

Aus der Fachrichtung Medizinische Biometrie, Epidemiologie und Medizinische Informatik

Direktor: Univ.-Prof. Dr. Stefan Wagenpfeil

Klinische Medizin

der Medizinischen Fakultät

der Universität des Saarlandes, Homburg/Saar

## **Quantile reference values for peak oxygen uptake:**

Cross-sectional study of 9,354 adult participants of cardiopulmonary exercise tests using cycle ergometry in three German cities.

Dissertation zur Erlangung des Grades eines Doktors der Medizin

der Medizinischen Fakultät

der Universität des Saarlandes

2018

Vorgelegt von: Daniel Rapp

Geboren am: 8 April 1987

Geboren in: Münsingen

**Dekan**

**Univ. - Prof. Dr. med. Michael D. Menger**

**1. Berichterstatter**

**Univ. - Prof. Dr. Stefan Wagenpfeil**

Institut für Medizinische Biometrie, Epidemiologie und  
Medizinische Informatik (IMBEI), Universität des Saar-  
landes, Homburg (Saar)

**2. Berichterstatter**

**apl. Prof. Dr. med. Jürgen Scharhag**

Institut für Sport- und Präventivmedizin, Universität des  
Saarlandes, Saarbrücken

Zentrum für Prävention und Sportmedizin, Klinikum rechts  
der Isar, Technische Universität München, München

# Contents

<b>Author's publications related to this dissertation</b>	<b>iii</b>
<b>Table of abbreviations</b>	<b>iv</b>
<b>Nomenclature</b>	<b>v</b>
<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>vii</b>
<b>Abstract</b>	<b>viii</b>
<b>Zusammenfassung</b>	<b>ix</b>
<b>1 Rationale and objectives</b>	<b>1</b>
<b>2 Background</b>	<b>4</b>
2.1 Cardiorespiratory fitness as a risk factor . . . . .	4
2.2 Exercise tests in preventive medicine . . . . .	6
2.3 Measurement of peak oxygen uptake . . . . .	7
2.4 Reference values for peak oxygen uptake . . . . .	11
<b>3 Material &amp; Methods</b>	<b>15</b>
3.1 Study design and participants . . . . .	15
3.2 Measurement of sample characteristics . . . . .	17
3.3 Data management . . . . .	18
3.4 Representativity . . . . .	21
3.5 Sample description . . . . .	22
3.6 Statistical methods . . . . .	24
3.7 Reporting . . . . .	29
<b>4 Results</b>	<b>30</b>
4.1 Descriptive statistics . . . . .	30
4.2 Representativity . . . . .	37
4.3 Nomograms and regression model validation . . . . .	39
4.4 Exercise test modalities and random sample . . . . .	50
4.5 Web application . . . . .	52
4.6 Multivariable analyses . . . . .	54

<b>5</b>	<b>Discussion</b>	<b>63</b>
5.1	Reference values . . . . .	64
5.2	Study population and representativity . . . . .	66
5.3	Exercise test modalities . . . . .	70
5.4	Statistical methods . . . . .	73
5.5	Multivariable analyses . . . . .	76
5.6	Summary of strengths and limitations . . . . .	78
<b>6</b>	<b>Conclusions and clinical implications</b>	<b>79</b>
<b>7</b>	<b>Statistical code</b>	<b>81</b>
7.1	Web application . . . . .	81
7.1.1	ui.R . . . . .	81
7.1.2	server.R . . . . .	83
7.2	Calculations for web application . . . . .	87
7.3	Statistical code . . . . .	91
7.3.1	Specifications . . . . .	91
7.3.2	Data . . . . .	91
7.3.3	Random sample . . . . .	94
7.3.4	Descriptive statistics . . . . .	96
7.3.5	Representativity . . . . .	100
7.3.6	Reference values . . . . .	102
7.3.7	Validation . . . . .	108
7.3.8	Multiple quantile regression . . . . .	114
7.3.9	Comparison of reference values . . . . .	117
<b>8</b>	<b>Bibliography</b>	<b>120</b>
<b>A</b>	<b>Appendix</b>	<b>129</b>
A.1	Random sample . . . . .	129
A.2	Selective dropout analysis . . . . .	130
A.3	Regression modelling . . . . .	136
A.4	Informed consent . . . . .	139
<b>B</b>	<b>Danksagung</b>	<b>140</b>

## Author's publications related to this dissertation

### Medical journals

Rapp D, Scharhag J, Wagenpfeil S, Scholl J (2018) Reference values for peak oxygen uptake: cross-sectional analysis of cycle ergometry-based cardiopulmonary exercise tests of 10 090 adult German volunteers from the Prevention First Registry. *BMJ Open* Mar 2018, 8 e018697. doi:10.1136/bmjopen-2017-018697

Rapp D, Scharhag J, Wagenpfeil S, Scholl J (2018) Reference values for peak oxygen uptake (secondary publication). *Dtsch Z Sportmed.* 2018; 69: 199-205. doi:10.5960/dzsm.2018.329

### Scientific conference presentation

Rapp D, Scholl J, Kurz P, Scharhag J, Wagenpfeil S (2016) Sex- and age-specific nomograms for maximal oxygen uptake: Quantile regression models in a sample of 10,189 individuals. Presentation at the Joint Workshop on High-dimensional and Functional Data Analysis of the International Biometric Society (IBS), 2nd of September, 2016, Ulm

### Poster presentation

Rapp D, Scholl J, Kurz P, Scharhag J, Wagenpfeil S (2016) Adjusted percentile reference values for peak oxygen uptake: Regression analysis in a sample of 10,189 individuals from primary preventive examinations. Presentation at HEC 2016 (Health - Exploring Complexity), 31st of August, 2016, Munich

### Websites

Interactive web application:

[https://vo2peak.shinyapps.io/vo2peak\\_calculator/](https://vo2peak.shinyapps.io/vo2peak_calculator/)  
[www.uks.eu/vo2peak](http://www.uks.eu/vo2peak)

Web appendix:

<https://github.com/rappdaniel/vo2peak>

## Table of abbreviations

ACCP	American College of Chest Physicians
ACSM	American College of Sports Medicine
AIC	Akaike Information Criterion
ATS	American Thoracic Society
APMHR	Age-predicted maximal heart rate
BMI	Body mass index
CI	Confidence interval
CPET	Cardiopulmonary exercise test
CRF	Cardiorespiratory fitness
DEGS	Studie zur Gesundheit Erwachsener in Deutschland
DGSP	Deutsche Gesellschaft für Sportmedizin und Prävention
fig.	figure
FRIL	Fine-grained record linkage software
HDL	High density lipoprotein
HR <sub>max</sub>	Maximal heart rate
HR <sub>rest</sub>	Heart rate at rest
HRR	Heart rate reserve
JBS3	Joint British Society 3
LDL	Low density lipoprotein
ME	Margin of error
MET	Metabolic equivalent
p.	Page
PROCAM	Prospective cardiovascular Münster study
RECORD	Reporting of studies conducted using observational routinely-collected data
RER	Respiratory exchange ratio
RKI	Robert Koch-Institut
RR	Risk ratio
SCORE	Systematic coronary risk evaluation
STROBE	Strengthening the reporting of observational studies in epidemiology
$\dot{V}CO_2$	Volume of carbon dioxide elimination per time unit
$\dot{V}O_2$	Volume of oxygen uptake per time unit
$\dot{V}O_{2peak}$	Peak oxygen uptake
$\dot{V}O_{2R}$	Oxygen uptake reserve
$\dot{V}O_{2rest}$	Oxygen uptake at rest
WHO	World Health Organization

Units of measurement are abbreviated according to SI-units.

## Nomenclature

$\alpha$	Type I error probability
$\hat{\beta}$	Estimated regression coefficient
$e$	Error term
$i$	$i$ -th vector element: $x_i$
$r$	Spearman's correlation coefficient
$n$	Sample size
$\hat{p}$	Expected sample proportion
$\tau$	Quantile of the quantile regression
$z$	Quantiles of the normal distribution

## List of Figures

3.1	Data sources and information. . . . .	21
3.2	Flow chart showing numbers of cases included in reference values. . . . .	24
4.1	Cumulative number of participants by date of examination. . . . .	31
4.2	Box plots of absolute and relative $\dot{V}O_{2\text{peak}}$ by age group. . . . .	35
4.3	Nomogram of absolute $\dot{V}O_{2\text{peak}}$ (Rüdesheim/Frankfurt). . . . .	41
4.4	Nomogram of relative $\dot{V}O_{2\text{peak}}$ (Rüdesheim/Frankfurt). . . . .	42
4.5	Coefficient plots of absolute $\dot{V}O_{2\text{peak}}$ . . . . .	45
4.6	Coefficient plots of relative $\dot{V}O_{2\text{peak}}$ . . . . .	45
4.7	Apparent validation box plots. . . . .	46
4.8	Apparent validation calibration plots. . . . .	47
4.9	External validation box plots. . . . .	48
4.10	External validation calibration plots. . . . .	49
4.11	Median regressions for inclusion versus exclusion of participants with no maximal effort. . . . .	51
4.12	Screenshot of web application. . . . .	53
4.13	Coefficient plots of multivariable quantile regression models for absolute $\dot{V}O_{2\text{peak}}$ in men. . . . .	55
4.14	Coefficient plots of multivariable quantile regression models for absolute $\dot{V}O_{2\text{peak}}$ in women. . . . .	56
4.15	Coefficient plots of multivariable quantile regression models for relative $\dot{V}O_{2\text{peak}}$ in men. . . . .	57
4.16	Coefficient plots of multivariable quantile regression models for relative $\dot{V}O_{2\text{peak}}$ in women. . . . .	58
5.1	Comparison of reference values for $\dot{V}O_{2\text{peak}}$ . . . . .	66
A.1	Histograms of continuous variables for men. . . . .	131
A.2	Histograms of continuous variables for women. . . . .	132
A.3	Correlation matrix of continuous variables for men. . . . .	133
A.4	Correlation matrix of continuous variables for women. . . . .	134
A.5	Apparent validation box plots: Comparing regression models. . . . .	136
A.6	Apparent validation calibration plots: Comparing regression models. . . . .	137
A.7	Informed consent . . . . .	139



## List of Tables

2.1	End criteria for maximal effort in exercise tests. . . . .	10
3.1	Inclusion and exclusion criteria of the present study. . . . .	16
3.2	Specifications of record linkage. . . . .	20
3.3	Eligible variables of the present study for comparison with DEGS1. . . . .	22
4.1	Descriptive statistics by sex and study center. . . . .	32
4.2	Quantiles of relative $\dot{V}O_{2\text{peak}}$ by sex and age group. . . . .	36
4.3	Quantiles of absolute $\dot{V}O_{2\text{peak}}$ by sex and age group. . . . .	36
4.4	Comparison of study population to results of DEGS1. . . . .	38
4.5	Regression coefficients and 95% confidence intervals of quantile regressions plotted in nomograms (absolute $\dot{V}O_{2\text{peak}}$ ). . . . .	43
4.6	Regression coefficients and 95% confidence intervals of quantile regressions plotted in nomograms (relative $\dot{V}O_{2\text{peak}}$ ). . . . .	44
4.7	Regression coefficients and R squared for linear regression in calibration data for apparent validation. . . . .	50
4.8	Regression coefficients and R squared for linear regression in calibration data for external validation. . . . .	50
4.9	Median regression using exertion and age as interaction terms. . . . .	52
4.10	Regression coefficients of multivariable quantile regressions for relative $\dot{V}O_{2\text{peak}}$ in men. . . . .	59
4.11	Regression coefficients of multivariable quantile regressions for relative $\dot{V}O_{2\text{peak}}$ in women. . . . .	60
4.12	Regression coefficients of multivariable quantile regressions for absolute $\dot{V}O_{2\text{peak}}$ in men. . . . .	61
4.13	Regression coefficients of multivariable quantile regressions for absolute $\dot{V}O_{2\text{peak}}$ in women. . . . .	62
5.1	Quality assessment of the measurement of $\dot{V}O_{2\text{peak}}$ . . . . .	73
A.1	Bivariate descriptive table of the full dataset and random sample. . . . .	129
A.2	Bivariate descriptive table of the final dataset and excluded cases. . . . .	130
A.3	Skewness and quantile-based skewness of all eligible quantitative variables	135
A.4	Regression coefficients and R squared for linear regression in calibration data for apparent validation. Comparison of three regression models. . . . .	138
A.5	Akaike Information Criterion AIC for different regression models. . . . .	138

# Abstract

## Background

Low cardiorespiratory fitness is a crucial risk factor for premature death and a plethora of health threats. It is determined by measuring the maximal volitional oxygen consumption ( $\dot{V}O_{2\text{peak}}$ ) in incremental cardiopulmonary exercise tests. The interpretation of an individual's  $\dot{V}O_{2\text{peak}}$  is only meaningful if sex-specific and age-specific reference values are considered. The primary goal of this study was to create reference values for  $\dot{V}O_{2\text{peak}}$  based on cardiopulmonary exercise tests using cycle ergometry.

## Methods

The data were acquired in the course of primary preventive health screenings. Overall, 9,354 German white-collar workers (6,063 men, 3,291 women) aged 25 to 69 years who performed cycle ergometry-based incremental exercise tests were included in the analysis. Three study centres recorded the data in a central database (Prevention First Registry) for an inquiry period between 2001 and 2015. Quantile regressions were used to create nomograms and an interactive web application was developed ([www.uks.eu/vo2peak](http://www.uks.eu/vo2peak)). Apparent and external validations of the regression fits were performed. The generalisability of this sample was assessed by comparing five characteristics to a study which was representative of the German population. Exercise test modalities were not recorded in the full dataset but were acquired retrospectively for a random sample with an a priori calculated sample size of 252 participants.

## Results

An estimated proportion of 97% of the recorded exercise tests was continued until exertion. The reference values showed a particularly high validity for the age groups from 30 to 64 years. 3/5 characteristics in men and 4/5 characteristics in women of this sample were significantly different from the German population, indicating a selection of healthy participants.

## Conclusions

The reference values presented by this study are based on one of the most extensive databases in this field. They can be used for participants of cycle ergometry-based exercise tests aged 25 to 69 years who are part of a population that is comparable to this study.

# Zusammenfassung

## Hintergrund

Eine niedrige kardiorespiratorische Fitness ist ein entscheidender modifizierbarer Risikofaktor für vorzeitiges Versterben sowie eine Vielzahl von weiteren Gesundheitsgefahren. Sie wird bestimmt, indem die maximal willkürliche Sauerstoffaufnahme ( $\dot{V}O_{2\text{peak}}$ ) bei einer spiroergometrischen Untersuchung gemessen wird. Die individuelle  $\dot{V}O_{2\text{peak}}$  ist jedoch nur aussagekräftig, wenn sie mit geschlechtsspezifischen und altersspezifischen Referenzwerten verglichen wird. Das primäre Ziel dieser Arbeit war die Erstellung von Referenzwerten für die  $\dot{V}O_{2\text{peak}}$ , welche im Rahmen von Spiroergometrien mit Fahrradergometern erhoben wurden.

## Methoden

Die Datenerhebung erfolgte im Rahmen von präventivmedizinischen Untersuchungen bei denen eine Spiroergometrie mit Fahrradergometern durchgeführt wurde. Insgesamt wurden 9,354 Teilnehmer (6,063 Männer, 3,291 Frauen) im Alter von 25 bis 69 Jahren in die Analyse eingeschlossen. Die Teilnehmer waren vor allem Büroangestellte mit einem bewegungsarmen Arbeitsumfeld. Untersuchungen aus drei deutschen Zentren wurden in einer zentralen Datenbank (Prevention First Registry) gespeichert. Quantilregressionen wurden berechnet um Nomogramme zu erstellen und eine Webapplikation wurde entwickelt ([www.uks.eu/vo2peak](http://www.uks.eu/vo2peak)). Apparente und externe Validierungen der Quantilregressionen wurden durchgeführt. Fünf Merkmale dieser Studienpopulation wurden mit einer für Deutschland repräsentativen Studie verglichen um die Repräsentativität der Ergebnisse zu untersuchen. Die Testmodalitäten der Spiroergometrien wurden nicht in der zentralen Datenbank gespeichert und wurden deshalb für eine Zufallsstichprobe mit einer a priori berechneten Fallzahl von 252 Teilnehmern erhoben.

## Ergebnisse

Bei einem geschätzten Anteil von 97% der Teilnehmer wurden Ausbelastungskriterien erfüllt. Die Referenzwerte zeigten eine besonders hohe Validität für die Altersgruppen von 30 bis 64 Jahre. 3/5 Merkmale bei Männern und 4/5 Merkmale bei Frauen dieser Studienpopulation zeigten signifikante Unterschiede zur deutschen Bevölkerung. Dies deutet auf eine Selektion von gesunden Teilnehmern hin.

## Schlussfolgerung

Die Referenzwerte dieser Studie basieren auf einer der umfangreichsten bisher publizierten Datengrundlagen in diesem Feld. Sie können für Teilnehmer von fahradergometrischen Spiroergometrien im Alter von 25 bis 69 Jahren verwendet werden, die einer Bevölkerungsgruppe angehören, welche vergleichbar mit dieser Studienpopulation ist.

# 1 Rationale and objectives

Cardiorespiratory fitness (CRF) is the ability of a person to perform physical activity for a prolonged period of time [53, p. 72]. During physical activity, the cardiorespiratory system reacts with an increased heart rate and breathing rate, which is necessary to supply the additional amount of oxygen. This is needed to sustain increased energetic requirements through the oxidation of metabolic substrates.

A comprehensive body of evidence shows that low endurance capacity is an important risk factor for a plethora of health threats such as cardiovascular diseases and premature death [35, 61, 41]. This means that the risk of dying and the risk of cardiovascular diseases is higher in persons with low endurance capacity. Besides cardiovascular diseases and all-cause mortality, CRF is also associated with diabetes mellitus [5, 77], some types of cancer [60] and - to a lower extent - with psychiatric diseases like depression or dementia [21].

CRF is a strong predictor for health threats in later life and its predictive power is comparable to well-accepted risk factors such as tobacco smoking or arterial hypertension. Kim et al. (2007) [33] stated that, “Exercise capacity is known to be one of the most important predictors of death for men and women alike.” However, despite the high predictive power, cardiovascular risk calculators like PROCAM [4]\*, European SCORE [12]†, Framingham [13] or JBS3 [28]‡ do not consider CRF for risk estimation. The American Heart Association (AHA) emphasised in 2016 that this is a major drawback and that it is crucial to consider CRF for cardiovascular risk predictions [58].

The assessment of CRF in primary preventive health examinations is crucial and the value of CRF in preventive medicine goes beyond the mere prediction of future health threats. As physical exercise increases CRF [39], it is assumed to be a modifiable risk factor. Participants of preventive health examinations are hence able to impact their future health by performing physical activity regularly. The AHA emphasised that CRF is an essential modifiable factor that should be addressed to reach the AHA’s 2020 goals of improving the cardiovascular health of US-Americans [40].

