changes to be interpreted relative to baseline-derived methodological noise, improving the distinction between broad group-level changes and more robust individual-level responses. At the same time, the prediction results indicate that baseline metabolomic information, at least in this cohort and modelling framework, was less informative for VO_2peak estimation than standard clinical and physiological variables.

8.2 Repeated Baseline Design and Methodological Noise

A central contribution of this thesis is the use of repeated resting baseline measurements to interpret acute exercise-induced metabolite changes. In many exercise metabolomics studies. Post exercise values are compared with a single pre-exercise sample. This approach is useful for identifying group-level differences, but it assumes that the baseline measurement provides a stable and sufficient representation of the resting state. In practice, metabolite concentrations may vary over short time intervals because of sampling conditions, analytical measurement, sample preparation and short-term biological fluctuation. Therefore, a single-baseline comparison cannot determine whether an observed post-exercise change is clearly larger than the variability that would be expected at rest[14–18].

The repeated-baseline design addressed this limitation by providing two independent resting measurements before the exercise stimulus. the strong agreement observed between Baseline 1 and Baseline 2 indicated that the capillary metabolome was generally reproducible in the present cohort. The global correlation between the two baseline samples was very high, and most metabolites showed moderate-to-high metabolite-level agreement. In addition, the t-SNE visualization suggested that repeated samples preserved subject-specific metabolic profiles. Together, these findings support the interpretation that the repeated baseline captured a relatively stable resting reference state rather than random measurement variation.

However, the baseline results also showed that resting stability was not identical across all metabolites. While the overall metabolomic profiles was reproducible, some metabolites displayed greater short-term fluctuation or small systematic differences between Baseline 1 and Baseline 2. This observation is important because it supports the use of metabolite-specific estimates of variability rather than a single global threshold for all metabolites. A metabolite with a large natural or methodological fluctuation requires a larger post-exercise change before that response can be interpreted as robust, whereas a metabolite with a stable baseline profile may provide more reliable individual-level information even when its absolute response is smaller.

For this reason, the term methodological noise was used in this thesis as an operational description of baseline-derived short-term variability. This estimate does not represent purely analytical error. Instead, it includes the combined contribution of pre-analytical variation, analytical measurement variability and short-term resting biological fluctuation. These components cannot be fully separated using two baseline samples alone. Nevertheless, their combined estimate remains useful because it represents the variation that must be exceeded before a post-exercise response ca be interpreted with greater confidence.

The signal-to-noise framework translated this concept into a practical metabolite-ranking strategy. Rather than considering all statistically significant post-exercise changes as equally informative, the analysis compared exercise reactivity with the baseline-derived methodological noise of each metabolite. This distinction was particularly relevant because the conventional single-baseline analysis identified a broad set of significant metabolite changes whereas the SNR analysis showed that these changes differed substantially in magnitude relative to baseline variability. Therefore, the repeated-baseline approach did not replace conventional statistical testing, but complemented it by adding an individual-level reliability perspective.

This framework is particularly relevant for capillary blood metabolomics. Capillary sampling is practical and minimally invasive, making it suitable for repeated measurements during exercise protocols[9–12,14]. At the same time, the reduced sample volume and sensitivity to collection conditions make a careful evaluation of short-term variability especially important. The present findings suggest that capillary blood can provide a reliable matrix for repeated targeted metabolomics when the interpretation of exercise-induced changes accounts for baseline-derived methodological noise.

More broadly, the repeated-baseline design provides an example of how experimental design can improve downstream data analysis. Rather than treating repeated baseline samples only as a technical quality-control procedure, they were used here as a source of information for estimating the reliability of the observed exercise response. This is a useful data-science perspective because it incorporates uncertainty directly into the analytical framework. In the context of acute exercise metabolomics, this approach can help distinguish broad population-level changes from metabolite response that are more robust relative to the variability observed within individuals before exercise[16–18].

The baseline reproducibility observed in this thesis is consistent with the broader concept of metabolic individuality[34,43,44]. Previous studies have shown that individuals can preserve partially stable metabolomic profiles over repeated measurements, even when individual metabolites differ in their degree of temporal stability. This is important because repeated-baseline metabolomics does not require

every metabolite to be perfectly stable. Rather, it requires that the baseline state provides enough information to estimate metabolite-specific fluctuation before the experimental stimulus is applied[14,15].

The present findings therefore support two complementary interpretations. First, the high global agreement between baseline 1 and baseline 2 indicates that the capillary metabolomic profile was not dominated by random measurement variation. Second, the heterogeneity of metabolite-specific variability is precisely what justifies the signal-to-noise framework. If all metabolites had shown identical baseline stability, a single global correction factor might have been sufficient. Instead, the repeated-baseline design allowed each metabolite response to be evaluated against its own resting variability.

This point is particularly relevant for capillary blood sampling[9–13]. Capillary sampling is attractive in exercise physiology because it is minimally invasive and compatible with repeated measurements, but it may also be more sensitive to collection conditions, small-volume handling and local peripheral variation than standard venous sampling. The repeated-baseline design partly addresses this issue by shifting the interpretation from absolute concentration alone to within-subject change relative to a short-term resting reference. In this sense, the approach is well suited to capillary metabolomics because it uses the feasibility of repeated sampling as an analytical advantage.

It is also important to emphasize that baseline-derived methodological noise should not be interpreted as a limitation of the dataset alone. In acute exercise studies, short-term variability is an inherent part of the measurement context. By estimating it directly, the study makes the uncertainty visible and incorporates it into the interpretation. This strengthens the biological interpretation of high-SNR metabolites and provides a more cautious interpretation of metabolites whose post-exercise changes are closer to resting fluctuation[14–18].

8.3 Acute Exercise Response and Noise-Adjusted Metabolite Ranking

The acute exercise analyses showed that maximal CPET induced a broad and coordinated perturbation of the capillary metabolome. The separation of post-exercise samples from the two resting baseline samples in the PCA indicated that the exercise stimulus affected the global metabolic profile rather than only a small number of isolated metabolites. This global shift was primarily driven by amino acids and amino acid-related metabolites along the first principal component, whereas fatty acids contributed more strongly to a secondary axis of variation. This pattern is consistent with the systematic metabolic demand induced by maximal exercise, which involves rapid changes in substrate use, glycolytic activation, amino acid turnover and lipid-related regulation[1–6].

The stronger contribution of amino acids and amino acid-related metabolites to the first principal component suggests that this class captured a coordinated systemic response to maximal exercise. During intense exercise, amino acid concentrations may be affected by muscle uptake and release, transamination reactions, changes in protein turnover and altered substrate availability. Therefore, the observed post-exercise amino acid-related shift should not be interpreted as a single pathway response, but as the combined result of several processes activated during maximal physiological stress[23].

This interpretation also helps explain why amino acid-related metabolites appeared prominently in the determinant analysis. If these metabolites are sensitive to the intensity and completeness of the exercise stimulus, then participants who did not reach a comparable maximal effort should be expected to show attenuated responses. The overlap between high-SNR metabolites and exhaustion-associated metabolites therefore supports the idea that his class reflects not only exercise exposure, but also the effective intensity of the stimulus.

Fatt acids showed a different pattern. Their stronger contribution to the second principal component suggests that they were more involved in secondary inter-individual variability than in the dominant post-exercise separation. This is physiologically plausible because circulating fatty acids are strongly influenced by nutritional state, adipose tissue lipolysis, substrate availability and the balance between

carbohydrate and lipid metabolism. During maximal exercise, carbohydrate metabolites and glycolytic activation are expected to dominate the immediate response, whereas fatty acid-related signals may vary more according to participant state and pre-exercise conditions[24–29,46–48].

The distinction between amino acid-related metabolites and fatty acids is therefore useful for interpreting the structure of the dataset. Amino acids and amino acid-related compounds appeared to represent he main coordinated response axis to maximal CPET, whereas fatty acids contributed more to heterogeneity across participants. This does not imply that fatty acids were biologically unimportant. Rather, their contribution may be more context-dependent and may require more standardize fasting conditions or additional recovery timepoints to be interpreted more clearly.

The conventional single-baseline analysis provided a complementary metabolite-level view of this response. Comparing Post-exercise with Baseline 1 identified significant changes in 40 of the 52 retained metabolites, confirming that the exercise response was widespread across the targeted panel. Lactate showed the strongest post-exercise change, as expected for a maximal exercise protocol[22]. Its marked increase supports the physiological validity of the CPET stimulus and confirms that the dataset captured a pronounced acute metabolic perturbation

However, the broad number of significant metabolites also illustrates a limitation of conventional post-exercise comparisons. Statistical significance indicates that a metabolite changed consistently at the group level, but it does not directly indicate whether the magnitude of that response was large relative to the short-term variability of the same metabolite. This distinction became evident when the results of the single-baseline analysis were compared with the signal-to-noise ranking.

The SNR analysis showed that exercise responsiveness was distributed along a continuum. A small group of metabolites showed very strong responses, with post-exercise changes clearly exceeding baseline-derived methodological noise. Other metabolites showed moderate or limited responses, despite reaching statistical significance in the conventional comparison. This result suggests that exercise-induced metabolite changes should not be interpreted as a binary distinction between significant and non-significant metabolites. Instead, response strength should be considered relative to the stability of the resting metabolite profile.

This ranking approach is particularly useful for identifying metabolites that may be more reliable markers of acute exercise response. Metabolites with high SNR values are not necessarily the only biologically relevant metabolites, but they provide stronger evidence that the observed post-exercise perturbation is larger than the variability observed before exercise. In contrast, metabolites with lower SNR values may still participate in the exercise response, but their interpretation should remain more cautious because their average change is closer to the expected baseline-derived variability.

The comparison between PCA, conventional single-baseline testing and SNR stratification therefore provides a layered interpretation of the acute exercise response. PCA showed that maximal exercise induces a coordinated global shift in the metabolome. The single-baseline analysis identifies the metabolites contributing to this broad group-level response. The SNR framework then refines this interpretation by indicating which metabolite changes most clearly exceed baseline-derived methodological noise.

Ad additional implication of these findings is that amino acids and amino acid-related metabolites appear particularly relevant for characterizing the acute exercise response in this dataset[23]. Their strong contribution to the main PCA axis, together with the prominence of several amino acid-related metabolites in the high-SNR and determinant analyses, suggests that this metabolite class captured an important component of the systemic response to maximal CPET. Fatty acids, by contrast, appeared to contribute more to secondary variation and to heterogeneous individual patterns, rather than to the dominant global shift[24–29].

Overall, the results support the interpretation that maximal exercise produces a broad but uneven metabolomic response. Some metabolites show robust changes that clearly exceed methodological noise, while others show more modest or variable perturbations. The repeated-baseline approach makes this distinction visible and provides a more nuanced framework for interpreting acute exercise metabolomics than conventional significance testing alone.

8.4 Participant-Level Determinants of Response Variability

Beyond the identification of robust metabolite responses, an important objective of this thesis was to investigate why the magnitude of the acute metabolic response differed across participants. The determinant analysis showed that inter-individual variability was not explained uniformly by all participant characteristics[5,16,18]. Instead, exhaustion status and fasting status emerged as the most relevant variables associated with noise-adjusted metabolite responses, whereas VO_2peak, body composition, age and sex showed fewer significant associations after multiple-testing correction.

Exhaustion status showed the strongest association pattern. Participants classified as potentially not exhausted generally displayed attenuated responses across a set of amino acids and amino acid-related metabolites. This finding is physiologically plausible because participants who do not attain a sufficiently high effective exercise intensity may experience a smaller systemic metabolic perturbation. Although all participants completed the same general CPET protocol, the metabolic stimulus was not necessarily equivalent across individuals. Differences in effort attainment, respiratory exchange ratio or agreement between incremental and verification-bout VO_2peak values may influence the extent to which challenges substrate metabolism and amino acid turnover.

The role of exhaustion status deserves particular attention because it links the physiological validity of the CPET with the interpretation of the metabolomic response. In this thesis, exhaustion status was not treated only as a technical exclusion criterion for VO_2peak prediction, but also as a determinant of the acute biochemical response. This is important because a participant who does not achieve a sufficiently intense effort may produce a lower measured VO_2peak and, at the same time, a weaker post-exercise metabolomic perturbation.