The most common approach for determining CRF is by performing incremental cardiopulmonary exercise testing (CPET) and measuring the participant’s oxygen uptake. During CPET, the participant performs a physical activity of increasing intensity which leads to increased requirements for oxygen. The participant tries to maintain the increased physical activity until the maximal volitional intensity is achieved. At this point of the exercise test, the oxygen uptake of the participant is assessed. If the intensity performed in the exercise test is close to the true exercise capacity of the participant, the determined

---

\*PROCAM = Prospective Cardiovascular Münster study

†SCORE = Systematic Coronary Risk Evaluation

‡JBS3 = Joint British Society 3

oxygen uptake is then stated to be the maximal oxygen uptake. The maximal oxygen uptake is widely considered to be the gross criterion of CRF [44, 59].

Gas exchange measurement during CPET usually determines the oxygen uptake of a participant as the difference between the volume of oxygen inhaled and exhaled per time unit. The result of this calculation is the volumetric flow rate ( $\dot{V}O$ ) of oxygen which is incorporated by the participant\*. According to the American Thoracic Society’s statement on cardiopulmonary exercise testing (2003), there are two notations of the highest oxygen uptake measured using CPET [1].  $\dot{V}O_{2\max}$  is considered to express the oxygen uptake in the case that the participant maintained the incremental exercise until the maximal volitional intensity. Maximal effort can be assumed based on end criteria measured during the exercise test. On the other hand,  $\dot{V}O_{2\text{peak}}$  denotes the oxygen uptake if it is not clear whether the participant performed the exercise test until maximal exertion. However, there is no strict distinction between the usage of  $\dot{V}O_{2\max}$  and  $\dot{V}O_{2\text{peak}}$ , and both terms are often used interchangeably. Therefore, the present study uses  $\dot{V}O_{2\text{peak}}$  to denote the highest oxygen uptake during CPET irrespective of whether end criteria were considered or not.

As mentioned above, a high CRF seems to be essential for improving cardiovascular health. However, the definition of what exactly can be defined as a high CRF is not apparent. CRF is known to be strongly dependent on sex and age [23, 50]. Other factors such as physical activity or tobacco smoking are also known to have an impact, which is why there is a wide inter-individual variability of CRF [50]. Considering the substantial role that sex plays as well as the decline of CRF associated with increasing age, it is critical to interpret the individual CRF of a CPET participant in the light of sex- and age-specific reference values from a comparable population. According to Kim et al. (2007), “defining normative values for EC [exercise capacity] is of utmost importance in accurate risk prediction after stress testing” [33]. Based on this, the primary and secondary goals of the present analysis were:

### Primary goals

1. **Primary goal 1:** To calculate percentile reference values for peak oxygen uptake.
2. **Primary goal 2:** To visualise reference values as nomograms and as an interactive web application.

---

\* $\dot{V}O$  = volume of oxygen per time unit =  $\frac{\delta V}{\delta t}$ , measured in litres of oxygen per minute

## Secondary goals

1. **Secondary goal 1:** To perform external validation of the reference values. The population used for external validation should be different from the population in which the reference values are calculated.
2. **Secondary goal 2:** To compare this study sample with the German population in order to detect selection bias.
3. **Secondary goal 3:** To perform an exploratory multivariable analysis aiming to find predictors for peak oxygen uptake.

In the following, the role of CRF in preventive medicine, the measurement of CRF using CPET and present reference values for peak oxygen uptake are outlined in chapter 2. The data source, test modalities and statistical methods of the present study are described in chapter 3. The results including reference values, visualisations as nomograms and an explorative multivariable analysis are displayed in chapter 4. Finally, the results are discussed in chapter 5 and conclusions are drawn in chapter 6.

## 2 Background

This chapter reviews the current literature on the association between cardiorespiratory fitness and health-related outcomes (section 2.1) followed by the value and the general framework of incremental exercise testing in primary preventive health examinations (section 2.2). Section 2.3 describes the methodological principles of the measurement process using incremental exercise tests. In section 2.4, existing reference values for peak oxygen uptake and quality standards of the measurement are reviewed.

### 2.1 Cardiorespiratory fitness as a risk factor

Low CRF is a risk factor for premature all-cause mortality and cardiovascular disease. A comprehensive body of scientific evidence supports an inverse dose-response relationship, meaning that the higher a person's CRF, the lower the person's risk of cardiovascular disease [53, 33]. Besides, other health problems such as neoplasia, metabolic or neurodegenerative diseases have been shown to be inversely associated with CRF. The evidence can be derived from meta-analyses as well as comprehensive prospective studies. Several studies have reported adjusted effects supporting the assumption that low CRF is a risk factor which is independent of widely accepted risk factors. The effects of CRF on cardiovascular disease and mortality have been recognized as decidedly strong. CRF is therefore regarded as one of the most important independent predictors of cardiovascular disease and all-cause mortality [33].

Kodama et al. (2009) performed a systematic review and meta-analysis of observational cohort studies that addressed the question of CRF being a predictor for cardiovascular disease [35]. Overall, 33 studies were eligible for the analysis including 102,980 subjects for analysis of all-cause mortality and 84,323 subjects for analysis of cardiovascular disease. Mean age at baseline of the included studies was 37 to 58 years. The meta-analysis revealed that an increase of maximal cardiorespiratory fitness decreased the risk for cardiovascular disease (risk ratio [RR] 0.85, 95% confidence interval [CI] 0.82 to 0.88 for an increase of maximal aerobic capacity by one MET\*) and the risk of all-cause mortality (RR 0.87, 95% CI 0.84 to 0.90 for an increase of maximal aerobic capacity by one MET). Another meta-analysis was performed by Löllgen et al. (2009), who analysed the association between physical activity and all-cause mortality [41]. In total, 38 prospective cohort studies including information on the intensity of physical activity were included in the analysis. Highly active subjects had a lower risk to die compared to subjects with

---

\*CRF can be described by metabolic equivalents (METs). 1 MET is assumed to be the oxygen consumption under resting conditions which is assumed to be 3.5 mL O<sub>2</sub>/min/kg. < 3 METs indicate low-intensity physical activity and ≥ 6 METs indicate vigorous-intensity physical activity.

low activity levels (men: RR 0.78, 95% CI 0.72 to 0.84, women: RR 0.69, 95% CI 0.53 to 0.90).

Shah et al. (2016) conducted a comprehensive cohort study and observed that the CRF of younger adults (18 to 30 years) was also a predictor of health outcomes in later life [61, 11]. They observed 4,872 participants of exercise tests for a median follow-up period of 27 years. One additional minute of exercise test duration at baseline reduced the hazard of death by 15% (hazard ratio 0.85, 95% CI 0.80 to 0.91). The authors assumed that CRF was an independent risk factor and was not mediated by other well-known risk factors such as coronary artery calcification. This assumption was supported by their adjusted analysis which considered confounding and effect modification by major risk factors. Furthermore, the authors could not find an association between CRF and coronary artery calcification at each follow-up measurement and hence suggested that coronary artery calcification did not mediate the protective effects of high CRF.

Aside from all-cause mortality and cardiorespiratory events, a low CRF was a risk factor for other health-related outcomes such as metabolic diseases [77], neoplasia [60], neurodegenerative diseases [21] or, with smaller effect size, affective disorders [51].

Zaccardi et al. (2015) described the association between CRF and type 2 diabetes mellitus [77]. The authors conducted a meta-analysis of prospective studies as well as their own prospective cohort study with a follow-up period of 23 years. 92,992 subjects were analysed in the meta-analysis including 8,564 cases with type 2 diabetes mellitus. Higher CRF was associated with a lower risk for type 2 diabetes mellitus (RR 0.95, 95% CI 0.93 to 0.98 per increased MET) in the meta-analysis.

Schmid & Leitzmann (2015) performed a meta-analysis of the association between CRF and total cancer mortality [60]. They considered six prospective studies, which included 71,654 subjects with a median follow-up period of 16 years, resulting in 2,002 total cancer mortality cases. CRF was measured using maximal or submaximal exercise tests. All included studies presented risk estimates that were adjusted for age and smoking status. The meta-analysis showed a strong, inverse association when the groups with highest and lowest CRF were compared (RR 0.55 95% CI 0.47 to 0.65).

The impact of physical activity on neurodegenerative diseases was studied by Hamer & Chida (2009) [21]. In their meta-analysis, the authors extracted data from 16 prospective cohort studies, including 163,797 participants with 2,731 dementia or Alzheimer's disease patients and 488 patients with Parkinson's disease. CRF was not measured directly in the included studies, but self-reported physical activity was used as the predictor. When subjects with the highest and the lowest physical activity were compared, Alzheimer's disease was found less often in the group with higher physical activity (RR 0.55, 95% CI 0.36 to 0.84). Other associations showed lower effects measures or non-significant results. Another association was analysed by Papasavvas et al. (2016) [51], who found a modest correlation (correlation coefficient -0.16 95% CI -0.21 to -0.10) between the severity of depression and CRF in their meta-analysis.

In summary, there is substantial evidence for an association between CRF and cardiovas-



cular disease as well as CRF and premature all-cause mortality. The associations were observed in prospective studies, meta-analysis and were validated in external populations. Hence, it is likely that there is a causal relationship between CRF and these health threats. It should also be emphasised that the effect sizes of these relationships are very high and comparable with other major risk factors such as tobacco smoking or dyslipidaemia. Therefore, CRF is a decisive modifiable risk factor which should be targeted by preventive medicine. Furthermore, adjusted analyses support the assumption that CRF is a risk factor that is independent of other major risk factors. Aside from cardiovascular disease and mortality, plenty of health threats are inversely associated with CRF. To conclude, CRF might be an essential target in the prevention of cardiovascular disease, mortality and other critical health threats.

## 2.2 Exercise tests in preventive medicine

This section outlines the general framework of CPET and the value of CPET in preventive medicine. Furthermore, the effort to increase CRF, which is a primary goal of preventive medicine and one of the important rationales to measure  $\dot{V}O_{2\text{peak}}$ , is outlined.

### General framework of exercise tests

CPET and the measurement  $\dot{V}O_{2\text{peak}}$  should be embedded into a more comprehensive assessment to rule out preexisting diseases that might affect patient safety [53, 1, 71]. This assessment includes anamnesis, resting electrocardiogram, blood pressure as well as the measurement of anthropometric characteristics and laboratory values such as carbon hydrate and lipid metabolism. During CPET, there is usually a recording of heart rate, blood pressure and the assessment of subjective symptoms. Blood pressure and exercise electrocardiogram are recorded to detect pathological reactions of the circulatory system and terminate CPET if appropriate. Subjective symptoms may either be the rating of perceived exertion (e.g. by Borg Scale, [53]), localised pain or angina pectoris. Taking capillary blood samples is also common to measure lactate levels at a given work rate. Based on lactate thresholds, precise recommendations for physical exercise can be conducted. A number of organisations have published extensive guidelines for CPET and its framework such as the ACSM (2014) [53], ATS/ACCP (2003) [1], AHA (2010) [7] and in German by Trappe & Löllgen (2000) [67] or DGSP\* (2007) [15].

### Exercise prescription

In chapter 2.1, evidence was presented to show that CRF is a substantial risk factor for a number of health threats. Garber et al. (2011) stated in their ACSM position that, “The scientific evidence demonstrating the beneficial effects of exercise is indisputable” [18].

---

\*Deutsche Gesellschaft für Sportmedizin und Prävention

From a public health perspective, it is interesting how CRF can be improved and whether and to what extent an improvement in CRF leads to a lower occurrence of adverse events. Therefore, a short overview of cardiopulmonary exercise and its health implications is outlined in the following.

CRF is largely dependent on physical activity and physical exercise. Improved fitness as a response to progressive exercise is a basic principle of training theory [16]. This was shown in a meta-analysis by Lin et al. (2015) [39], who used 27 studies of men and 25 studies of women. The exercise arrangements were different in the analysed studies including moderate and vigorous training. The mean overall response of relative  $\dot{V}O_{2\text{peak}}$  was quantified at 5.4 mL $O_2$ /min/kg (95% CI 4.3 to 6.5) in men and 3.2 mL $O_2$ /min/kg (95% CI 2.6 to 3.9) in women. The effect of exercise was stronger in subjects with a sedentary lifestyle compared to subjects with an active lifestyle, and it also increased with exercise duration per week. The authors also observed that exercise leads to desired effects on lipid and glucose metabolism. Those findings were confirmed by a meta-analysis of Huang et al. (2016) [24], who conducted a meta-analysis of older adults with a mean age of 68 years.

The ACSM's recommendations for cardiorespiratory endurance exercise in healthy individuals depend on the intensity of exercise [18, 53]. When moderate intensity is performed, 30 minutes per day on five days a week are recommended. Exercise with vigorous intensity should be performed on three days a week for 20 minutes per day. The intensity of endurance exercise should be defined based on heart rate reserve (HRR) or on  $\dot{V}O_{2\text{peak}}$  reserve ( $\dot{V}O_{2R}$ )\*. Using the oxygen uptake method, moderate exercise is defined as  $\dot{V}O_{2\text{rest}}$  plus 40% to < 60% of  $\dot{V}O_{2R}$  and vigorous exercise at  $\dot{V}O_{2\text{rest}}$  plus 60% to < 90% of  $\dot{V}O_{2R}$ . Persons with diseases such as heart failure, or athletes with high CRF are likely to benefit from different training arrangements which can be found in [53, p. 161] and [18].

To conclude, it is clear that cardiorespiratory fitness is a substantial independent risk factor. However, CRF is critically dependent on cardiorespiratory exercise and training [37, 18]. It is reasonable to measure CRF by acquiring  $\dot{V}O_{2\text{peak}}$  in preventive medicine to assess the participant's risk and in order to arrange endurance exercise systematically. For the interpretation of a participant's  $\dot{V}O_{2\text{peak}}$  as well as for exercise prescription, sex-specific and age-specific reference values based on a comparable population are essential.

## 2.3 Measurement of peak oxygen uptake

This chapter outlines the measurement tools needed to assess  $\dot{V}O_{2\text{peak}}$ . Different types of ergometers, test protocols, and end criteria are summarised. A special focus is placed on how the results of CPET are affected by its test modalities.

---

\*Heart rate reserve is defined as  $HRR = HR_{\text{max}} - HR_{\text{rest}}$ , where  $HR_{\text{max}}$  denotes the maximal heart rate and  $HR_{\text{rest}}$  denotes heart rate at rest.  $\dot{V}O_{2\text{peak}}$  reserve is calculated as  $\dot{V}O_{2R} = \dot{V}O_{2\text{peak}} - \dot{V}O_{2\text{rest}}$ , where  $\dot{V}O_{2\text{peak}}$  denotes maximal oxygen consumption and  $\dot{V}O_{2\text{rest}}$  denotes oxygen consumption at rest [53, p. 170].

### Cardiopulmonary exercise tests

There are several test arrangements that aim to assess  $\dot{V}O_{2\text{peak}}$ . Open circuit spiroergometry that is performed until maximal effort of the participant is a standard setting and has been described as the gold standard in measuring  $\dot{V}O_{2\text{peak}}$  [1, 7, 53]. Spiroergometry is the measurement of gas exchange during physical activity that is performed on an ergometer. Gas exchange is measured by wearing an airtight mask with a built-in pneumotachograph, which is a low-resistance valve through which respiration is conducted.  $\dot{V}O_2$  is obtained using respiratory rate, tidal volume and the composition of inhaled and exhaled gas. Recent spiroergometry systems provide a breath-by-breath analysis of the composition of gas. In order to measure maximal oxygen uptake, the work rate is increased over time until the maximal volitional work rate of the participant is finally achieved.  $\dot{V}O_{2\text{peak}}$  is then calculated as the average  $\dot{V}O_2$  of the final exercise period in order to decrease noise. Past studies have suggested using an average over at least the final 30 seconds of the exercise test [50, 1]. In some settings, CPET cannot be performed until exertion. In that case, exercise is performed until a predefined termination point and  $\dot{V}O_{2\text{peak}}$  is extrapolated based on e.g. heart rate at a given work rate using prediction equations. Such submaximal testing is an option if maximal testing is not safe for patients with preexisting cardiovascular disease [53]. However, the focus of the following is on maximal exercise tests, as those were used in the present study.

### Type of ergometer

Graded exercise testing requires that the participant performs a standardised dose of work. Two options that were applied frequently in past studies are cycle ergometers and treadmills [53, 1, 71]. Some characteristics and differences of cycle and treadmill ergometers are summarised below.

The work rate is adjusted differently in cycle and treadmill ergometers. In treadmills, this is usually done using speed and elevation of the device, whereas in cycle ergometers, the resistance is modified. In electronically braked cycle ergometers, the work rate can be adjusted very accurately by controlling the resistance of the device. When treadmill ergometers are used, on the other hand, the work rate can only be estimated based on the speed and weight of the subject.

There is a substantial difference in the assessed  $\dot{V}O_{2\text{peak}}$  between cycle and treadmill ergometers.  $\dot{V}O_{2\text{peak}}$  was estimated to be 5 to 10% higher when measured on treadmill ergometers [1, p. 218]. The difference between both ergometer types was even greater in two studies of the same population using both types of ergometers. A 35 year-old man showed a median relative  $\dot{V}O_{2\text{peak}}$  of 42 mL $O_2$ /min/kg using treadmill ergometers and 30 mL $O_2$ /min/kg using cycle ergometers [30, 31]. It was assumed that these differences occur because CPET on cycle ergometers is often terminated due to local muscle fatigue. Additionally, a larger number of muscles are active in treadmill tests, which increases  $O_2$  consumption. Because of the large difference between cycle and treadmill ergometer, it is

essential to be informed about the type of ergometer that was used for reference values. Another difference in ergometers concerns additional testing during the exercise test like electrocardiograms or drawing capillary blood samples. It is more feasible and there are less artefacts when cycle ergometers are used because the whole body of the subject is in motion on treadmill ergometers. To address this issue, it becomes necessary to use specific test protocols to treadmill ergometers. When capillary blood samples and treadmill ergometry are desired, discontinuous protocols are often applied and the blood samples are drawn during phases without exercise.

In conclusion, ATS/ACCP Guidelines recommend cycle ergometers as the preferred mode of exercise [53]. In order to increase external validity, treadmill or field test may be desired, nevertheless.

### Protocols of incremental exercise tests

Exercise tests are usually based on an incremental work rate. The ATS/ACCP guidelines (2003) [1, p. 224] describe four different general types of CPET protocols. The protocols are particularly defined by the amount of increment per time unit.