This dual role has methodological consequences. If potentially not-exhausted participants are included in acute-response analyses without adjustment, the resulting variability may be incorrectly interpreted as inter-individual metabolic heterogeneity. Conversely, excluding all potentially not -exhausted participants from every analysis may remove biologically and experimentally informative variation. The approach used in this thesis was therefore intermediate: exhaustion status was retained in the response analyses as an explanatory variable, while the prediction analysis excluded potentially not-exhausted participants to reduce uncertainty in the target variable.

This distinction is relevant for future exercise metabolomics studies. When the exercise protocol is intended to induce a maximal metabolic perturbation, the quality and completeness of effort should be documented and incorporated into the statistical analysis. Variables such as peak RER, verification-bout agreement, test duration and ventilatory thresholds can help determine whether participants received a comparable physiological stimulus. without this information, differences in metabolite response magnitude may be difficult to separate from differences in actual exercise intensity[16–18].

This result has important implications for exercise metabolomics studies. Exhaustion status is often considered mainly as a quality-control criterion for interpreting maximal oxygen uptake. The present findings suggest that it can also influence the metabolomic response itself. If potentially not-exhausted participants are included without accounting for their test performance, attenuated metabolic responses may be interpreted as intrinsic biological heterogeneity rather than as a consequence of a weaker effective exercise stimulus. therefore, assessing effort attainment is relevant not only for the validity of VO_2peak measurement, but also for the interpretation of downstream biochemical responses.

Fasting status showed a more selective but still relevant association pattern. The response of eight metabolites was associated with whether participants performed the test in a fasted or non-fasted state, and most of these metabolites belonged to amino acid or amino acid-related classes. This observation is consistent with the fact that pre-exercise nutritional state influences substrate availability, circulating amino acid concentrations and the balance between nutrient mobilization and utilization during exercise. The overlap between fasting-associated and exhaustion-associated metabolites suggests that some components of the acute amino acid response are sensitive both to nutritional state and to the effective intensity of the exercise stimulus[25,46,48].

The predominance of exhaustion status and fasting status also helps explain why more general participant characteristics showed limited effects in the determinant analysis. VO_2peak, age, sex and relative fat mass are all biologically relevant variables, but in the present dataset they did not produce broad associations across the noise-adjusted metabolite panel. This does not imply that these variables are unrelated to exercise metabolites. Rather, their effects may be more metabolite-specific, smaller in magnitude, or more difficult to detect after correction for multiple testing. In addition, the cohort consisted mainly of young adults and was not balanced by sex, which may have reduced the range of variation available for detecting age- or sex-related effects[24,26,27,45,49].

The limited association pattern for VO_2peak is particularly noteworthy. Cardiorespiratory fitness is expected to influence exercise physiology, but the present findings suggest that fitness level was not the dominant determinant of the immediate noise-adjusted metabolomic response within this cohort. One possible interpretation is that the acute metabolic response to maximal CPET depends more strongly on the actual intensity attained during the test and on pre-exercise nutritional state than on fitness alone. Another possibility is that fitness-related effects may be distributed across multiple metabolites with relatively small effect sized, making them difficult to detect in univariable models after multiple-testing correction.

Another important point is that exhaustion status and fasting status should not be interpreted as completely independent biological explanations. In the present cohort, fasting was not experimentally randomized, and the number of fasted participants was smaller than the number of non-fasted participants. Therefore, fasting-related differences may partly overlap with other participant characteristics of behavioural factors. Similarly, participants classified as potentially not exhausted may differ from exhausted participants in motivation, tolerance of maximal effort, training background or other unmeasured characteristics[25,46–48].

This means that the determinant analysis identifies associations rather than causal mechanisms[16–18]. The fact that exhaustion status and fasting status were the strongest predictors of response variability indicates that these variables captured relevant structure in the data, but it does not prove that they were the only underlying causes of the observed differences. More controlled experimental designs would be needed to separate the independent effects of nutritional state and effort attainment. For example, future studies could standardize fasting status before testing and apply the same verification protocol across all participants, allowing exhaustion-related and nutrition-related effects to be modelled with less confounding.

Despite this limitation, the present analysis remains informative because it highlights variables that should not be ignored. Even when fasting status and exhaustion status are imperfect proxies, they identify important sources of variability that can affect the interpretation of acute metabolomic responses. Their inclusion therefore improves the transparency of the analysis and prevents over-attribution of response heterogeneity to stable biological differences between participants.

Overall, the determinant analysis indicates that acute metabolomic response variability is shaped by both physiological state and experimental conditions. I particular, the effective exercise stimulus and the nutritional condition before exercise appear to be more relevant than broad demographic and

anthropometric characteristics for explaining the strongest patterns in the noise-adjusted response. These results reinforce the importance of recording and modelling contextual variables when interpreting acute exercise metabolomics data.

8.5 Exploratory Responder Patterns

The baseline-envelope and responder analyses provided an additional individual-level perspective on the acute exercise response. While the signal-to-noise framework quantified the average strength of each metabolite response relative to baseline-derived methodological noise, the baseline-envelope approach examined whether post-exercise values moved outside the resting range observed for each participant. This distinction is useful because a metabolite can show a robust average response at the group level while still displaying heterogeneous response directions across individuals[16–18].

For the metabolites highlighted by the determinant analysis, post-exercise values frequently fell outside the individual baseline envelope. This finding supports the interpretation that the selected metabolites were not only significant at the group level, but also often showed changes that exceeded each participant's own repeated-baseline range. In particular, several amino acids and amino acid-related metabolites displayed strong directional patterns, indicating that the acute exercise stimulus produced responses that were detectable beyond subject-specific resting variability in a substantial proportion of participants[25,46,48].

At the same time, the screening of metabolites with SNR values greater than 1 showed that a small subset displayed bidirectional response patterns. For these metabolites, some participants showed increases whereas others showed decreases relative to their individual baseline envelope. This result highlights that the direction of the acute metabolic response was not uniform for every metabolite. Such heterogeneity may reflect differences in nutritional status, substrate availability, exercise intensity, metabolic regulation or other participant-specific characteristics that were not fully represented in the available covariates.

Among the bidirectional metabolites, arginine showed the clearest association with fasting status. Participants with increasing arginine responses were more frequently represented in the fasted condition, suggesting that pre-exercise nutritional state may influence not only the magnitude but also the direction of the arginine response. This observation is consistent with the earlier continuous determinant analysis, in which arginine was also associated with fasting status. Although this result should not be interpreted as evidence of a causal mechanism, it supports the idea that fasting-related metabolic conditions may shape individual response trajectories.

The additional responder analysis further explored selected metabolites with robust SNR values and intermediate responder proportions. This approach was designed to avoid highly unbalanced responder groups and to focus on metabolites for which both responded and non-responder patterns were sufficiently represented. The selected metabolites overlapped with the set identified in the determinant analyses, suggesting that the responder analysis was not entirely disconnected from the broader response structure observed in the dataset.

Tryptophan showed the clearest exploratory association with physiological performance variables. Participants classified as tryptophan responders showed higher peak RER and higher VT1 than non-responders. This may suggest that tryptophan response status captured part of the variability related to effective exercise intensity of aerobic performance. However, these findings should be interpreted cautiously. Responder groups were defined post hoc, multiple exploratory comparisons were considered, and the analysis was not designed as a confirmatory test of a pre-specified physiological hypothesis.

The responder analyses therefore add value mainly as a hypothesis-generating extension of the repeated-baseline framework. They show that individual deviations from baseline-derived reference ranges can be used to explore heterogeneous response patterns that may not be fully visible in conventional group-

level analyses. At the same time, the limited number of robust associations indicates that much of the observed inter-individual variability remains unexplained.

A key limitation of responder analyses is that the classification depends strongly on how the response threshold is defined. In the present thesis, responder status was based on deviations from a baseline-derived reference range and was used as an exploratory tool. However, when thresholds are defined after inspecting the data, responder groups may partly reflect sampling variability, measurement noise or regression-to-the-mean effects rather than stable biological phenotypes[16–18].

Regression to the mean is particularly relevant when individuals are classified according to deviations from a small number of baseline measurements. Participants with relatively high or low baseline values may appear to show larger changes simply because repeated measurements tend to move closer to the average. The repeated-baseline design reduces this risk compared with a single baseline but it does not eliminate it completely, especially when only two resting measurements are available.

For this reason, the responder results should be interpreted as a way to generate hypotheses about individual metabolic response patterns rather than as a definitive classification of participants. A confirmatory responder framework would require pre-specified metabolite-specific thresholds, repeated resting measurements across different days, prospective validation and ideally an independent cohort[16–18]. Such a design would make it possible to determine whether responder status is reproducible within individuals and whether it predicts meaningful physiological or training-related outcomes.

The tryptophan responder findings are a useful example. The association with peak RER and VT1 suggests that tryptophan deviations may be related to exercise intensity or aerobic performance, but the analysis was exploratory and involved post hoc group definition. Therefore, the result should be presented as a promising signal for future investigation rather than as evidence of a validated tryptophan responder phenotype.

Overall, these analyses reinforce the idea that acute exercise metabolomics should not be interpreted only in terms of average metabolite changes. Repeated-baseline information also makes it possible to examine whether individual participants respond outside their own expected resting range, whether those responses are directionally consistent and whether selected response patterns may relate to physiological or experimental characteristics. This perspective may be particularly useful in future studies with larger samples, more standardized nutritional conditions and pre-specified responder hypotheses.

8.6 Prediction of VO_2peak

The predictive modelling analysis extended the thesis from the interpretation of acute exercise responses toward the estimation of a functional physiological outcome. The main question was whether baseline capillary metabolomic features could improve the prediction of VO_2peak beyond information already available from demographic, anthropometric and activity-related variables. This question is relevant because a practical metabolomics-based model would ideally provide additional information about cardiorespiratory fitness before exercise testing is performed.

In the present cohort, participant-level clinical features provided the strongest predictive information. the clinical-only model achieved moderate out-of-fold performance, with an R^2 of 0.606 ad a mean absolute error of approximately 4.3 mL min^{-1} kg^{-1}. The leave-one-feature-out analysis indicated that body mass index, sex and weekly endurance sport activity made the largest independent contributions of this performance. This result is physiologically plausible because VO_2peak is strongly related to body size, body composition, sex-related physiological characteristics and habitual endurance training exposure[50].

By contrast, the baseline metabolomic branch showed limited predictive value. The omics-only model based on the selected 18-metabolite subset performed poorly and did not provide a useful standalone

estimate of VO_2peak. In addition, combining the metabolomic and clinical branches through an ensemble strategy did not meaningfully improve performance beyond the clinical-only model. The ensemble weights strongly favoured the clinical model, indicating that the predictive contribution od the selected metabolite subset was small.

The residual omics strategy provided a more stringent test of potential metabolomic value. In this approach, the metabolomic model was asked to predict the component of VO_2peak not already explained by clinical variables. The lower performance of the residual omics model suggests that the selected baseline metabolites did not capture a substantial independent residual signal. This result is important because it reduces the possibility that the limited contribution of the omics branch was simply masked by the stronger clinical predictors in the ensemble model.

The full-panel sensitivity analysis helped clarify the interpretation of these findings. Expanding the metabolomic candidate set from 18 to 52 metabolites improved the standalone omics-only model, indicating that the full targeted metabolite panel contained more information related to VO_2peak than the restricted subset selected from the response-determinant analysis. However, the full-panel omics model remained clearly weaker than the clinical-only model, and the combined clinical plus omics models still did not provide meaningful improvement in predictive performance. Therefore, the limited incremental contribution of baseline metabolomics cannot be attributed only to the initial restriction to 18 metabolites.

Several explanations may account for this result. First, the selected targeted metabolite panel was designed primarily to characterize exercise-related metabolic changes, not specifically to capture determinants of cardiorespiratory fitness. Baseline metabolite concentrations may therefore contain only indirect or weak information about VO_2peak. Second, the main analysis intentionally used pre-exercise baseline values. This was conceptually appropriate for a genuine pre-exercise prediction task, but it excluded dynamic information from post-exercise changes that may be more closely related to exercise capacity. Third, the available samples size was limited for a metabolomics-based prediction problem, particularly after excluding participants classified as potentially not exhausted.