1. In progressive or continuous incremental protocols (also: “ramp protocol”), the work rate is adjusted in short periods of time or continuously.
2. In multistage incremental protocols, the work rate is constant for a defined period of time (e.g. three minutes) and increased when the time period is over.
3. In protocols using constant work rate, the work rate is constant over a period of usually < 30 minutes.
4. In discontinuous protocols, the work rate is constant over a period of time (e.g. three minutes), then a resting period is inserted after which an increased work rate is applied. This type of protocol is sometimes necessary when lactate measurements from capillary blood samples are desired and treadmill is the preferred mode of physical activity.

In clinical practice, however, plenty of versions of the protocols mentioned above are used. The ATS/ACCP guidelines (2003) [1, p. 224] provide an overview of the most common protocols.

The amount of increments per time has to be selected carefully and has to be adjusted to the participant’s level of fitness. An increase in work rate leads indeed to an increased oxygen consumption, but the oxygen consumption increases with some delay. Therefore, a rapid increase in work rate might lead to inaccurate results in gas exchange measurement. According to ATS/ACCP guidelines (2003), increments should be between 5 to 25 Watt/minute [1]. This should yield a period of 8 to 12 minutes for CPET, overall. By the use of predicted  $\dot{V}O_{2peak}$ , the increase in work rate can be adjusted to achieve an estimated overall time of 10 minutes [1, p. 225].

### Termination of exercise tests

The exercise test should be continued until the participant reaches the maximal volitional effort. This is required because  $\dot{V}O_{2\text{peak}}$  cannot be interpreted as the highest volitional oxygen uptake if CPET was terminated due to other reasons. In order to quantify the participant's effort, several measures have been used. Midgley et al. (2007) [45] reviewed and commented end criteria for maximal effort that were used in past studies. Some of the commonly used criteria were based on the slope of  $\dot{V}O_2$  during incremental exercise, the heart rate, the respiratory exchange ratio ( $\text{RER} = \frac{\dot{V}CO_2}{\dot{V}O_2}$ ) or blood lactate level. The most common criteria of the reviewed studies are displayed in table 2.1.

**Table 2.1:** End criteria for maximal effort in exercise tests.

Criterion	Value used
Heart rate	$\geq 90\%$ of APMHR
RER	$\geq 1.1$
Blood lactate	$\geq 8$ mmol/L
$\dot{V}O_2$ plateau	$\leq 150$ mL $O_2$ /min

**Note:** APMHR = age-predicted maximal heart rate  
RER = respiratory exchange ratio  
Table modified according to Midgley et al. (2007) [45]

The estimation of the age-predicted maximal heart rate (APMHR) can be useful as an end criterion, but several calculation methods have been used. Common estimation methods have been  $\text{APMHR} = 220 - \text{age in years}$  or the less rigorous equation  $\text{APMHR} = 200 - \text{age in years}$  [59, 15]. The latter criterion might be more suitable for cycle ergometry since the exercise test is usually terminated earlier compared to treadmills (section 2.3). A more data-driven approach was published by Tanaka et al. (2001) [66], who performed a meta-analysis of studies with an overall sample size of 18,712 participants. The linear regression model  $\text{APMHR} = 208 - 0.7 * \text{age in years}$  showed a high goodness-of-fit and a correlation coefficient of  $r = -0.9$ .

In addition to end criteria for maximal effort, there are health-related contraindications for the continuation of the exercise test [53, p. 87] such as i) symptoms of hypoxia (e.g. dyspnoea, angina pectoris, cyanosis) ii) unphysiological response of blood pressure or heart rate (decreasing blood pressure or heart rate despite increasing work rate, excess of blood pressure) or arrhythmia in the electrocardiogram.

## 2.4 Reference values for peak oxygen uptake

It is essential to compare the individual results of exercise testing with reference values that are drawn from a comparable population and are stratified by at least age and sex [33]. This is due to the strong impact of sex and age on the CRF. Exercise test results are only diagnostically conclusive if sex-specific and age-specific reference values from a comparable population are considered [1, 7, 50].

Several attempts have been made to produce reference values for CPET. Kim et al. (2007) [33] summarised eleven studies that provided reference values for exercise capacity and conducted an external validation of the models. The authors had access to CPET results of 13,089 men and 9,177 women and were able to assess the subsequent deaths of the study subjects using Social Security Death Index for a median follow-up time of five years. The authors concluded that all reference values were more or less accurate for the prediction of premature all-cause mortality. This was done by classifying the individual results of a CPET participant in relation to the reference values. Low  $\dot{V}O_{2\text{peak}}$  in comparison to the reference values was then used as the predictor of premature all-cause mortality. Reference values that were adjusted for age and sex, nevertheless, performed better compared to simple cut-off values. The best predictions of death for men and women were based on a Veterans Affairs cohort [46] and the St. James Take Heart Project [20], respectively.

Paap and Takken (2014) performed a systematic review of studies that presented reference values for  $\dot{V}O_{2\text{peak}}$  in adults [50]. The key points of this review are summarised below: Overall, 35 studies from 1985 to 2013 were included in the analysis. In the reviewed studies, the sample size ranged from 25 to 2,263 and the age from 4 to 95 years, respectively. 23/32 (71.9%) studies with a reported design were prospective, and 17/32 (53.1%) were population-based. The study design was unclear in three studies. Most studies were conducted in European and North American countries, resulting in predominantly Caucasian participants. The authors described the need of reference values especially for South American, Middle Eastern, African and Asian populations. 22/35 (62.9%) studies used only cycle ergometers, 12/35 (34.3%) studies only treadmill ergometers and 1/35 (2.9%) study used both ergometers. The reviewed studies used different CPET protocols. The most prevalent protocols were multistage or ramp protocols, but different individualised protocols were also present. Most studies (21/32; 65.6%) used either breath-by-breath or mixing chamber systems with an at least 30-second time-averaging of  $\dot{V}O_2$  of the final exercise test period to determine  $\dot{V}O_{2\text{peak}}$ . Only 7/23 (30.4%) studies with reported time averaging used < 30 seconds. Tobacco smokers were excluded by 7/29 (24.1%) studies in which smoking status was reported, and only one study took smoking status into account in the analysis. The authors created a rating system based on ATS/ACCP guidelines for exercise testing [1] to grade the quality of each reviewed study. They listed 14 dichotomous criteria (yes/no) and coded yes as 1 and no as 0. Studies with  $\geq 10$  points were considered “high quality”, 7 to 9 “moderate quality” and < 7 “low quality”. Some of those criteria were also used to assess the quality of the present study (table 5.1). The

14 items can be categorised into four groups:

- Study design
  1. Prospective design
  2. Proper randomization
  3. Community-based sampling of the study population
  4. The number of study subjects is at least as high as calculated in the sample size estimation
- Characteristics of CPET
  5. Measurement of gas exchange data and  $\text{VO}_{2\text{peak}}$  is averaged over time to avoid noise (preferably  $\geq 30$  seconds intervals)
  6. CPET was performed using breath-by-breath or mixing chamber analysis according to ATS/ACCP guidelines [1]
  7. Quality control was performed according to ATS/ACCP guidelines [1]
- Important background reported
  8. Level of physical activity reported
  9. Exercise testing protocol described
- Data analysis and reporting
  10. External validation of the statistical model
  11. Adequate fitting of the regression model was performed
  12. Analysis was stratified by racial group
  13. Smokers were excluded
  14. Confidence limits were given for descriptive statistics

The study that scored the most points (11) was conducted by Itoh et al. (2013) [26]. The publication of Edvardsen et al. (2013) [17] scored 10 points. For this reason, these two studies are summarised below:

Itoh et al. (2013) performed a prospective, community-based, multi-centre study in Japan. The allocation of study subjects into groups was randomised. The final sample consisted of 749 healthy Japanese participants aged 20 to 78 years. The authors excluded smokers, subjects with a body mass index (BMI) of  $< 17.6 \text{ kg/m}^2$  or  $\text{BMI} > 28.6 \text{ kg/m}^2$ , subjects who exercise regularly ( $> 2$  times per week), subjects with cardiopulmonary pathologies and more. The measurement was done using treadmill and cycle ergometers, and ramp protocols were applied.  $\text{VO}_{2\text{peak}}$  was calculated as the average of the last 30 seconds of exercise. RER was measured, and subjects with poor effort ( $\text{RER} < 1$ ) were excluded from

the analysis. Differences in protocol and cycle versus treadmill ergometer were compared. Multiple linear regression was used to model the effects of age, sex and the type of test protocol. The results were displayed as two-dimensional line charts and scatter plots. The analysis using linear ordinary least square regression, however, was a limitation of this study because it only allows the estimation of conditional means and not percentiles. This approach does not consider the distribution of  $\text{VO}_{2\text{peak}}$  for a given age and sex. A further limitation was the extensive exclusion of participants which lead to a final study population which was constructed artificially and might decrease external validity. The authors obviously wanted to derive reference values from healthy individuals, who did not exercise on a regular basis or have any potential risk factors.

A study performed by Edvardsen et al. (2013) [17] was based on a multi-centre, population-based random sample from Norway. The final sample size after exclusion of study subjects with poor effort ( $\text{RER} < 1.10$  or Borg score  $< 17$ ) was 759 men and women aged 20 to 85 years. Smokers were not excluded in this study. CPET was performed only on treadmill ergometers. Gas exchange measures were reported as the mean  $\text{VO}_2$  of the final 30 seconds of the exercise test. The reference values were also presented as the result of linear ordinary least square regression, and no percentiles were reported.

The results of two recent studies that were based on a large sample were not included in the systematic review of Paap and Takken [50]. Kaminsky et al. (2015) and Kaminsky et al. (2017) [30, 31] presented the results of the “Fitness Registry and the Importance of Exercise National Database (FRIEND)” database. The database was based on CPET data from several laboratories in the USA. Men and women who participated in an exercise programme or research study at the age of 20 to 79 years were eligible for the study. Percentile values of treadmill [30] as well as cycle [31] ergometer were published. Kaminsky et al. (2015) [30] recorded CPET results from eight laboratories in the time frame from 2014 to 2015. Treadmill ergometers were used, and individuals who achieved exertion ( $\text{RER} > 1.0$ ) were included in reference values. Overall, 7,783 subjects (4,611 men, 3,172 women) were eligible for the analysis. On the other hand, Kaminsky et al. (2017) [31] published reference values based on CPETs using cycle ergometers. 4,494 subjects (1,717 men, 2,777 women) from ten laboratories participated in data acquisition between 2014 to 2016. Exertion was assumed if  $\text{RER} > 1.1$ .

Reference values based on a German sample were also derived. Koch et al. (2009) [34] conducted a population-based study in Pomerania, a north-eastern region of Germany. A representative sample of 3,300 subjects was drawn from the whole population ( $n = 212,157$ ). Of those 3,300, 1,708 subjects agreed to participate in CPET on a cycle ergometer. The volunteers were younger, healthier and tobacco smoking was less prevalent compared to the Pomeranian population. This might be an indicator of selection bias. The selection of healthy participants might lead to increased reference values. A rigorous exclusion of cases such as smokers, obese subjects and subjects with cardiac arrhythmia led to a final sample size of 534 (253 men and 281 women). CPET was performed using a stepwise protocol on cycle ergometers with an increment of 16 W/min. The analysis



was done using quantile regression with maximal oxygen uptake as dependent and age as independent variables. Age was modelled as a categorical factor and was adjusted for sex and BMI. Regression coefficients and plots with normal ranges were supplied.

To conclude, several attempts have been made to create reference values. The quality of the existing reference values was diverse. Nevertheless, reference values were valuable for the prediction of all-cause mortality.

## 3 Material & Methods

The material and methods of the present study are outlined in this chapter. Section 3.1 describes the study design as well as the composition and structure of this study's participants. Furthermore, it summarises the general set-up of data acquisition. The measurement of  $\dot{V}O_{2\text{peak}}$  is described in section 3.2. Data management and the construction of the final datasets is described in section 3.3. Lastly, the methods to assess the generalisability of this sample, and the statistical methods are outlined in the sections 3.4, and 3.6.

### 3.1 Study design and participants

#### General framework

The data acquisition of the present study was conducted by Prevention First<sup>®</sup>, a quality network of institutions offering primary preventive health screenings. There were three study centres in the German cities Rüdeshheim, Frankfurt and Munich. Data acquisition ended at all three sites in 2015 but started in different years. Rüdeshheim was the first centre to systematically record the data in 2001, followed by Frankfurt in 2006 and Munich in 2008.

#### Study design

This was a cross-sectional, registry-based study analysing routine files recorded in the course of primary preventive health screenings.

#### Participants

A proportion of 95% of all recorded CPET participants of this study was acquired in the course of workplace health promotion programmes. More than 100 local companies such as mid-sized companies, banks, insurance companies or business consulting participated in these programmes. The individuals in this group were predominantly white-collar workers with office jobs and a sedentary working environment. The other 5% of the recorded individuals comprised persons with private health insurance or persons who purchased the health screening as direct payers.

#### Inclusion criteria and exclusion criteria

All participants underwent pre-exercise evaluation before qualifying for participation in CPET. According to ACSM guidelines [53, p. 40], medical history, physical examination and laboratory tests were applied. Subjects who did not meet clinical exclusion criteria

(e.g. hypertensive emergency, acute infection) and who were free of acute complaints subsequently performed CPET with the goal of maintenance until exertion. If exclusion criteria were present, CPET was offered at a later time. Experienced test instructors and physicians supervised CPET. table 3.1 outlines inclusion and exclusion criteria of the present study. Some participants contacted Prevention First<sup>®</sup> more than one time, but for this cross-sectional analysis, only the first contact of a participant and Prevention First<sup>®</sup> was considered.

**Table 3.1:** Inclusion and exclusion criteria of the present study.

Inclusion criteria	Exclusion criteria
<ul style="list-style-type: none"> <li>●Men and women aged 25 to 69 years</li> </ul>	<ul style="list-style-type: none"> <li>●Age &lt; 25 or age ≥ 70 years</li> </ul>
<ul style="list-style-type: none"> <li>●First contact of the participant and Prevention First<sup>®</sup></li> </ul>	<ul style="list-style-type: none"> <li>●Follow-up examinations of the participant and Prevention First<sup>®</sup></li> </ul>
<ul style="list-style-type: none"> <li>●Participants of workplace health promotion programmes, members of private health insurance or direct payers</li> </ul>	<ul style="list-style-type: none"> <li>●Persisting contraindications for CPET such as acute myocardial infarction or unstable angina pectoris. For a comprehensive list, see [1, p. 227]</li> </ul>
<ul style="list-style-type: none"> <li>●Participation in preventive health screenings including CPET at Prevention First<sup>®</sup> in Rüdeshheim, Frankfurt or Munich</li> </ul>	<ul style="list-style-type: none"> <li>●Participants who did not agree with the use of personal data for scientific purposes or did not provide informed consent</li> </ul>
	<ul style="list-style-type: none"> <li>●Missing value in one of the characteristics i) age or ii) peak oxygen uptake or iii) study centre (location of data acquisition: Rüdeshheim, Frankfurt or Munich)</li> </ul>

### Ethics approval and informed consent

Participants who agreed with the use of their personal data for scientific purposes and provided informed consent (appendix A.4) were considered for this study. Their CPET results and pre-exercise evaluations were recorded in a computerised database.

According to the AGENS\* guidelines for secondary data analysis, ethical approval was not needed as this was a secondary data analysis using routine files [3, p. 3]. The participant's data were pseudonymised before data analysis, and there was no particular treatment of participants who were analysed in this study.

---

\*AGENS = Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) und der Deutschen Gesellschaft für Epidemiologie (DGEpi)

## 3.2 Measurement of sample characteristics

### Primary outcome measures: peak oxygen uptake

This study determined CRF as the highest volitional oxygen consumption during CPET. The consumption of oxygen was assessed as a volumetric flow rate ( $\dot{V}O_2 = \frac{\delta V}{\delta t}$ ). End criteria of the exercise test were not recorded in the main dataset but were acquired retrospectively for a random sub-sample (section 3.3). Therefore, the term  $\dot{V}O_{2\text{peak}}$  rather than  $\dot{V}O_{2\text{max}}$  was used to describe the highest oxygen uptake during CPET. This was in concordance with the recommendations of CPET guidelines [1].  $\dot{V}O_{2\text{peak}}$  was assessed as an absolute value (absolute  $\dot{V}O_{2\text{peak}}$  measured in  $\text{LO}_2$  per minute) and relative to the participant's body weight in kilograms (relative  $\dot{V}O_{2\text{peak}}$  measured in  $\text{mLO}_2$  per minute per kilogram of body weight).  $\dot{V}O_{2\text{peak}}$  was defined as the mean  $\dot{V}O_2$  of the last 10 seconds of the exercise test. Such averaging was recommended to avoid noise and artefacts of measurement [1, 50].

### Further sample characteristics

The sample characteristics were acquired in the course of the pre-exercise evaluation prior to the exercise test. The age of the participants was recorded in years. Tanita TBF 410 (Tanita, Tokyo, Japan) body composition analysers were used to assess body weight and to estimate body fat via bioelectrical impedance analysis. Body fat was also estimated by measuring skinfold thickness using Lange Skinfold Calipers (Beta Technology, Cambridge, Maryland, USA). Skinfold thickness was measured at three sites according to Jackson Pollock (1985) [27] (men: chest, abdomen, thigh; women: triceps, suprailium, thigh). Blood pressure was acquired after the participant rested in a sitting position for at least 5 minutes. Manometric blood pressure gauges by BOSO (Jungingen, Germany) were used, and the average of two consecutive measurements was recorded to the nearest 2 mmHg. Overweight, obesity, and hypertension were defined according to the World Health Organization (WHO) (overweight:  $\text{BMI} \geq 25 \text{ kg/m}^2$ , obesity:  $\text{BMI} \geq 30 \text{ kg/m}^2$ , hypertension: either systolic blood pressure  $\geq 140 \text{ mmHg}$ , or diastolic blood pressure  $\geq 90 \text{ mmHg}$  [76, 74]). Participants were instructed not to eat on the morning prior to their exercise test in order to draw fasting blood samples. The analyses of the blood samples were performed by two accredited laboratories (study centres Rüdeshheim and Frankfurt: Labor Dr. Riegel, Wiesbaden, Germany; study centre Munich: Synlab, Augsburg, Germany). Dyslipidaemias were defined according to guidelines by the ESC/EAS\* [9, table 10]. Elevated low-density lipoprotein (LDL) cholesterol was defined as fasting LDL cholesterol levels  $\geq 115 \text{ mg/dL}$ , reduced high-density lipoprotein (HDL) cholesterol was defined as HDL cholesterol  $\leq 40 \text{ mg/dL}$ , elevated triglycerides were defined as triglycerides  $\geq 150 \text{ mg/dL}$ . Diabetes mellitus was defined according to the WHO as either fasting blood

---

\*European Society of Cardiology (ESC) and European Atherosclerosis Society (EAS)

glucose levels  $\geq 126$  mg/dL or glycated haemoglobin (HbA<sub>1c</sub>)  $\geq 6.5\%$  [57].