The finding that clinical variables outperformed metabolomic features should not be interpreted as evidence that the capillary metabolome is unrelated to cardiorespiratory fitness. Rather, it indicates that, in this cohort and with the present modelling framework, baseline metabolomic information did not add sufficient independent predictive value beyond variables that already capture major sources of VO_2peak variability. This is consistent with the broader results of the thesis, in which metabolomics was more informative for characterizing acute response patterns and their variability than for predicting baseline fitness status.

The prediction results also highlight an important distinction between association and prediction. A metabolite may be consistently selected by a regularized model or show biological relevance in the exercise-response analyses, but this does not necessarily mean that it improves out-of-sample prediction of VO_2peak. The selection-frequency results should therefore be interpreted as exploratory model structure rather than evidence of validated metabolomic biomarkers[39–42].

The absence of meaningful incremental predictive value from baseline metabolomics is itself an informative result. In omics studies, there is often an expectation that high-dimensional molecular data should improve prediction beyond standard clinical or physiological variables. However, this is not guaranteed, especially when the clinical variables already capture major determinants of the outcome. In the present thesis, sex, body mass index and endurance sport activity provided strong information about VO_2peak, leaving limited residual variance for baseline metabolites to explain.

This finding also highlights the distinction between biological relevance and predictive utility[37,39–42]. A metabolite may be involved in the acute exercise response or selected by a regularized model without improving out-of-sample prediction. Predictive value depends not only on biological plausibility, but

also on effect size, redundancy with other predictors, measurement stability, sample size and generalizability. Therefore, the weak incremental contribution of baseline metabolomics should not be interpreted as evidence that the metabolome is unrelated to exercise physiology. Rather, it indicates that the baseline metabolite concentrations measured in this cohort did not provide sufficient independent information beyond simpler participant-level variables.

From a data-science perspective, this result strengthens the thesis because it avoids overclaiming[41,42]. The repeated nested cross-validation strategy, residual omics model and full-panel sensitivity analysis all tested whether baseline metabolites added predictive value in different ways. The convergence of these analyses toward the same conclusion makes the negative prediction result more credible. It suggests that the main value of the metabolomic data in this study lies in characterizing the acute response and its variability, rather than replacing or improving standard predictors of cardiorespiratory fitness.

The full-panel sensitivity analysis is particularly important in this interpretation. The improvement of the omics-only model from the 18-metabolite subset to the full 52-metabolite panel indicates that the metabolome did contain some information related to VO_2peak. However, the lack of improvement in the combined clinical-plus-omics models shows that this information was not sufficiently complementary to the clinical features. This distinction is important because it prevents the conclusion from being too strong. The appropriate interpretation is not that baseline metabolomics contains no fitness-related signal, but that its incremental predictive value was limited in the present cohort and modelling framework.

Overall, the predictive analysis provides a cautious but useful conclusion. Baseline capillary metabolomics showed limited standalone and incremental predictive value of VO_2peak, whereas standard participant-level variables provided the dominant predictive information. The main contribution of the metabolomic data in this thesis was therefore not the development of a high-performing VO_2peak prediction model, but the characterization of acute-exercise responses through repeated-baseline and noise-aware analysis.

8.7 Strengths and Limitations

This thesis has several strengths. First, the repeated-baseline design provided an internal estimate of short-term metabolite variability before the exercise stimulus. This allowed post-exercise changes to be interpreted relative to baseline-derived methodological noise rather than only through conventional group-level significance testing. The use of two resting baseline samples therefore strengthened the interpretation of acute metabolomic responses and represented the central methodological contribution of the work.

Second, the study combined multiple analytical perspectives. Baseline reliability, multivariate exploration, conventional single-baseline testing, signal-to-noise stratification, participant-level determinant analysis, baseline-envelope response classification and predictive modelling were integrated within a single analytical framework. This made it possible to move from broad metabolomic patterns to metabolite-specific response ranking, individual-level heterogeneity and prediction of a functional physiological outcome.

Third, the prediction workflow included several procedures intended to reduce overfitting. Feature selection, model tuning, clinical imputation and ensemble-weight optimisation were performed within the training data of the cross-validation framework. In addition, the full 52-metabolite panel was analysed as a sensitivity analysis, allowing the main findings from the selected 18-metabolite model to be interpreted in a broader context.

Several limitations should also be considered. The study included 100 healthy adults, but the effective sample size was lower for specific analyses. In particular, the prediction analysis was restricted to 77

participants after excluding those classified as potentially not exhausted. This exclusion was methodologically justified because potentially unreliable VO_2peak values could distort the prediction task, but it reduced the amount of data available for model training statistical power for detecting small associations and may have contributed to the variability observed across validation folds[41,42].

The cohort was also relatively homogeneous. Participants were mainly young healthy adults, which limited the range of age-related variation and restrict generalizability to older individuals, sedentary populations or clinical groups. In addition, the cohort was not balanced by sex. After excluding potentially not-exhausted participants, the prediction cohort included 51 males and 26 females. Sex was an important predictor of VO_2peak, but the female subgroup was too small to support reliable sex-specific predictive modelling. Therefore, the combined prediction models may partly reflect the sex distribution of the present cohort.

Another limitation concerns the experimental conditions. Participants were tested under mixed fasting and non-fasting conditions. Although fasting status was included in the determinant analysis and emerged as a relevant source of responsive variability, this heterogeneity may have increased unexplained variation in baseline metabolite concentrations and acute exercise responses. More standardized nutritional conditions would improve the ability to distinguish exercise-related effects from differences in pre-exercise metabolic state.

Another limitation concerns the use of univariable models in the participant-level determinant analysis. This approach was chosen to provide an interpretable screening of associations between participant characteristics and noise-adjusted metabolite responses. However, univariable models cannot separate the independent contribution of correlated predictors. For example, fasting status, exhaustion status, body composition and habitual exercise may partly overlap with one another or with unmeasured behavioural variables. Therefore, the determinant results should be interpreted as association patterns rather than independent causal effects[16–18].