### Exercise test modalities

Open circuit spiroergometry using cycle ergometers was performed to apply an incremental workload to the participant. The CPET protocol was selected according to the requirements of each measurement and was adapted to the estimated level of fitness of the participant. Multi-stage protocols were used if capillary blood samples for measuring blood lactate levels were drawn and ramp protocols were used if no capillary blood samples were drawn. The duration of CPET was intended to be 4 to 6 stages of 3 minutes each in multistage protocols or 12 to 18 minutes in ramp protocols. The increments in work rate were selected by an experienced CPET instructor or physician before the CPET according to the participant's level of fitness to fulfil these conditions. Before the first application of a workload, the participant pedalled with no resistance for 3 minutes.

Gas exchange was measured with breath-by-breath analysis using the Ganshorn Powercube system (Ganshorn Medizin Electronic GmbH, Niederlauer, Germany). Ganshorn LF8 V8.5 software and the previous versions were used for analysis of the results. Calibration of the gas exchange measurement system was performed daily in concordance with the manufacturer's instructions, and approximately 3 to 4 CPETs were performed per day using the same calibrated system. Cycle ergometers were calibrated once a year and met the German directives for medical devices. At all study centres, quality control was performed according to a DIN EN ISO 9001 certified quality management system.

If there were no contraindications during the exercise test [53, p. 87], the increments of work were continued until the maximal volitional work rate was achieved. Blood lactate levels, RER, and maximal heart rate were recorded to get information about the participant's effort. Adequate exertion was assumed if one of the end criteria was achieved: i) blood lactate levels  $\geq 8$  mmol/L or ii) RER  $\geq 1.10$  or iii) maximal heart rate  $\geq 90\%$  of the age-predicted maximal heart rate ([45], section 2.3). The measures of maximal effort were used to evaluate the effort of the participant instantly and were noted in the participant's medical chart. However, the end criteria were not recorded in the main study database.

## 3.3 Data management

### Data sources

Health screening data of all eligible participants who met the inclusion and exclusion criteria (table 3.1) were recorded in a computerised database called "PF Studie". All data were pseudonymised using a unique identification number. A key file to link the personal data and the unique identification number was stored at the Prevention First<sup>®</sup> centre in Rüdeshheim. For this cross-sectional analysis, only the baseline data was exported to the main dataset. If there was more than one contact of a subject and Prevention

First<sup>®</sup>, only the first contact was included in the dataset. The variable “study centre” and some CPET modalities were not recorded in the “PF Studie” database. CPET modalities were documented in medical records and the study centre was recorded in a second database of Prevention First<sup>®</sup> (“PF Patient”). This additional information was added retrospectively. The variable study centre could be added using probabilistic record linkage as it was stored in a database, but CPET modalities had to be acquired manually from medical records. Due to a large number of observations, the additional CPET test results could only be provided for a random subset of the main dataset with an a priori calculated sample size of  $n = 252$  participants (section 3.3). An overview of the data sources is displayed in fig. 3.1.

### Random sample

Exercise test modalities were not recorded in the main study dataset; rather, these were included in medical records that were stored separately. This information had to be acquired manually. As the acquisition of the information for the entire main study dataset would have been excessively time-consuming, a random sample was drawn from the main study dataset. The variables were added manually and a second dataset (“random sample”) was created. Sample size calculation was used to estimate the size of the random sample. The minimum sample size was calculated for a proportion and 95% CI of subjects who achieved exertion ( $\geq 90\%$  of age-predicted maximal heart rate or  $RER \geq 1.10$  or blood lactate levels  $\geq 8$  mmol/L). The margin of error (ME) for a single proportion with the assumption of normal approximation was used [72]:

$$ME = z_{1-\frac{\alpha}{2}} * \frac{\hat{p} * (1 - \hat{p})}{n} \quad (3.1)$$

Where ME denotes the margin of error,  $z_{1-\frac{\alpha}{2}}$  is the critical value of the standard normal distribution that corresponds to the level of confidence,  $\hat{p}$  is the point estimator for the expected sample proportion and  $n$  is the a priori sample size of the random sample. Using ME of 0.05 and 95% CI ( $\alpha=0.05$ ), the sample size is calculated as:

$$n = \frac{1.96^2 * \hat{p} * (1 - \hat{p})}{0.05^2} \quad (3.2)$$

For a proportion of 80% of all individuals reaching exertion,  $n = 246$  and for 90% of all individuals reaching  $VO_{2peak}$ ,  $n = 139$ . At the start of data analysis, the main outcome variable (relative  $VO_{2peak}$ ) included 218/10189 (2.1%) missing values. A subsequent data acquisition yielded the described numbers of non-missing cases (fig. 3.2). Therefore, the sample size was adjusted for the amount of missing values. Hence, the final sample size was calculated as:

$$n = \frac{246}{1 - 0.021} \approx 252 \quad (3.3)$$

In order to test the representativity of the random sample for the full dataset, a set of characteristics (A.1) were compared between the random sample and the full dataset. However, none of the compared characteristics were significantly different between the entire main study dataset and the random sample.

### Record linkage

In contrast to the CPET modalities, information on the study centre was acquired from the second database using record linkage. The information on the study centre was stored in the database “PF Patient” and had to be linked with the main dataset, which was based on the database “PF Studie”. However, as there was no key identifier to link both databases, record linkage was necessary to merge “PF Studie” and “PF Patient”. As the data were typed into the databases separately by hand, it was also necessary to consider typing errors. Probabilistic record linkage was selected to address these conditions.

The participant’s name and sex were present in both databases and could be used for record linkage. Birthdate was recorded in “PF Patient” but not in “PF Studie”, which only included the year of the first contact with Prevention First<sup>®</sup> as well as the age in years at the time of this contact. Based on this, the participant’s birth year could be determined with an inaccuracy of one year. The calculated birth year was correct if the contact with Prevention First<sup>®</sup> was later than the subjects birthday and it was one year too low if the contact was before the subjects birthday. This inaccuracy was also necessary to consider in probabilistic record linkage. In light of the above-mentioned requirements for record linkage, two software programmes were considered: i) the R package `recordlinkage` [8] and ii) “Fine-grained record linkage (FRIL)” software [29]. FRIL software performed better using the specifications outlined in 3.2.

**Table 3.2:** Specifications of record linkage.

Variable	Linkage algorithm	Weight
First name	Edit distance	40%
Last name	Edit distance	40%
Year of birth	Numeric distance	10%
Sex	Equal fields	10%

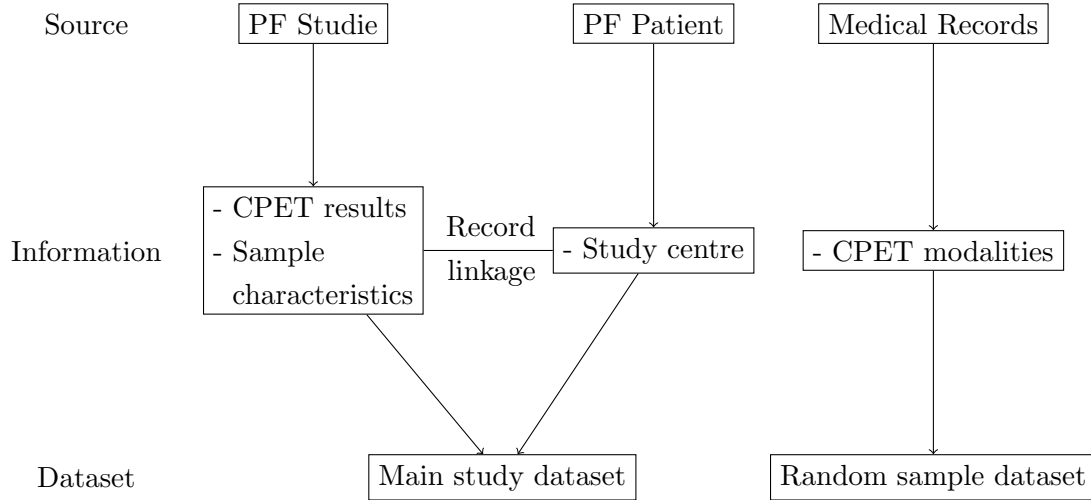
**Note:** Acceptance level was set at 90; FRIL software was used for record linkage [29].

### Additional data acquisition

After drawing the random sample (section 3.3), additional data acquisition was conducted for the full dataset aiming to reduce the number of missing values in the variables age

as well as absolute and relative  $\dot{V}O_{2\text{peak}}$ . This yielded the final main study dataset as described in fig. 3.2.

**Figure 3.1:** Data sources and information.



### 3.4 Representativity

The representativity of the present study population was analysed by comparing eligible characteristics to data from DEGS1\* [19]. DEGS1 is a study based on a representative sample of the German population and was performed by the German governmental organisation which is responsible for disease control and prevention (RKI<sup>†</sup>). The information needed for the comparison with the present study was based on a cross-sectional study sample with an enquiry period from 2008 to 2011. The study acquired a sample of 8,152 adult subjects using a multistage-sampling process. The participants of the study were more than 18 years of age, and the sample was representative of the German population from 2011.

Variables from the present study were eligible for comparison with DEGS1 if i) the information was recorded in “PF Studie” as well as in DEGS1, ii) the variable was binary, coded with yes/no, iii) the unit of measurement of this characteristic was identical or could be transformed to be identical, iv) 95% confidence intervals were provided for the proportions in DEGS1, and v) the results were reported separately for men and women in DEGS1. Eligible variables according to these criteria are displayed in table 3.3.

\* “Studie zur Gesundheit Erwachsener in Deutschland”

† “Robert Koch-Institut”



**Table 3.3:** Eligible variables of the present study for comparison with DEGS1.

Characteristic	DEGS1 source
Smoking status	Lampert et al. (2013) [38]
Ex-smoker	Lampert et al. (2013) [38]
Overweight	Mensink et al. (2013)[43]
Obesity	Mensink et al. (2013) [43]
High blood pressure	Neuhauser et al. (2013) [48]
<b>Note:</b> Overweight: BMI $\geq 25$ kg/m <sup>2</sup>	
Obesity: BMI $\geq 30$ kg/m <sup>2</sup>	
Hypertension: systolic blood pressure $\geq 140$ mmHg or diastolic blood pressure $\geq 90$ mmHg	

The variables of the present study were transformed to directly age-standardised proportions using the R package `epitools` [2] to compare the results of DEGS1 to the present population. The German population as reported in the census 2011 [62] was selected as the standard population for age standardisation. This was done because the results of DEGS1 were age-standardised for this population [38, 43, 48]. Consequently, only the proportions of the present population had to be transformed to achieve comparability.

### 3.5 Sample description

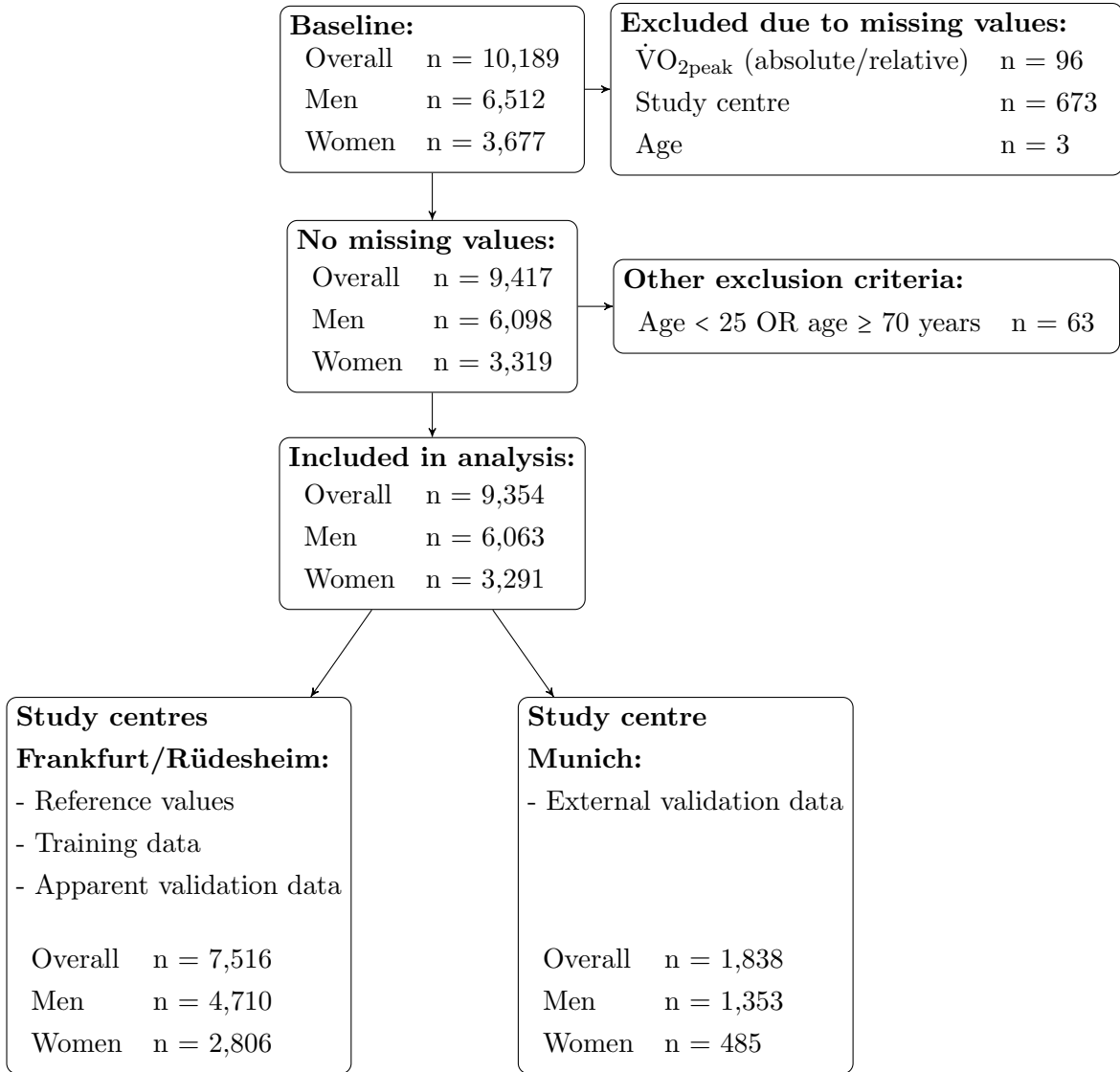
To analyse differences in the study centres and to perform external validation, the full dataset was separated into two datasets using i) data recorded in Rüdeshheim or Frankfurt and ii) data recorded in Munich. The separation was conducted in this form as there were participants who could be assigned to either Rüdeshheim or Frankfurt but not definitely to one of both locations. This was because some participants appeared in both locations because they might have visited both study centres due to the spatial proximity of Rüdeshheim and Frankfurt. Participants who could not be assigned definitely to either Rüdeshheim/Frankfurt or Munich, on the other hand, were excluded from reference values. Furthermore, participants with missing values in the variables age or absolute as well as relative  $\dot{V}O_{2\text{peak}}$  were excluded.

In their systematic literature review, Paap & Takken (2014) [50] recommended to exclude smokers, and earlier studies [34, 26] also excluded obese subjects in order to produce reference values that are representative for a healthy population. Other reference values did not exclude smokers or obese subjects [30, 17, 30, 53]. For purpose of this study, smokers and obese subjects were not excluded in the main analysis, but subgroup analyses were performed. A flow chart of excluded cases is displayed in fig. 3.2.

Participants aged 25 to 69 years were included in the present analysis. Participants aged  $\geq 70$  years or  $< 25$  years were excluded. The participants were excluded because the integer variable “age in years” had to be transformed to a categorical variable with 5-year age classes to perform the apparent and external validation of the present study (section 3.6). In this analysis, age class as an ordinally scaled variable was modelled as a

metric predictor. This approach was critically discussed in the literature and required the assumption of a metrically scaled variable [68]. Comparatively few cases of the present study were recorded in the age groups of  $< 25$  years and  $\geq 70$  years, and therefore the reduction of the sample size was approved to perform the validation analysis as mentioned above.

The final study dataset was compared with the cases that were excluded due to missing values to analyse if there was a selective dropout of participants. Selective dropout could bias the final results and would be present if participants with specific characteristics were more likely to be excluded. Selective dropout can be analysed by comparing the excluded and the included cases (table A.2). In the present study, some statistically significant differences were present between both groups. However, the extents of the differences were low, and the significant results were likely due to high numbers of cases. Selective dropout that severely impacts the results of the present study was therefore unlikely.

**Figure 3.2:** Flow chart showing numbers of cases included in reference values.

## 3.6 Statistical methods

### General framework

The statistical analyses were performed using R software version 3.3.1 [54]. Analyses were stratified for sex and adjusted for age if adequate. All 95% confidence intervals were approximated using ordinary non-parametric bootstrapping with 10,000 pseudo-random bootstrap samples [14, p. 120-123]. Statistical significance was assumed for P values of less than 0.05 or when 95% confidence intervals did not overlap using a two-sided significance level of  $\alpha = 0.05$ . There was no correction of P values for multiple testing. The computational code of statistical analyses should be published to ascertain reproducible research [25, 52]. The code of the present analysis is outlined in chapter 7 and can be accessed via the online appendix of this study at <https://github.com/rappdaniel/vo2peak>.

### Descriptive statistics

A subset of all variables in the original dataset was selected for descriptive statistics and for multivariable quantile regression modelling. Quantitative variables were described as median [1st quartile; 3rd quartile] and qualitative variables as n (%). Comparisons of two groups were conducted using Mann-Whitney U test for quantitative variables and  $\chi^2$  test for qualitative variables. Fisher's exact test was not considered as expected cell counts with  $n < 5$  under the null hypothesis were not present. Two-way descriptive tables were performed using the R package `compareGroups` [65]. For quantitative variables, the assumption of normal distribution was analysed using (supplementary tables, [49]):

1. box plots and histograms
2. skewness
3. quantile based skewness [22, p. 14]

Kolmogorov-Smirnov and Shapiro-Wilk tests were not applied as the sample size was large and particularly Shapiro-Wilk tests cannot be computed for samples with  $n > 5000$  using the described version of the R software [54].

### Quantile regression modelling and conducting nomograms

Several approaches were used to arrange age-adjusted reference values as nomograms. Kaminsky et al. [30, 31] categorised age in 10-year classes and calculated quantiles within each class. The quantiles were also visualised using box plots for each age class. Well-known and widely used nomograms are the WHO's Child Growth Standards for paediatrics [75]. For those nomograms, age was modelled as a continuous predictor, and cubic spline smoothing was used to fit the regression models to the data [73]. Koch et al. (2009) [34] published quantile reference values and nomograms for maximal oxygen uptake and used polynomial quantile regression. Accordingly, the present study used quantile regressions (R package `quantreg` [36]) to model conditional quantiles and to estimate predicted quantile values of maximal oxygen uptake depending on age as a continuous predictor. Age as a linear and polynomial predictor as well as b-spline smoothing (R package `splines` [54]) were compared using Akaike Information Criterion (AIC) and also in apparent (section 3.6). For nomograms, the quantiles 0.05, 0.1, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8, 0.9, and 0.95 were selected.

### Validation of regression models

After fitting quantile regression, the models' performance was tested and compared with empirical data. Validation of the statistical models was also a quality criterion in the systematic review of Paap & Takken (2014) [50]. Two major approaches were applied in the underlying data analyses: i) in apparent validation, the regression fits were compared

with the original data to which the regressions were fitted [64, p. 300]. This usually leads to results that are too optimistic. Therefore, ii) in external validation the regression fits were compared with new subjects who were not used to fit the regression models. A graphical technique of model validation is the calibration plot [56] where predicted and observed values are plotted. Additionally, linear ordinary least square regression was performed in the calibration data using observed and predicted values of  $\dot{V}O_{2\text{peak}}$  as dependent and independent variables, respectively.

To perform external validation in the present analysis, the dataset was split by study centre prior to regression modelling. Data from Rüdeshheim and Frankfurt were used as training data to fit quantile regressions and data from Munich as validation data for external validation. Compared to other predictive models, there were some challenges in the validation of the present regression models. i) There were 13 regression models (for the quantiles 0.05, 0.1, ..., 0.95) for men and women each, ii) the goal of the present analysis was not to get the best prediction of maximal oxygen uptake for a given age but to predict quantiles for a given age. Due to these challenges, regular calibration plots could not be performed. To address these problems, age was recoded into 5-year age classes for model validation and empirical quantiles of  $\dot{V}O_{2\text{peak}}$  were calculated within each age class. Subsequently, quantile regression models were fitted using the recoded age classes as a metrically scaled predictor (section 3.5). That way, the predicted quantile could be compared with the empirical quantile for each age class, and it was also possible to derive observed and predicted values for all age classes and quantiles (figs. 4.7 to 4.10). It was also possible to acquire one single calibration plot for all age classes and quantiles. For model validation, the quantiles 0.25, 0.5, and 0.75 were selected and were coded in graphs as  $\bullet$  0.25,  $\blacktriangle$  0.5 and  $\blacksquare$  0.75 (figs. 4.7 to 4.10). For the calibration plot, predictions from all three quantile regressions (quantiles 0.25, 0.5, and 0.75) were compared to observed data and plotted within a single calibration plot. Linear regression was performed using the same calibration data to get one overall regression result for all quantiles.

### Further nomograms and subgroup analyses

In addition to the described reference values, additional nomograms were conducted. On the one hand, this was done to ascertain comparability with past studies that excluded smokers and obese participants [50]. On the other hand, subgroup analyses were conducted using only participants from either Frankfurt and Rüdeshheim or participants from Munich. All nomograms can be accessed via the online appendix at <https://github.com/rappdaniel/vo2peak>.

### Measures of exertion in random sample

As described in sections 3.2 and 3.3, the measures of exertion were acquired for a random sample of 252 subjects in the present study. It was not possible to exclude cases with poor effort as the measures of exertion were not recorded in the entire main study dataset. This

is a critical issue because participants who did not exert maximal effort are not valid and should be excluded from the analysis. If, for example, a participant terminated CPET immediately after the start because of anxiety caused by wearing the gas exchange mask, the measured maximal oxygen uptake would be close to the resting oxygen consumption. Such cases would distort the reference values. Therefore, this study aimed to estimate the proportion of participants who did not continue CPET until the maximal volitional effort. On the other hand, the goal was to estimate how the reference values were affected by keeping such participants in the analysis. These analyses were performed in the random sample in which the measures of maximal effort were gathered retrospectively. First, the proportion and 95% confidence intervals of participants who did not reach the end criteria (section 3.2) was calculated. Secondly, a visual approach was used: Median regression lines were plotted using i) all cases of the random sample and ii) only cases which reached exertion in CPET. Peak oxygen uptake was used as the dependent and age in years as the independent variable, respectively. Additionally, an analysis of covariance was applied using i) age in years, ii) exertion (yes/no) as well as iii) an interaction term of both as independent variables in median regressions. This was done to derive P values for the variable “exertion” and to see if the regression was significantly altered by keeping subjects with poor effort in the data.

### **Multivariable regression modelling**

The sample characteristics of the present dataset were considered for multivariable quantile regression modelling. In contrast to the calculation of nomograms, the quantiles 0.25, 0.50, and 0.75 were used. A correlation matrix of all quantitative variables using Spearman’s correlation coefficient was calculated before the variables were included in the regression model to avoid collinearity. If there was a high correlation between two variables, one was considered to be eliminated heuristically. Histograms, as well as measures of skewness, were used to check for normality of the distribution of quantitative variables. If a quantitative variable appeared to follow a non-normal distribution, the variable was recoded into a binary variable using dummy coding (0 = no, 1 = yes). All variables were added to an AIC-based stepwise variable selection using backward and forward selection. Furthermore, a pseudo R squared according to Hao et al. (2007) [22, p. 52] was calculated for all quantile regression models.

### **Quantile calculator**

The goal of the present analysis was to present reference values for peak oxygen uptake that can be interpreted by physicians and CPET participants. An interactive web application was created (section 7.1) to facilitate doctor-patient communication and increase the value of the presented reference values in clinical practice ([www.uks.eu/vo2peak](http://www.uks.eu/vo2peak)). This web application followed four goals:

1. Plotting the individual results of a CPET participant that are obtained in the course of a preventive health screening on a nomogram produced by the present study.
2. Calculating a sex-specific and age-specific percentile for the individual's result. It was aimed to present a percentile which is exact within one percent.
3. Optional plotting of 95% confidence intervals to visualise the uncertainty of the estimation.
4. Interactive subgroup analyses by excluding smokers and obese participants from the reference values.

The methods that were used for this web application were slightly different from the rest of present analysis (chapter 3.3). The overall sample size in this analysis was 10,090 instead of 9,354 because participants who provided no information on the study centre (n=673) and participants who were not 25 to 70 years old (n=63) were not excluded. The material and methods for this analysis are also described in Rapp et al. (2018) [55]. The web application was created using the R software package `shiny` [10]. A screenshot is displayed in fig. 4.12.

### Parallelisation

Some of the statistical procedures of the present study were computationally intensive. Especially the calculation of 95% confidence intervals using 10,000 bootstrap samples was time-consuming. Therefore, parallel computing was applied to decrease the computation time using the R software package `parallel` [54], and mostly 4 CPUs. The calculation of 95% confidence intervals in the online web application required the greatest amount of time. The hardware used for these calculations was a DELL-Server (DELL PowerEdge R720) with two 8-core-processors (CPUs) and 256GB RAM memory using a Debian-based Linux operating system. Despite using 12 of the 16 CPUs in this case, the overall computation process lasted more than 150 hours.

### Missing data

Participants with missing values in one of the primary variables of interest (peak oxygen uptake, age and sex) were excluded from the data (fig. 3.2). In multiple quantile regression, casewise deletion was used. Casewise deletion means that cases with a missing value in the dependent variable or one of the independent variables are excluded from the regression model in which they appear. In the R-code, casewise deletion is specified by the term `na.action = "na.omit"`.

## 3.7 Reporting

The results of the present study were reported according to the RECORD\* guidelines [69], which are an extension of the STROBE† reporting guidelines. A complete RECORD checklist can be accessed via the online appendix [49].

---

\*Reporting of studies conducted using observational routinely-collected data

†Strengthening the reporting of observational studies in epidemiology



## 4 Results

### 4.1 Descriptive statistics

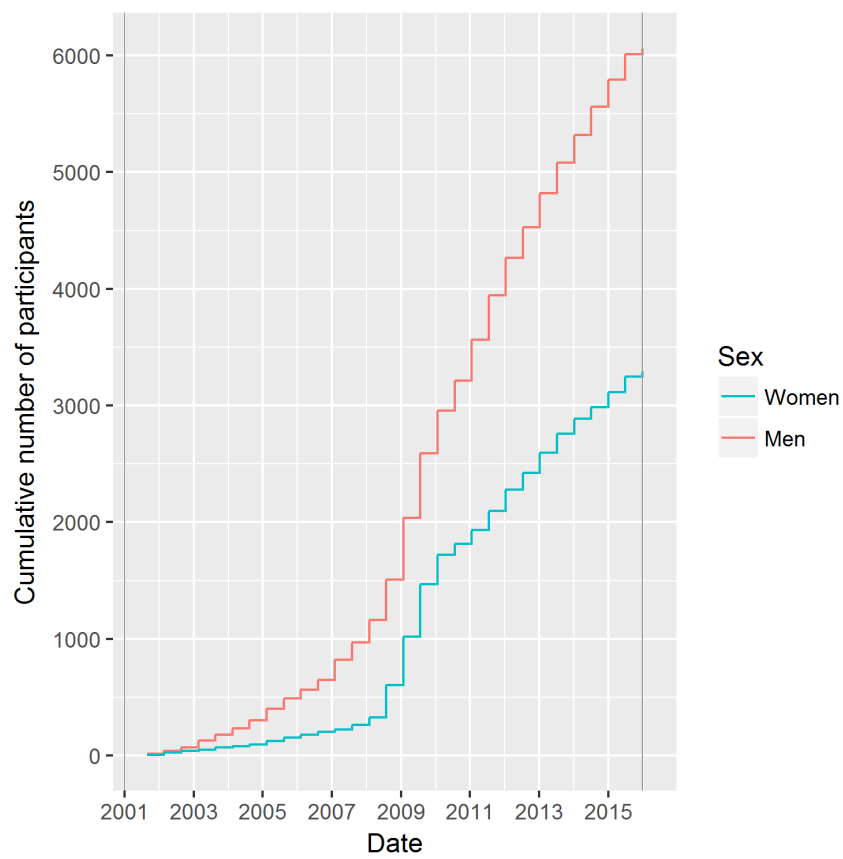
The time period of data collection ranged from 18 July 2001 to 20 November 2015. Rüdeshheim was the first study centre to record the data in 2001, followed by Frankfurt in 2006 and Munich in 2008. The cumulative number of participants by examination date is plotted in fig. 4.1.

Overall, 9,354 participants (6,063 men, 3,291 women) were included in the analysis. 7,516 subjects (4,710 men, 2,806 women) were recorded in Rüdeshheim or Frankfurt and 1,838 (1,353 men, 485 women) subjects were recorded in Munich (table 4.1).

The overall median age was 45 [41; 50] years for men and also for women (supplementary table 1 [49]). Despite a high number of observations overall, the marginal age groups were sparse. For example, only one women in the age class [30; 35) years was recorded in Munich (table 4.1). The distributions of the participants' ages are displayed in figs. A.1 and A.2.

Median absolute  $\dot{V}O_{2\text{peak}}$  was 3.0 LO<sub>2</sub>/min [2.6; 3.4] for men and 1.9 LO<sub>2</sub>/min [1.7; 2.2] for women (supplementary table 1 [49]). Relative  $\dot{V}O_{2\text{peak}}$  was 35.3 mL O<sub>2</sub>/min/kg [30.3; 40.5] for men and 28.7 mL O<sub>2</sub>/min/kg [24.4; 33.2] for women. The differences between men and women in both, absolute and relative  $\dot{V}O_{2\text{peak}}$  were statistically significant. However, there was no statistically significant difference of peak oxygen uptake between study centres except for absolute  $\dot{V}O_{2\text{peak}}$  in men (table 4.1).

Peak oxygen uptake was lower in older participants. In men, median relative  $\dot{V}O_{2\text{peak}}$  was 37.2 mL O<sub>2</sub>/min/kg in the age group [25; 30) and 29.3 mL O<sub>2</sub>/min/kg in the age group [65; 70). A similar decrease was observed in women. Quantiles of relative and absolute  $\dot{V}O_{2\text{peak}}$  by sex and age group are displayed in tables 4.2 and 4.3, respectively. The distribution of absolute and relative  $\dot{V}O_{2\text{peak}}$  is visualised in figure 4.2. Furthermore, all quantitative characteristics were plotted as histograms and scatter plot matrices in figs. A.1 to A.4.



**Figure 4.1:** Cumulative number of participants from Rüdeseheim, Frankfurt and Munich by date of examination. The overall inquiry period ranged from 18 July 2001 to 20 November 2015.

**Table 4.1:** Descriptive statistics by sex and study center.

	Men			Women		
	Frankfurt/Rüdesheim N=4710	Munich N=1353	P value	Frankfurt/Rüdesheim N=2806	Munich N=485	P value
<b><math>\dot{V}O_{2peak}</math></b>						
Relative [mLO <sub>2</sub> /min/kg]	35.2 [30.3;40.5]	35.5 [30.3;40.6]	0.846	28.6 [24.3;33.1]	28.9 [24.7;33.8]	0.168
Absolute [LO <sub>2</sub> /min]	3.02 [2.64;3.41]	2.95 [2.57;3.36]	<0.001	1.95 [1.68;2.20]	1.90 [1.66;2.19]	0.186
<b>Anthropometric</b>						
Age [years]	45.0 [40.0;50.0]	46.0 [41.0;51.0]	<0.001	45.0 [41.0;50.0]	45.0 [41.0;49.0]	0.863
Age class [years]						
[25,30)	22 (0.47%)	3 (0.22%)		9 (0.32%)	4 (0.82%)	
[30,35)	103 (2.19%)	21 (1.55%)		51 (1.82%)	1 (0.21%)	
[35,40)	412 (8.75%)	132 (9.76%)		106 (3.78%)	25 (5.15%)	
[40,45)	1816 (38.6%)	413 (30.5%)		1208 (43.1%)	197 (40.6%)	
[45,50)	1056 (22.4%)	374 (27.6%)		655 (23.3%)	137 (28.2%)	
[50,55)	721 (15.3%)	247 (18.3%)		439 (15.6%)	80 (16.5%)	
[55,60)	401 (8.51%)	117 (8.65%)		232 (8.27%)	31 (6.39%)	
[60,65)	143 (3.04%)	38 (2.81%)		81 (2.89%)	8 (1.65%)	
[65,69]	36 (0.76%)	8 (0.59%)		25 (0.89%)	2 (0.41%)	
Weight [kg]	85.0 [78.0;94.0]	83.0 [76.0;91.0]	<0.001	67.0 [60.0;76.0]	64.0 [59.0;73.0]	0.001
Height [cm]	181 [177;186]	181 [177;185]	0.414	167 [163;172]	167 [163;172]	0.806
BMI [kg/m <sup>2</sup> ]	25.7 [23.8;28.2]	25.2 [23.5;27.4]	<0.001	23.8 [21.4;26.9]	23.1 [21.1;25.8]	0.001
Body fat Caliper [%]	23.0 [19.0;27.0]	22.6 [19.0;26.7]	0.173	31.0 [26.0;36.5]	28.5 [24.6;33.7]	<0.001
Overweight			<0.001			0.001
no	1865 (39.6%)	625 (46.2%)		1729 (61.7%)	337 (69.5%)	
yes	2842 (60.4%)	728 (53.8%)		1075 (38.3%)	148 (30.5%)	
Obesity			<0.001			0.014
no	4078 (86.6%)	1228 (90.8%)		2474 (88.2%)	447 (92.2%)	
yes	629 (13.4%)	125 (9.24%)		330 (11.8%)	38 (7.84%)	

*continued on next page*

Table 4.1 – continued from previous page

	Men			Women		
	Frankfurt/Rüdesheim N=4710	Munich N=1353	P value	Frankfurt/Rüdesheim N=2806	Munich N=485	P value
<b>Blood pressure</b>						
Systolic [mmHg]	128 [120;138]	124 [118;134]	<0.001	120 [110;130]	116 [106;128]	0.006
Diastolic [mmHg]	82.0 [80.0;90.0]	80.0 [78.0;88.0]	<0.001	80.0 [70.0;82.0]	78.0 [70.0;82.0]	0.051
Hypertension			0.007			0.014
no	3193 (67.9%)	971 (71.8%)		2303 (82.1%)	421 (86.8%)	
yes	1508 (32.1%)	381 (28.2%)		501 (17.9%)	64 (13.2%)	
<b>Glucose metabolism</b>						
Blood glucose [mg/dL]	96.0 [90.0;102]	93.0 [88.2;99.0]	<0.001	91.0 [86.0;98.0]	89.5 [83.0;94.0]	<0.001
HbA1 <sub>c</sub> [%]	5.40 [5.10;5.60]	5.40 [5.20;5.60]	<0.001	5.30 [5.10;5.60]	5.40 [5.20;5.50]	0.054
Diabetes mellitus			0.061			0.636
no	4524 (96.9%)	1305 (97.9%)		2738 (98.3%)	478 (98.8%)	
yes	146 (3.13%)	28 (2.10%)		46 (1.65%)	6 (1.24%)	
<b>Lipid metabolism</b>						
Total cholesterol [mg/dL]	216 [191;242]	208 [185;232]	<0.001	209 [186;235]	200 [180;222]	<0.001
HDL cholesterol [mg/dL]	52.0 [45.0;60.0]	54.0 [47.0;63.0]	<0.001	66.0 [56.0;75.0]	72.0 [61.0;84.0]	<0.001
LDL cholesterol [mg/dL]	135 [114;158]	128 [107;151]	<0.001	122 [102;145]	110 [94.0;131]	<0.001
Triglycerides [mg/dL]	117 [85.0;164]	104 [75.0;149]	<0.001	87.0 [68.0;119]	75.0 [58.0;100]	<0.001

continued on next page

Table 4.1 – continued from previous page

	Men			Women		
	Frankfurt/Rüdesheim N=4710	Munich N=1353	P value	Frankfurt/Rüdesheim N=2806	Munich N=485	P value
<b>Smoking status</b>						
Smoker			<0.001			<0.001
no	3990 (85.1%)	1226 (91.4%)		2352 (84.3%)	434 (90.8%)	
yes	696 (14.9%)	116 (8.64%)		439 (15.7%)	44 (9.21%)	
Ex-smoker			<0.001			<0.001
no	3391 (72.4%)	1102 (82.1%)		2045 (73.3%)	395 (82.6%)	
yes	1295 (27.6%)	240 (17.9%)		746 (26.7%)	83 (17.4%)	

**Note:** Overweight: BMI  $\geq 25\text{kg/m}^2$ , obesity: BMI  $\geq 30\text{kg/m}^2$ , hypertension: systolic blood pressure  $\geq 140$  mmHg or diastolic blood pressure  $\geq 90$  mmHg

Quantitative characteristics are displayed as median [1st quartile; 3rd quartile], qualitative characteristics as n (%).

No P values were calculated for the characteristic “age class” because this resulted in 2-by-9 contingency tables with expected cell counts with  $n < 5$  under the null hypothesis were present.