A multivariable modelling strategy could, in principle, provide a more direct estimate of independent effects. However, the present sample size limited the complexity of models that could be fitted reliably, especially after accounting for multiple metabolites and multiple predictors. Including several correlated predictors in metabolite-wise models would have increased the risk of unstable coefficients and reduced statistical power. For this reason, the simple univariable framework was appropriate for the exploratory objective of this thesis, but future studies with larger cohorts should test whether the exhaustion- and fasting-related associations remain stable in multivariable models.

The definition of methodological noise also requires careful interpretation. The repeated baseline comparison provided an operational estimate of short-ter variability, but it could not separate analytical error, pre-analytical variation and short-term biological fluctuation. Therefore, the estimated baseline-derived methodological noise should not be interpreted as a pure technical measurement error. Rather, it represents the combined variability that must be considered when assessing whether a post-exercise response exceeds the resting reference state.

A further limitation is that two baseline measurements provide only a short-term within-visit estimate of variability. This is sufficient to improve the interpretation of immediate post-exercise responses, but it does not capture longer-term day-to-day variation[14,15]. Metabolite concentrations may differ across days because of sleep, diet, recent training, circadian rhythm or other uncontrolled factors. Therefore, the baseline-derived methodological noise estimated in this thesis should be interpreted as a within-session reference rather than as a complete estimate of individual biological variability.

This distinction affects the interpretation of individual baseline envelopes. A post-exercise value outside the interval defined by two resting samples indicates a deviation beyond the observed short-term baseline range, but it does not necessarily imply that the value would be outside a broader multi-day personal reference interval. Future studies with repeated resting samples collected across multiple days

could provide more robust individual reference ranges and allow the separation of within-session variability from day-to-day biological variation.[14,15,17]

The responder analyses also have important limitations. Responder and non-responder groups were defined using post hoc baseline-envelope criteria, and multiple exploratory comparisons were performed across metabolites and participant characteristics. Although these analyses provided useful hypothesis about individual response patterns, they were not designed as confirmatory tests and should not be interpreted as definitive evidence of metabolite-specific responder phenotypes[16,18].

Finally, the predictive analyses lacked independent external validation. Cross-validation provides an internal estimate of out-of-sample performance, but it cannot fully establish whether the models would generalize to a separate population, different laboratory conditions or another metabolomics platform. This is particularly important for the metabolomic feature-selection results, which should be viewed as exploratory model-derived patterns rather than validated biomarkers of VO_2peak[41,42].

Overall, the limitations mainly affect the generalizability and confirmatory strength of the findings rather than the internal logic of the analytical framework. The study provides a coherent and data-driven evaluation of repeated-baseline capillary metabolomics, but larger and more standardized cohorts will be needed to validate the observed response patterns and to determine whether metabolomic information can provide clinically meaningful prediction of cardiorespiratory fitness.

8.8 Future Perspectives

The findings of this thesis suggest several directions for future research. First, future studies should include larger and more heterogeneous cohorts. A larger samples size would improve the precision of metabolite-level estimates, increase the power to detect smaller participant-level associations and allow more reliable stratified analyses by sex, age group, fitness level or nutritional status. In particular, balanced recruitment of males and females would make it possible to investigate whether repeated-baseline metabolomic responses and VO_2peak prediction models differ across sex-specific physiological profiles.

Second, greater standardization of pre-exercise conditions would improve the interpretation of acute metabolomic responses. The present study showed that fasting status was associated with part of the response variability, especially among amino acids and amino-acid-related metabolites. Future protocols could standardize dietary intake, fasting duration, time of day, recent physical activity and hydration status before testing. Alternatively, these factors could be measured in greater detail and incorporated directly into multivariable models.

A further perspective concerns the repeated-baseline design itself. Two baseline samples were sufficient to estimate short-term variability and to construct a practical signal-to-noise framework. However, additional resting measurements collected on different days could provide a more complete characterization of intra-individual metabolite variability. This would allow the separation of within-visit fluctuation from day-to-day biological variation and could improve the estimation of metabolite-specific reference ranges[14,15].

Future work could also extend the present analysis beyond immediate post-exercise response. Including additional recovery timepoints would make it possible to examine not only the magnitude of the acute perturbation, but also the temporal dynamics of metabolite recovery. Some metabolites may show rapid transient changes immediately after exercise, whereas others may change later during recovery. A longitudinal post-exercise design could therefore identify response trajectories rather than only single post-exercise deviations.

From an analytical perspective, future studies could compare the current signal-to-noise approach with alternative modelling strategies. Mixed-effects models, Bayesian hierarchical models or variance-component approaches could explicitly model repeated measurements and separate between-subject

from within-subject variation. These methods may provide more flexible estimates of individual response uncertainty, particularly in larger datasets with multiple baseline and recovery measurements.

The prediction analysis could also be expanded in several ways. Larger training datasets and independent external validation cohorts would be necessary to determine whether baseline metabolomic information can provide stable incremental value for VO_2peak prediction. Future models could incorporate additional physiological data, such as resting heart rate, ventilatory thresholds, training history, nutritional variables or wearable-derived activity measures. Dynamic features derived from post-exercise changes may also improve prediction of exercise capacity, although such models would address a different question from genuine pre-exercise prediction.

Finally, the responder analyses could be developed into more formal precision-exercise frameworks. Rather than defining responder status post hoc, future studies could pre-specify metabolite-specific response thresholds based on repeated resting measurements and evaluate them prospectively. This would allow the identification of response phenotypes that may be linked to training adaptation, nutritional interventions or differences in exercise tolerance[16–18].

Another future direction concerns the translation of repeated-baseline metabolomics into practical study designs. In its simplest form, a repeated-baseline approach requires only one additional pre-exercise sample, but it can substantially improve the interpretability of post-exercise changes. This makes it attractive for studies in which the objective is not only to detect average exercise-induced perturbation, but also to identify metabolites that may be reliable at the individual level[16–18]. Future work should evaluate whether the additional sampling burden is justified by improved response classification, stronger reproducibility and better identification of meaningful individual response patterns.

This perspective may be particularly relevant for personalised exercise assessment[16–18]. If repeated-baseline metabolomics can identify robust individual response signatures, it could support the classification of metabolic response phenotypes or the monitoring of adaptation over time. However, such applications require stronger evidence than is available from the present exploratory study. They would require prospective designs, repeated testing, external validation and clinically r physiologically meaningful outcomes linked to the metabolomic response.