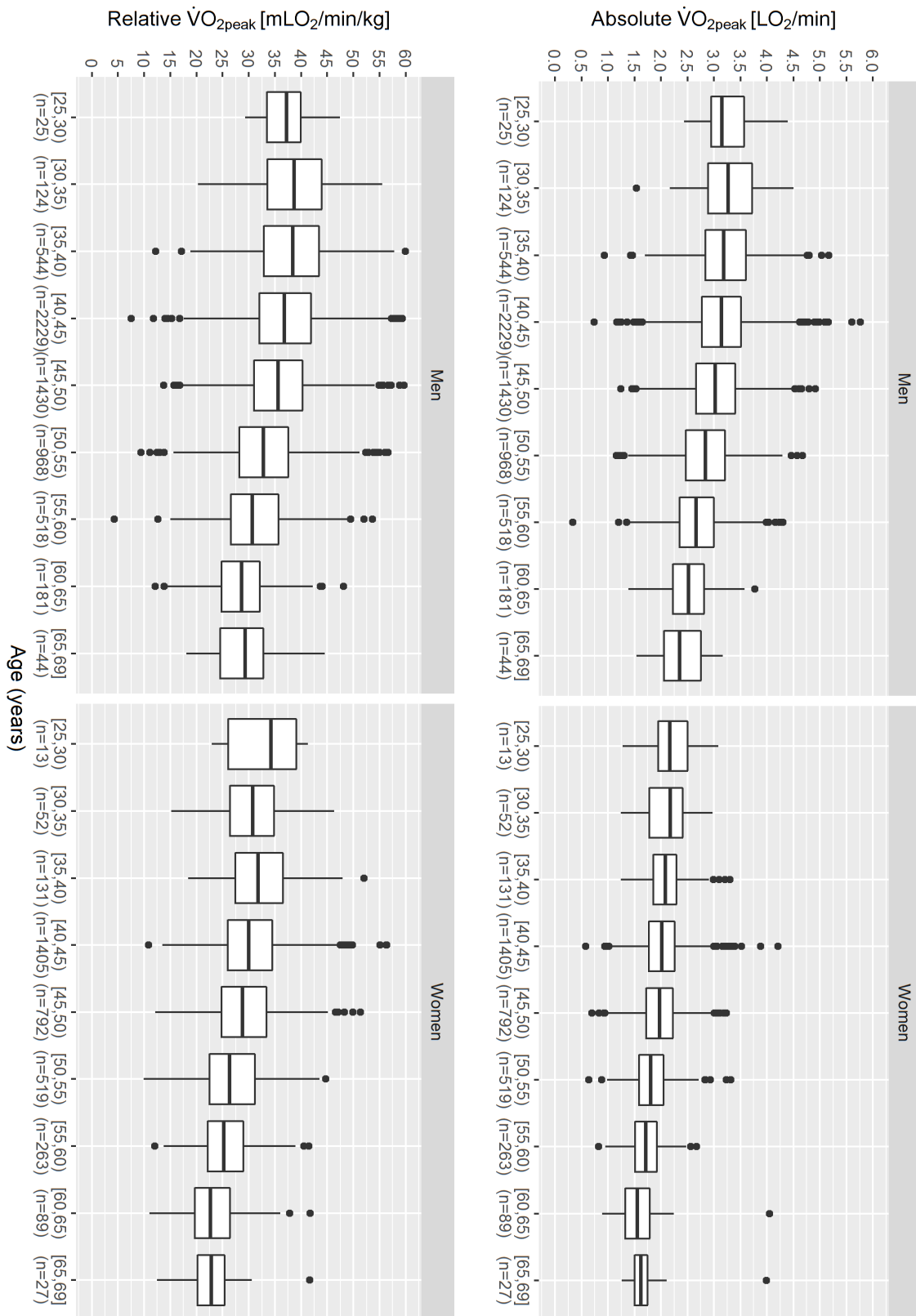


Figure 4.2: Box plots of absolute and relative  $\dot{V}O_{2peak}$  by age group.

**Table 4.2:** Quantiles of relative  $\dot{V}O_{2\text{peak}}$  by sex and age group.

Age class	Quantiles													n
	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	
<b>Female</b>														
[25,30)	23.1	23.5	25.1	26.1	27.9	32.5	34.2	35.2	37.4	39.1	39.8	40.6	40.9	13
[30,35)	20.3	24.2	25.9	26.4	27.2	30.0	30.8	32.5	34.0	34.8	36.4	40.1	43.5	52
[35,40)	22.9	25.1	26.7	27.4	28.6	30.1	31.8	33.8	35.0	36.5	37.8	40.5	44.1	131
[40,45)	20.4	22.4	24.9	26.0	27.0	28.6	30.0	31.7	33.3	34.5	35.4	38.4	40.4	1405
[45,50)	19.4	21.6	23.7	24.8	25.6	27.1	28.8	30.6	32.5	33.4	34.6	37.3	39.7	792
[50,55)	17.2	19.0	21.5	22.5	23.5	25.1	26.3	28.0	30.3	31.2	32.2	34.7	36.9	519
[55,60)	16.0	18.9	21.7	22.2	22.8	23.6	25.2	26.8	28.2	28.9	29.4	32.0	33.8	263
[60,65)	15.4	16.9	18.8	19.7	19.9	21.6	22.7	24.2	25.1	26.4	27.3	30.1	32.3	89
[65,69]	15.6	18.8	19.7	20.2	20.6	21.2	22.8	23.6	24.9	25.4	26.2	27.0	29.7	27
<b>Male</b>														
[25,30)	29.7	30.9	33.1	33.5	33.7	35.9	37.2	37.8	39.3	39.9	40.3	42.1	44.0	25
[30,35)	26.8	28.5	31.7	33.6	34.1	36.5	38.7	40.4	42.8	43.9	45.3	48.0	50.9	124
[35,40)	26.2	29.5	31.8	32.9	34.2	36.1	38.5	40.3	42.5	43.4	44.8	47.8	50.7	544
[40,45)	25.4	28.0	30.9	32.0	33.1	34.9	36.8	38.7	40.7	41.9	43.2	46.5	49.5	2229
[45,50)	25.1	27.1	30.0	31.0	32.1	33.9	35.6	37.5	39.3	40.3	41.6	44.6	47.5	1430
[50,55)	21.7	24.0	27.3	28.2	29.2	30.9	32.8	34.9	36.7	37.6	39.2	42.7	46.2	968
[55,60)	20.8	23.2	25.4	26.6	27.3	29.1	30.7	32.7	34.7	35.7	36.7	39.7	42.8	518
[60,65)	19.5	21.7	24.1	24.8	25.7	27.7	28.6	29.8	31.7	32.1	33.3	37.7	40.7	181
[65,69]	19.7	22.0	23.7	24.5	26.1	27.9	29.3	30.1	30.6	32.8	34.8	36.4	38.0	44

**Note:** Relative  $\dot{V}O_{2\text{peak}}$  was measured in mL $O_2$ /min/kg, age in years, participants from Frankfurt, Rüdeshheim and Munich included.

**Table 4.3:** Quantiles of absolute  $\dot{V}O_{2\text{peak}}$  by sex and age group.

Age class	Quantiles													n
	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	
<b>Female</b>														
[25,30)	1.47	1.66	1.91	1.95	1.97	1.99	2.17	2.32	2.41	2.50	2.64	2.86	2.97	13
[30,35)	1.49	1.61	1.73	1.78	1.91	2.08	2.17	2.25	2.29	2.41	2.45	2.62	2.70	52
[35,40)	1.55	1.64	1.80	1.86	1.90	1.99	2.08	2.13	2.23	2.29	2.37	2.60	2.84	131
[40,45)	1.43	1.55	1.71	1.77	1.83	1.93	2.02	2.11	2.20	2.26	2.32	2.50	2.70	1405
[45,50)	1.37	1.50	1.66	1.72	1.78	1.88	1.97	2.06	2.17	2.23	2.29	2.47	2.65	792
[50,55)	1.25	1.36	1.54	1.59	1.63	1.72	1.81	1.90	2.00	2.05	2.12	2.30	2.43	519
[55,60)	1.16	1.34	1.46	1.51	1.55	1.63	1.71	1.80	1.86	1.92	2.02	2.10	2.25	263
[60,65)	1.10	1.19	1.29	1.33	1.38	1.47	1.55	1.65	1.74	1.79	1.86	2.00	2.07	89
[65,69]	1.37	1.43	1.48	1.50	1.52	1.57	1.62	1.68	1.73	1.75	1.77	2.03	2.11	27
<b>Male</b>														
[25,30)	2.62	2.75	2.80	2.95	3.09	3.14	3.15	3.27	3.37	3.57	3.65	3.89	4.06	25
[30,35)	2.31	2.49	2.72	2.89	3.00	3.12	3.27	3.41	3.60	3.72	3.80	4.15	4.25	124
[35,40)	2.32	2.54	2.77	2.84	2.90	3.05	3.18	3.35	3.50	3.60	3.69	3.93	4.16	544
[40,45)	2.22	2.41	2.66	2.77	2.85	2.99	3.14	3.27	3.43	3.51	3.60	3.87	4.09	2229
[45,50)	2.17	2.37	2.59	2.66	2.74	2.89	3.02	3.17	3.31	3.40	3.49	3.75	3.95	1430
[50,55)	1.96	2.13	2.38	2.47	2.55	2.70	2.84	3.00	3.11	3.21	3.31	3.53	3.73	968
[55,60)	1.80	2.05	2.22	2.35	2.40	2.54	2.66	2.80	2.90	3.00	3.13	3.37	3.53	518
[60,65)	1.77	1.95	2.11	2.23	2.28	2.42	2.52	2.64	2.76	2.81	2.86	3.05	3.30	181
[65,69]	1.61	1.81	2.00	2.06	2.12	2.23	2.35	2.49	2.64	2.75	2.87	2.93	3.06	44

**Note:** Absolute  $\dot{V}O_{2\text{peak}}$  was measured in LO $_2$ /min, age in years, participants from Frankfurt, Rüdeshheim and Munich included.

## 4.2 Representativity

To get information on the representativity for the German population, some characteristics were compared between the underlying dataset and results from 'Studie zur Gesundheit Erwachsener in Deutschland' (DEGS1) [32]. Five binary characteristics (smoker, ex-smoker, overweight, obesity, hypertension) were measured equally in both sources and were therefore eligible for comparison (table 4.4). Direct age standardisation was applied to achieve comparable results (chapter 3.4).

Especially in women, the proportions of the underlying data differed from DEGS1 results. The only non-significant difference was in ex-smoking. The differences between the present study and DEGS1 were smaller in men. The proportions of smokers, overweight and obese men were significantly lower in the present study. Almost all of these compared risk factors were more prevalent in DEGS1 compared to the present study. Only hypertension in men and ex-smoking in women were less prevalent in DEGS1 (table 4.4). Large differences were observed for obesity in women as well as for smoking status in men. 23.9% of the women in DEGS1 and only 12.1% in the present study were obese. In DEGS1, 26.1% of men were smokers whereas this proportion was only 14.7% in the present study.



**Table 4.4:** Comparison of study population to results of DEGS1.

	Male		Female	
	PF Studie	DEGS1	PF Studie	DEGS1
Smoker	<b>14.7 (12.7 to 16.9)</b>	<b>26.1 (24.0 to 28.2)</b>	<b>14.7 (12.1 to 17.6)</b>	<b>21.4 (19.7 to 23.1)</b>
Ex-smoker	29.9 (27.3 to 32.5)	33.7 (31.9 to 35.5)	26.9 (23.6 to 30.4)	22.8 (21.4 to 24.2)
Overweight	<b>61.0 (58.3 to 63.6)</b>	<b>67.1 (65 to 69.2)</b>	<b>38.4 (35.0 to 42.1)</b>	<b>53.0 (50.8 to 55.1)</b>
Obesity	<b>13.9 (12.1 to 15.9)</b>	<b>23.3 (21.2 to 25.4)</b>	<b>12.1 (10.0 to 14.3)</b>	<b>23.9 (22.0 to 25.9)</b>
Hypertension	36.4 (33.8 to 39.2)	33.3 (31.1 to 35.6)	<b>20.7 (18.2 to 23.5)</b>	<b>29.9 (28.1 to 31.9)</b>

**Note:** Frequencies are displayed as % [95% confidence interval]. For this analysis, participants from all study centres (Frankfurt/Rüdesheim and Munich) were included (N = 9,354).

Overweight: BMI  $\geq 25$  kg/m<sup>2</sup>, obesity: BMI  $\geq 30$  kg/m<sup>2</sup>, hypertension: systolic blood pressure  $\geq 140$  mmHg or diastolic blood pressure  $\geq 90$  mmHg.

DEGS1 data extraction for smoking and ex-smoking [38], overweight and obesity [43] and hypertension [48].

Direct age standardisation was performed for results from the present study with the R package `epitools` [2] using the German age structure of 2011 as standard population [63]. Proportions in which 95% confidence intervals do not overlap are printed bold.

### 4.3 Nomograms and regression model validation

Quantile regressions were fitted using peak oxygen uptake as the dependent variable and age in years of the participant as the independent variable to plot age-dependent nomograms of peak oxygen uptake. Three different approaches of how to model age as the independent variable were described in section 3.6. Age could be modelled as i) a linear predictor, ii) a polynomial predictor or iii) using b-spline smoothing. All approaches were compared using Akaike Information Criterion (AIC) as well as in apparent validation of the models (figs. A.5 and A.6 and tables A.4 and A.5). Calibration data included three quantile regressions for the quantiles 0.25, 0.5, and 0.75. Box plots of observed data and predicted quantiles per age class were plotted (fig. A.5) and in calibration plots, the results of all three regressions were compared with observed quantiles in each age group (fig. A.6). The results from linear regressions that were fitted to calibration data are displayed in table A.4.

Quantile regressions with b-spline smoothing showed the most accurate results in apparent validation. Age as a linear predictor yielded the worst results. Age as a second-degree polynomial predictor fell in between these approaches. As an example, the results from linear regressions of calibration data should be outlined for men and relative  $\dot{V}O_{2\text{peak}}$ : In apparent validation, intercept, slope and R squared were 6.21, 0.80 and 0.89 for linear regression, 5.94, 0.83 and 0.92 for polynomial regression as well as 1.16, 0.97 and 0.98 for regression with b-splines, respectively (table A.4). AICs in relative  $\dot{V}O_{2\text{peak}}$  of men for median regressions were 8,175, 8,152, and 8,145 for linear regression, polynomial regression and regressions with b-splines, respectively (A.5). Overall, spline regressions performed best. Nevertheless, spline regression models often yield complex regression equations and cannot be formulated like univariable, multivariable or polynomial regression models [42]. As the final models should be outlined in a table, the more feasible approach was selected over the most accurate and final nomograms were conducted using age as a predictor with a second-degree polynomial. The final models were defined as:

$$\dot{V}O_{2\text{peak } i, \tau} = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau) * \text{age}_i + \hat{\beta}_2(\tau) * \text{age}_i^2 + e_{i, \tau} \quad (4.1)$$

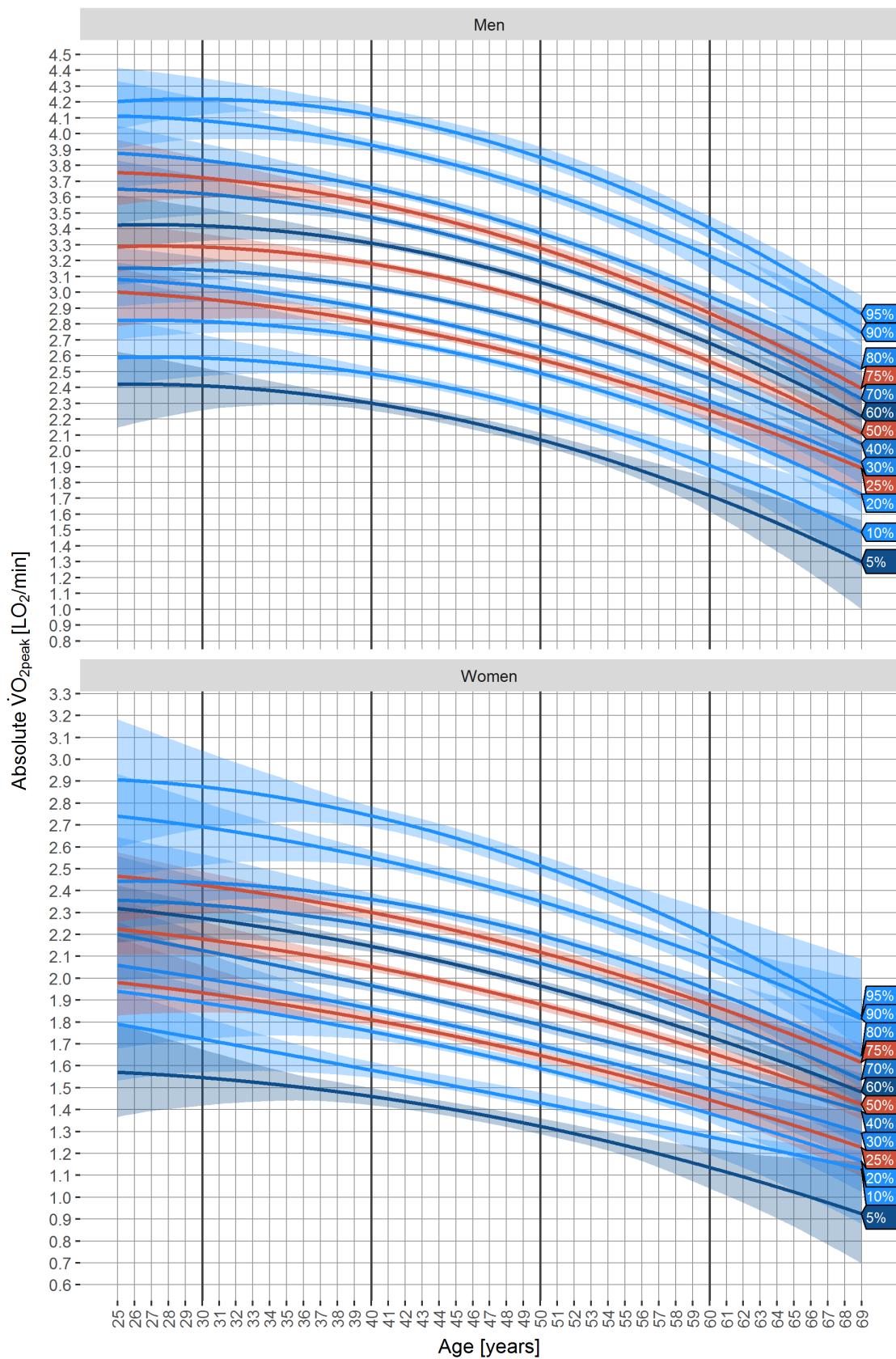
where  $\tau$  denotes the  $\tau$ -quantile,  $i = 1, \dots, n$  the  $i$ -th participant, age indicates the age in years of the participant  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\beta}_2$  the regression coefficients to be estimated and  $e$  the error term.

The validation of polynomial quantile regressions is displayed in figs. 4.7 to 4.8. Figure 4.7 shows predicted quantiles based on Frankfurt/Rüdesheim in comparison to observed quantiles from Frankfurt/Rüdesheim. The 95% confidence intervals of the predicted values (red shades) include a substantial proportion of the observed quantiles. This means that there was no statistically significant difference between observed and predicted values.

The predicted values were particularly accurate in central age classes and less accurate in marginal age classes. Significant differences between predicted and observed quantiles appeared especially in the age classes 25 to 29 years and 65 to 69 years. These results were also observed in the calibration plot (fig. 4.8) with results of R squared close to 1, intercept close to 0 and slope close to 1. As to be expected, the results of external validation showed less accurate results. However, especially the predictions in the central age classes ranging from 30 to 59 years were still accurate. In the marginal age classes, it has to be noted that the numbers of participants were low, leading to distorted empirical quantiles (fig. 4.9). This could also be observed in the calibration plots where the linear regressions showed some deviation from the angle bisector (fig. 4.10).

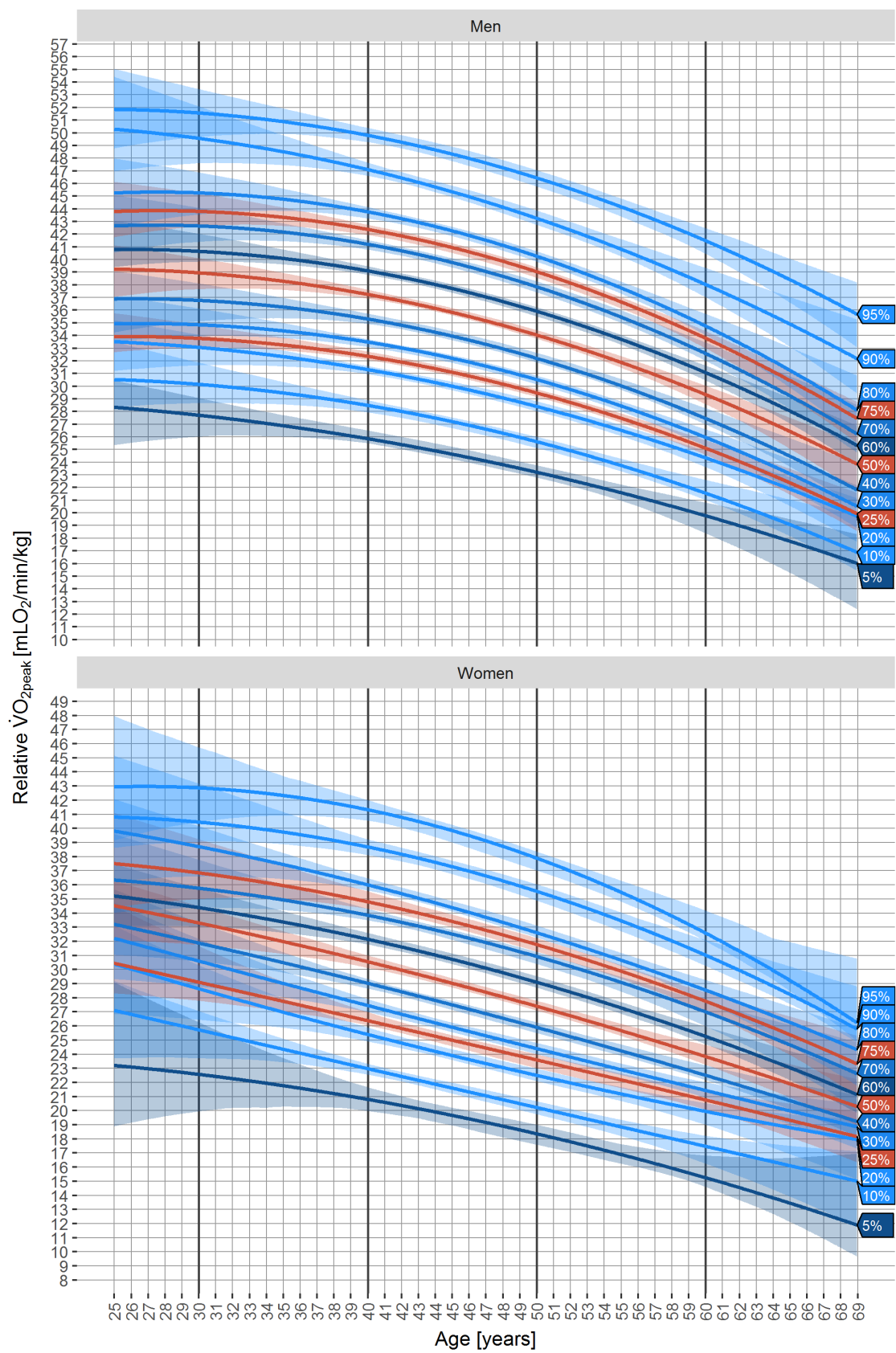
Regression coefficients of the final polynomial quantile regression models are displayed in tables 4.5 and 4.6 and plotted in figs. 4.5 and 4.6.

In addition to the presented results, subgroup analyses were performed. Reference values and nomograms were conducted using the entire dataset from Rüdeshheim, Frankfurt and Munich but also using only non-smokers and non-obese participants (supplementary tables and figures, [49]).



**Figure 4.3:** Nomogram of absolute  $\dot{V}O_{2peak}$  (Rüdesheim/Frankfurt)

**Note:** Participants from Rüdesheim and Frankfurt were included (2,806 women , 4,710 men).  
Shades are 95% confidence intervals.



**Figure 4.4:** Nomogram of relative  $\dot{V}O_{2peak}$  (Rüdesheim/Frankfurt)

**Note:** Participants from Rüdesheim and Frankfurt were included (2,806 women , 4,710 men).  
Shades are 95% confidence intervals.

**Table 4.5:** Regression coefficients and 95% confidence intervals of quantile regressions plotted in nomograms (absolute  $\dot{V}O_{2\text{peak}}$ ).

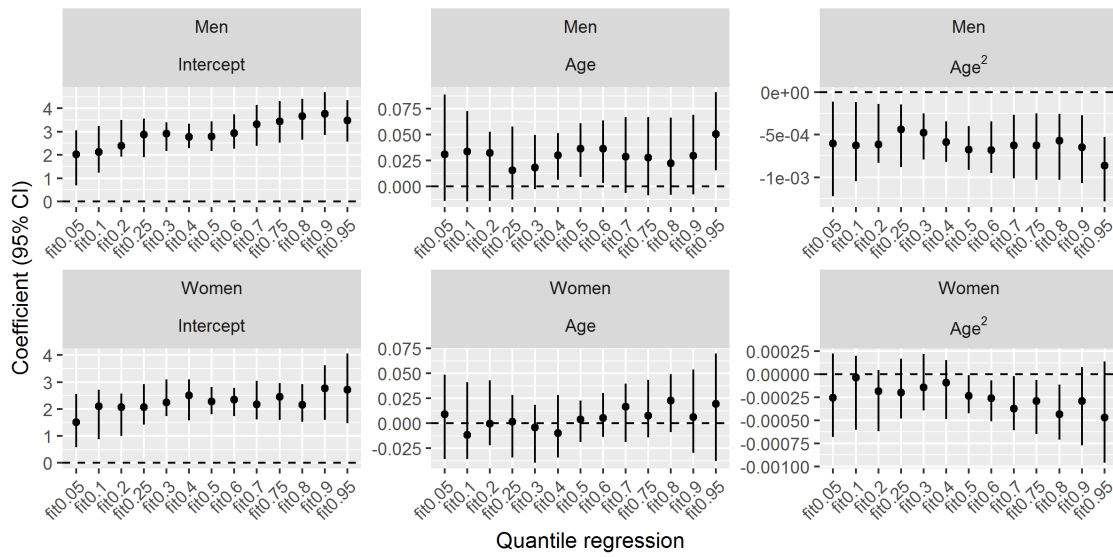
Quantile regression	Intercept			Age			Age <sup>2</sup>		
	Coefficient	95% CI		Coefficient	95% CI		Coefficient	95% CI	
		Lower bound	Upper bound		Lower bound	Upper bound		Lower bound	Upper bound
<b>Women</b>									
fit0.05	1.49913043	0.57906056	2.55129982	0.00916667	-0.03576120	0.04884738	-0.00025362	-0.00068089	0.00022411
fit0.1	2.10653061	0.87801975	2.70606696	-0.01180272	-0.03565934	0.04133534	-0.00003401	-0.00060396	0.00020259
fit0.2	2.06037037	0.99757677	2.58000000	-0.00018519	-0.02199574	0.04324667	-0.00018519	-0.00061728	0.00004548
fit0.25	2.06284600	1.41904762	2.92625754	0.00168129	-0.03425926	0.02863636	-0.00019981	-0.00047890	0.00017037
fit0.3	2.24454545	1.73953912	3.10143878	-0.00380808	-0.03949346	0.01845158	-0.00014478	-0.00039112	0.00022059
fit0.4	2.49637500	1.58473106	3.09074986	-0.00950000	-0.03444444	0.02871359	-0.00009375	-0.00048486	0.00015670
fit0.5	2.27391994	1.80357829	2.81044196	0.00387164	-0.01857143	0.02294894	-0.00023464	-0.00042357	-0.00000700
fit0.6	2.34365265	1.72597017	2.78857293	0.00549134	-0.01330923	0.03057967	-0.00026099	-0.00051054	-0.00006443
fit0.7	2.17450000	1.61368034	3.04000000	0.01662500	-0.01879120	0.03977625	-0.00037500	-0.00061006	-0.00001852
fit0.75	2.45054545	1.58603838	2.96052881	0.00778961	-0.01390476	0.04346154	-0.00028831	-0.00064935	-0.00006250
fit0.8	2.14655462	1.52331800	2.91912279	0.02270308	-0.00892053	0.04914586	-0.00043417	-0.00070996	-0.00011311
fit0.9	2.76728938	1.59563467	3.61307576	0.00614286	-0.02926714	0.05407547	-0.00028938	-0.00076925	0.00007775
fit0.95	2.70967546	1.46509311	4.05118261	0.01959770	-0.03775323	0.06983979	-0.00046991	-0.00096202	0.00013889
<b>Men</b>									
fit0.05	2.01681818	0.69031328	3.04755405	0.03117045	-0.01423629	0.08840770	-0.00060227	-0.00121296	-0.00011252
fit0.1	2.12961538	1.22997586	3.24016740	0.03401282	-0.01423148	0.07290991	-0.00062821	-0.00104848	-0.00011396
fit0.2	2.38682051	1.88285642	3.51403260	0.03267873	-0.01388917	0.05393145	-0.00061237	-0.00083575	-0.00013689
fit0.25	2.88086957	1.89875652	3.55889773	0.01565217	-0.01273666	0.05764334	-0.00043478	-0.00087558	-0.00013974
fit0.3	2.91333333	2.16329239	3.39361014	0.01857143	-0.00232158	0.05000046	-0.00047619	-0.00080000	-0.00025000
fit0.4	2.76311037	2.29830183	3.34193204	0.03023226	0.00621205	0.05076022	-0.00058900	-0.00080953	-0.00034446
fit0.5	2.79016667	2.19178512	3.41613125	0.03676786	0.01027467	0.06064167	-0.00067560	-0.00091471	-0.00040070
fit0.6	2.93021008	2.27226144	3.74169030	0.03679622	0.00338730	0.06380689	-0.00068277	-0.00095586	-0.00034259
fit0.7	3.31865801	2.39498904	4.13723057	0.02893939	-0.00540859	0.06700078	-0.00062771	-0.00101121	-0.00027174
fit0.75	3.44690821	2.52533731	4.33076923	0.02795743	-0.00880982	0.06764484	-0.00062651	-0.00103451	-0.00025132
fit0.8	3.66909091	2.65130000	4.39115654	0.02250000	-0.00773616	0.06597359	-0.00056818	-0.00102779	-0.00025298
fit0.9	3.77000000	2.82190745	4.69615087	0.02984848	-0.00780413	0.07077050	-0.00064738	-0.00107629	-0.00026695
fit0.95	3.47483516	2.59020000	4.36079514	0.05066904	0.01467532	0.08978920	-0.00086296	-0.00127605	-0.00051435

**Note:** Absolute  $\dot{V}O_{2\text{peak}}$  was measured in  $\text{LO}_2/\text{min}$ .

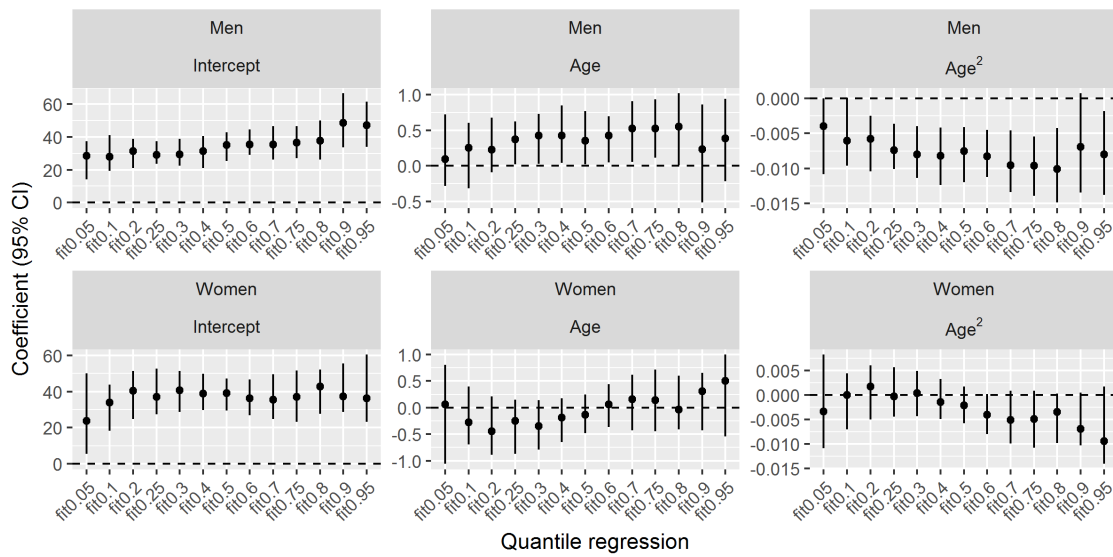
**Table 4.6:** Regression coefficients and 95% confidence intervals of quantile regressions plotted in nomograms (relative  $\dot{V}O_{2peak}$ ).

Quantile regression	Intercept			Age			Age <sup>2</sup>		
	Coefficient	95% CI		Coefficient	95% CI		Coefficient	95% CI	
		Lower bound	Upper bound		Lower bound	Upper bound		Lower bound	Upper bound
<b>Women</b>									
fit0.05	23.84835165	5.59087363	50.10395565	0.05785714	-1.05766005	0.80808081	-0.00335165	-0.01087569	0.00828160
fit0.1	33.97500000	18.19006800	43.70057692	-0.27500000	-0.69098656	0.39722149	-0.00000000	-0.00702743	0.00443350
fit0.2	40.43443223	24.81967857	51.45452772	-0.44472527	-0.88907457	0.20955888	0.00172161	-0.00500019	0.00608467
fit0.25	36.92186235	27.44654913	52.76745790	-0.25205321	-0.87316774	0.14559597	-0.00028918	-0.00439815	0.00571448
fit0.3	40.60000000	28.60000000	51.31336184	-0.34722222	-0.79009010	0.14013260	0.00046296	-0.00434783	0.00493336
fit0.4	38.79090909	29.83139244	49.74464580	-0.18954545	-0.64419192	0.17418103	-0.00136364	-0.00492071	0.00328683
fit0.5	39.08561208	29.59427572	47.29127948	-0.13156598	-0.48431904	0.24256936	-0.00204293	-0.00571978	0.00172630
fit0.6	36.30400000	26.91236870	46.57357367	0.05600000	-0.36562469	0.44546489	-0.00400000	-0.00796333	0.00023389
fit0.7	35.43421053	24.80867857	49.55557875	0.16206140	-0.42500000	0.61746128	-0.00504386	-0.00991429	0.00092639
fit0.75	37.06990553	23.17532322	51.68217906	0.14028340	-0.44874416	0.71469368	-0.00492578	-0.01077501	0.00087933
fit0.8	42.74829932	27.61014348	52.20000000	-0.03265306	-0.41126892	0.59666723	-0.00340136	-0.00980392	0.00036332
fit0.9	37.39480519	28.70107143	55.42313862	0.30930736	-0.42363415	0.65714547	-0.00692641	-0.01031061	0.00047619
fit0.95	36.22440000	23.26127906	60.39901657	0.50340000	-0.54513889	0.99632070	-0.00940000	-0.01404791	0.00176304
<b>Men</b>									
fit0.05	28.48571429	13.51236584	37.21760006	0.09285714	-0.27608749	0.74345431	-0.00396825	-0.01093750	-0.00000000
fit0.1	27.88750000	19.37541893	41.02363248	0.25555556	-0.31394850	0.60809598	-0.00601852	-0.00960426	0.00000000
fit0.2	31.39411765	20.99703725	39.25569048	0.22835294	-0.10408144	0.67366977	-0.00576471	-0.01041098	-0.00236799
fit0.25	29.13552632	23.26838235	37.53445760	0.37580409	0.01533095	0.63142962	-0.00738304	-0.01012987	-0.00365365
fit0.3	29.32000000	22.28888889	38.71196815	0.42400000	0.03333180	0.73387758	-0.00800000	-0.01134086	-0.00400942
fit0.4	31.32521008	20.88379840	40.66076087	0.42710084	0.03901366	0.84573873	-0.00819328	-0.01233775	-0.00414456
fit0.5	35.04000000	25.28893872	43.13656566	0.35500000	0.01778344	0.77170597	-0.00750000	-0.01192318	-0.00406143
fit0.6	35.34126984	29.02319610	44.60513183	0.42412698	0.04187771	0.70001282	-0.00825397	-0.01118308	-0.00443175
fit0.7	35.49010989	26.26422054	46.28282552	0.52516484	0.06362250	0.90667045	-0.00956044	-0.01335110	-0.00467829
fit0.75	36.55503247	27.04126082	46.66736161	0.52857143	0.11196168	0.93800238	-0.00957792	-0.01388286	-0.00540391
fit0.8	37.67414530	26.01267857	50.32868654	0.55384615	0.00917467	1.03611742	-0.01004274	-0.01497867	-0.00428571
fit0.9	48.66599327	33.83227273	66.70263062	0.23686869	-0.50892308	0.86269943	-0.00690236	-0.01339288	0.00070355
fit0.95	47.24897959	33.48281704	61.63508967	0.38350340	-0.21507143	0.96266353	-0.00799320	-0.01387266	-0.00182853