Overall, the present work supports the continued use of repeated-baseline metabolomics in exercise physiology. The main future opportunity is to combine carefully standardized experimental designs, repeated longitudinal measurements and robust statistical modelling in larger cohorts such approaches could improve the identification of reliable markers and clarify whether individual metabolomic profiles can contribute meaningfully to personalised exercise assessment.

9 Conclusion

This thesis investigated the capillary blood metabolome in relation to acute maximal exercise and cardiorespiratory fitness, with particular emphasis on the value of repeated baseline sampling. The central objective was to determine whether two resting baseline measurements could improve the interpretation of post-exercise metabolite changes by providing an estimate of baseline-derived methodological noise.

The results showed that the capillary metabolome was generally reproducible across repeated resting measurements. Baseline 1 and Baseline 2 showed strong overall agreement, most metabolites displayed moderate-to-high reliability, and repeated samples retained subject-specific metabolic patterns. These findings supported the use of the repeated baseline as a stable reference for evaluating acute exercise-induced changes.

Maximal CPET induced a broad and structured metabolic perturbation involving amino acids, amino acid-related metabolites, fatty acids and lactate. Conventional single-baseline analysis identified significant post-exercise changes in most retained metabolites, confirming the presence of a strong group-level exercise response. However, the repeated-baseline signal-to-noise framework showed that these changes differed substantially in robustness relative to baseline-derived methodological noise. This highlighted that statistical significance alone is not sufficient to determine whether an observed metabolite response clearly exceeds expected resting variability.

The participant-level analyses further showed that response heterogeneity was partly explained by exhaustion status and fasting status. Participants classified as potentially not exhausted generally showed attenuated amino acid-related responses, indicating that the effective intensity of the exercise stimulus influenced the magnitude of the observed metabolic perturbation. Fasting status also contributed to variability in selected metabolites responses, supporting the importance of pre-exercise nutritional conditions when interpreting acute exercise metabolomics data.

The individual baseline-envelope and exploratory responder analyses provided an additional subject-specific perspective. Several metabolites frequently showed post-exercise values outside their individual repeated-baseline range, whereas a smaller group displayed bidirectional response patterns. Arginine was the most interpretable bidirectional metabolite, with response direction associated with fasting status. Tryptophan responder status showed exploratory associations with peak RER and VT1, suggesting that selected metabolite response patterns may reflect differences in exercise intensity or aerobic performance. These findings remain hypothesis-generating and require confirmation in larger cohorts.

The prediction analysis showed that VO_2peak could be estimated with moderate accuracy from participant-level clinical features, particularly body mass index, sex and endurance sport activity. In contrast, baseline metabolomic information provided limited standalone and incremental predictive value. Neither the selected 18-metabolite subset nor the full 52-metabolite panel meaningfully improved prediction beyond the clinical-only model. This suggests that, within the present cohort and modelling framework, baseline capillary metabolomics was more informative for characterizing acute exercise-response variability than for predicting cardiorespiratory fitness.

Overall, this work supports repeated-baseline capillary metabolomics as a useful framework for acute exercise studies. The main contribution of the thesis is methodological: repeated baseline sampling makes it possible to evaluate post-exercise metabolite changes relative to short-term resting variability and therefore to distinguish broad group-level perturbations form responses that more clearly exceed baseline-derived methodological noise. Future studies using larger, more balanced and more standardized cohorts will be necessary to validate the observed response patterns, refine individual reference ranges and determine whether metabolomic information can provide clinically meaningful support for personalised exercise assessment

10 References

1. Walzik, D. *et al.* Systemic metabolite kinetics mirror skeletal muscle energy metabolism during acute aerobic exercise. *Am. J. Physiol.-Cell Physiol.* **330**, C111–C118 (2026).

2. Morville, T., Sahl, R. E., Moritz, T., Helge, J. W. & Clemmensen, C. Plasma metabolome profiling of resistance exercise and endurance exercise in humans. *Cell Rep.* **33**, (2020).

3. Contrepois, K. *et al.* Molecular choreography of acute exercise. *Cell* **181**, 1112–1130 (2020).

4. Lewis, G. D. *et al.* Metabolic Signatures of Exercise in Human Plasma. *Sci. Transl. Med.* **2**, (2010).

5. Nayor, M. *et al.* Metabolic Architecture of Acute Exercise Response in Middle-Aged Adults in the Community. *Circulation* **142**, 1905–1924 (2020).

6. Schranner, D., Kastenmüller, G., Schönfelder, M., Römisch-Margl, W. & Wackerhage, H. Metabolite Concentration Changes in Humans After a Bout of Exercise: a Systematic Review of Exercise Metabolomics Studies. *Sports Med. - Open* **6**, 11 (2020).

7. Khoramipour, K. *et al.* Metabolomics in Exercise and Sports: A Systematic Review. *Sports Med.* **52**, 547–583 (2022).

8. Roberts, L. D., Souza, A. L., Gerszten, R. E. & Clish, C. B. Targeted Metabolomics. *Curr. Protoc. Mol. Biol.* **CHAPTER**, Unit30.2 (2012).

9. Saito, K. *et al.* Overall Similarities and a Possible Factor Affecting Plasma Metabolome Profiles Between Venous and Capillary Blood Samples From 20 Healthy Human Males. *J. Pharm. Sci.* **108**, 3737–3744 (2019).

10. Reichel, T. *et al.* Acute response of biomarkers in plasma from capillary blood after a strenuous endurance exercise bout. *Eur. J. Appl. Physiol.* **123**, 179–189 (2023).

11. Catala, A., Culp-Hill, R., Nemkov, T. & D'Alessandro, A. Quantitative metabolomics comparison of traditional blood draws and TAP capillary blood collection. *Metabolomics* **14**, 100 (2018).

12. Hameed, A. *et al.* Comparison of the capillary and venous blood plasma lipidomes: validation of self-collected blood for plasma lipidomics. *J. Lipid Res.* **66**, (2025).

13. Nishiumi. Comparison of venous and fingertip plasma using non-targeted proteomics and metabolomics. *Talanta* **192**, 182–188 (2019).

14. Siopi, A. & Mougios, V. Metabolomics in Human Acute-Exercise Trials: Study Design and Preparation. in *Metabolic Profiling* (eds. Theodoridis, G. A., Gika, H. G. & Wilson, I. D.) vol. 1738 279–287 (Springer New York, New York, NY, 2018).

15. Kühn, T. *et al.* Mid- and long-term correlations of plasma metabolite concentrations measured by a targeted metabolomics approach. *Metabolomics* **12**, 184 (2016).

16. Ross, R. *et al.* Precision exercise medicine: understanding exercise response variability. https://doi.org/10.1136/bjsports-2018-100328 (2019) doi:10.1136/bjsports-2018-100328.

17. Chrzanowski-Smith, O. J., Piatrikova, E., Betts, J. A., Williams, S. & Gonzalez, J. T. Variability in exercise physiology: Can capturing *intra* -individual variation help better understand true *inter* -individual responses? *Eur. J. Sport Sci.* **20**, 452–460 (2020).

18. Hecksteden, A. *et al.* Individual response to exercise training - a statistical perspective. *J. Appl. Physiol.* **118**, 1450–1459 (2015).

19. Midgley, A. W., McNaughton, L. R., Polman, R. & Marchant, D. Criteria for Determination of Maximal Oxygen Uptake: A Brief Critique and Recommendations for Future Research. *Sports Med.* **37**, 1019–1028 (2007).

20. ATS/ACCP Statement on Cardiopulmonary Exercise Testing. *Am. J. Respir. Crit. Care Med.* **167**, 211–277 (2003).

21. Costa, V. A. B. *et al.* Is a verification phase useful for confirming maximal oxygen uptake in apparently healthy adults? A systematic review and meta-analysis. *PLOS ONE* **16**, e0247057 (2021).

22. Brooks, G. A. The lactate shuttle during exercise and recovery: *Med. Sci. Sports Exerc.* **18**, 360–368 (1986).

23. Rennie, M. J. & Tipton, K. D. Protein and amino acid metabolism during and after exercise and the effects of nutrition. *Annu. Rev. Nutr.* **20**, (2000).

24. Romijn, J. A. *et al.* Regulation of endogenous fat and carbohydrate metabolism in relation to exercise intensity and duration. *Am. J. Physiol.-Endocrinol. Metab.* **265**, E380–E391 (1993).

25. Zouhal, H. *et al.* Exercise Training and Fasting: Current Insights. *Open Access J. Sports Med.* **11**, 1 (2020).

26. Ralston, J. C. *et al.* Ethnic- and sex-specific associations between plasma fatty acids and markers of insulin resistance in healthy young adults. *Nutr. Metab.* **10**, 42 (2013).

27. Arner, P. & Rydén, M. Fatty Acids, Obesity and Insulin Resistance. *Obes. Facts* **8**, 147 (2015).

28. Zimmer, P. *et al.* High Frequency Saliva-Based Targeted Multi-Omics as Screening Tool in Professional Soccer – The Practice Perspective. *Ger. J. Sports Med.* **77**, 7–14 (2026).

29. Weber, D., Ferrario, P. G. & Bub, A. Exercise intensity determines circulating levels of Lac-Phe and other exerkines: a randomized crossover trial. *Metabolomics* **21**, 63 (2025).

30. Kelly, R. S., Kelly, M. P. & Kelly, P. Metabolomics, physical activity, exercise and health: A review of the current evidence. *Biochim. Biophys. Acta BBA-Mol. Basis Dis.* **1866**, 165936 (2020).

31. Kamlage, B. *et al.* Quality Markers Addressing Preanalytical Variations of Blood and Plasma Processing Identified by Broad and Targeted Metabolite Profiling. *Clin. Chem.* **60**, 399–412 (2014).

32. Bernini, P. *et al.* Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *J. Biomol. NMR* **49**, 231–243 (2011).

33. Stekhoven, D. J. & Bühlmann, P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118 (2012).

34. Assfalg, M. *et al.* Evidence of different metabolic phenotypes in humans. *Proc. Natl. Acad. Sci.* **105**, 1420–1424 (2008).

35. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, (2008).

36. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300 (1995).

37. Shmueli, G. To Explain or to Predict? *Stat. Sci.* **25**, (2010).

38. Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLOS ONE* **14**, e0224365 (2019).

39. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **58**, 267–288 (1996).

40. Zou, H. & Hastie, T. Regularization and Variable Selection Via the Elastic Net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).

41. Varma, S. & Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* **7**, 91 (2006).

42. Meinshausen, N. & Bühlmann, P. Stability selection. Journal of the Royal Statistical Society: Series B 72, 417–473 (2010).

43. Wenzel, C. *et al.* Targeted saliva multi-omics is a reliable, non-invasive method to capture physiological stress and recovery. Preprint at https://doi.org/10.64898/2026.01.28.26345066 (2026).

44. Kurgan, N. *et al.* Body Fluid Proteomic Landscape of Acute Exercise. 2025.05.28.656705 Preprint at https://doi.org/10.1101/2025.05.28.656705 (2025).

45. Nelson, A. B. *et al.* Acute aerobic exercise reveals that FAHFAs distinguish the metabolomes of overweight and normal-weight runners. *JCI Insight* **7**, (2022).

46. Pozefsky, T. *et al.* Effects of brief starvation on muscle amino acid metabolism in nonobese man. *J. Clin. Invest.* **57**, (1976).

47. Vendelbo, M. H. *et al.* Fasting Increases Human Skeletal Muscle Net Phenylalanine Release and This Is Associated with Decreased mTOR Signaling. *PLOS ONE* **9**, e102031 (2014).

48. Edinburgh, R. M., Koumanov, F. & Gonzalez, J. T. Impact of pre-exercise feeding status on metabolic adaptations to endurance-type exercise training. *J. Physiol.* **600**, 1327 (2021).

49. Moosavi, D., Vuckovic, I., Kunz, H. E. & Lanza, I. R. Metabolomic response to acute resistance exercise in healthy older adults by 1H-NMR. *PLOS ONE* **19**, e0301037 (2024).

50. Bassett, D. R. Limiting factors for maximum oxygen uptake and determinants of endurance performance: *Med. Sci. Sports Exerc.* **32**, 70 (2000).