**Note:** Relative  $\dot{V}O_{2peak}$  was measured in mL $O_2$ /min/kg.

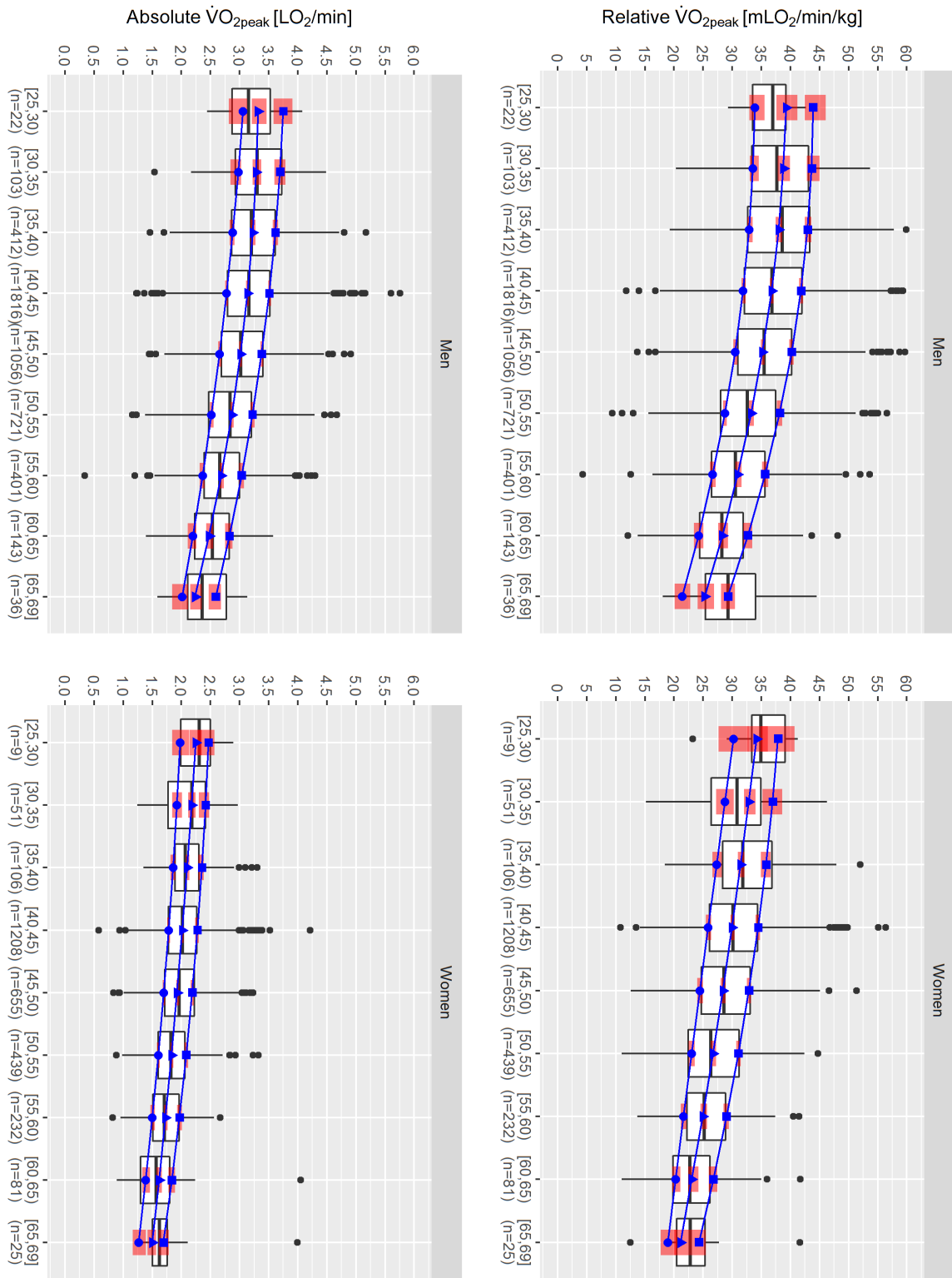


**Figure 4.5:** Coefficient plots of all quantile regressions coefficients plotted in nomograms (absolute  $\dot{V}O_{2peak}$ , table 4.5).



**Figure 4.6:** Coefficient plots of all quantile regressions coefficients plotted in nomograms (relative  $\dot{V}O_{2peak}$ , table 4.6).

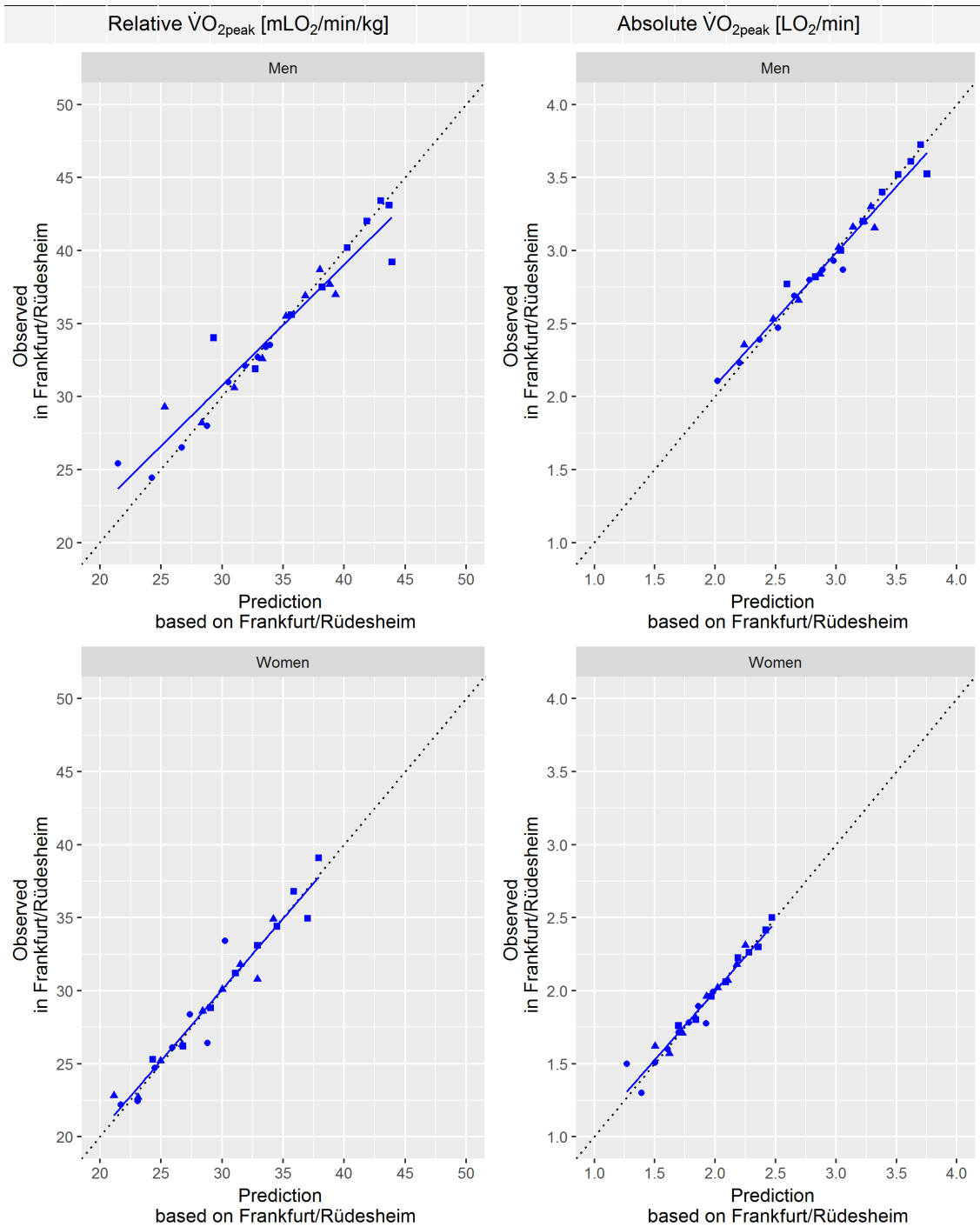




**Figure 4.7:** Apparent validation of quantile regression models.

**Note:** Quantiles are displayed as  $\bullet = 0.25$ ,  $\blacktriangle = 0.5$ ,  $\blacksquare = 0.75$ .

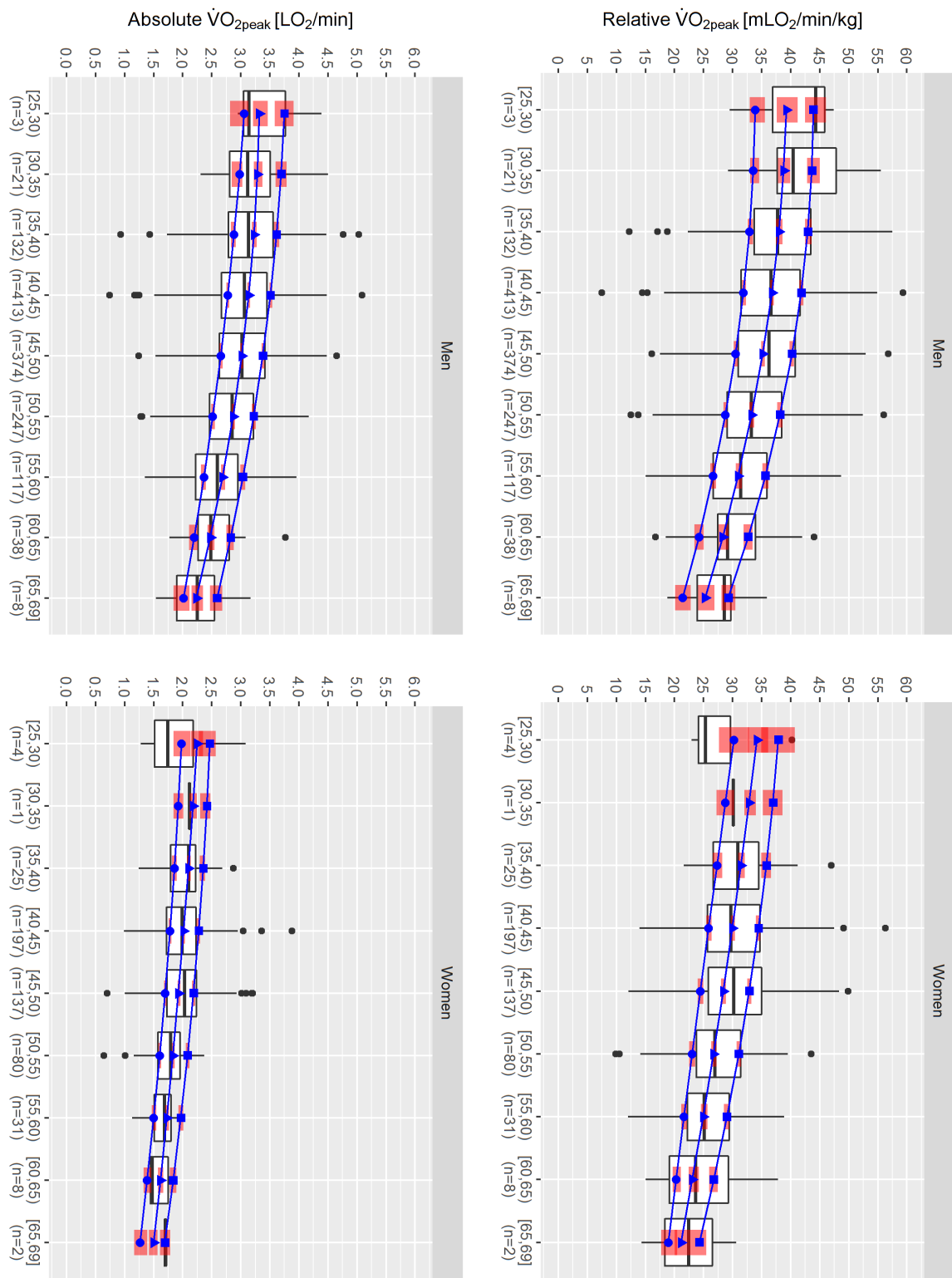
For apparent validation, the box plots as well as the regression predictions were based on participants from Frankfurt/Rüdesheim. The 95% confidence intervals for predicted values are plotted in red; age was modelled in classes.



**Figure 4.8:** Apparent validation calibration plots.

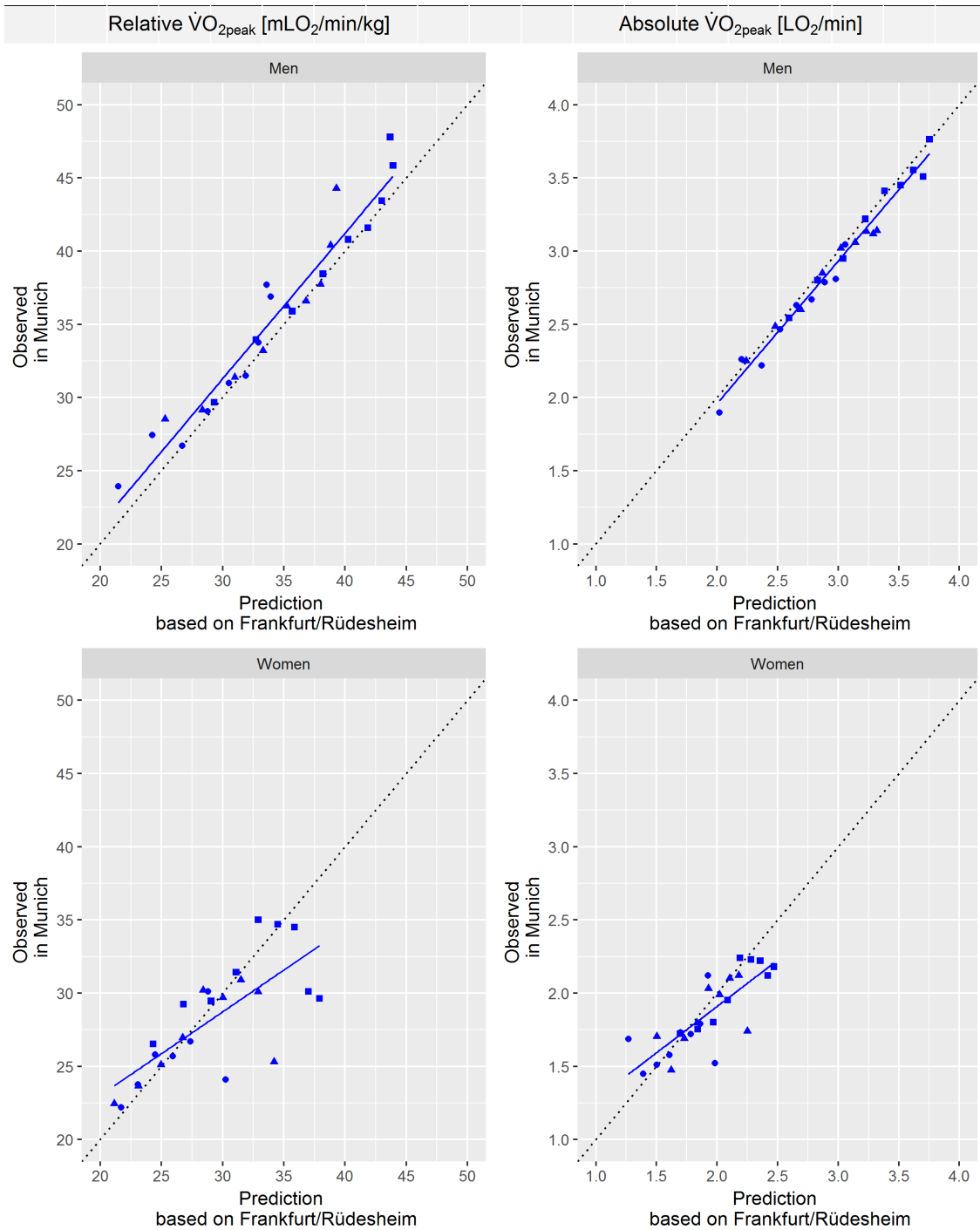
**Note:** Quantiles are displayed as ● = 0.25, ▲ = 0.5, ■ = 0.75.

For apparent validation, the observed values as well as the regression predictions were based on participants from Frankfurt/Rüdesheim.



**Figure 4.9:** Apparent validation of quantile regression models.

**Note:** Quantiles are displayed as  $\bullet = 0.25$ ,  $\blacktriangle = 0.5$ ,  $\blacksquare = 0.75$ . For external validation, participants from Munich were used for the box plots and participants from Frankfurt/Rüdesheim were used to obtain predicted values. The 95% confidence intervals for predicted values are plotted in red; age was modelled in classes.



**Figure 4.10:** External validation calibration plots.

**Note:** Quantiles are displayed as ● = 0.25, ▲ = 0.5, ■ = 0.75.

For external validation, the observed values were based on participants from Munich and the predicted values were based on participants from Frankfurt/Rüdesheim.

**Table 4.7:** Regression coefficients and R squared for linear regression in calibration data for apparent validation.

	Absolute $\dot{V}O_{2peak}$	Relative $\dot{V}O_{2peak}$
<b>Women</b>		
Intercept	0.11 [-0.12 to 0.33]	1.06 [-1.29 to 3.35]
Slope	0.94 [0.83 to 1.05]	0.97 [0.88 to 1.05]
R squared	0.95 [0.92 to 0.99]	0.95 [0.9 to 0.98]
<b>Men</b>		
Intercept	0.26 [0.07 to 0.45]	5.94 [1.35 to 9.91]
Slope	0.91 [0.84 to 0.97]	0.83 [0.71 to 0.96]
R squared	0.97 [0.95 to 0.99]	0.92 [0.85 to 0.97]

95% confidence intervals in square brackets.

**Table 4.8:** Regression coefficients and R squared for linear regression in calibration data for external validation.

	Absolute $\dot{V}O_{2peak}$	Relative $\dot{V}O_{2peak}$
<b>Women</b>		
Intercept	0.64 [0.26 to 0.96]	8.54 [2.48 to 14.63]
Slope	0.63 [0.46 to 0.83]	0.67 [0.45 to 0.91]
R squared	0.64 [0.4 to 0.85]	0.65 [0.4 to 0.87]
<b>Men</b>		
Intercept	0.01 [-0.18 to 0.22]	1.47 [-2.73 to 4.61]
Slope	0.97 [0.9 to 1.04]	0.99 [0.9 to 1.12]
R squared	0.98 [0.96 to 0.99]	0.94 [0.9 to 0.97]

95% confidence intervals in square brackets.

## 4.4 Exercise test modalities and random sample

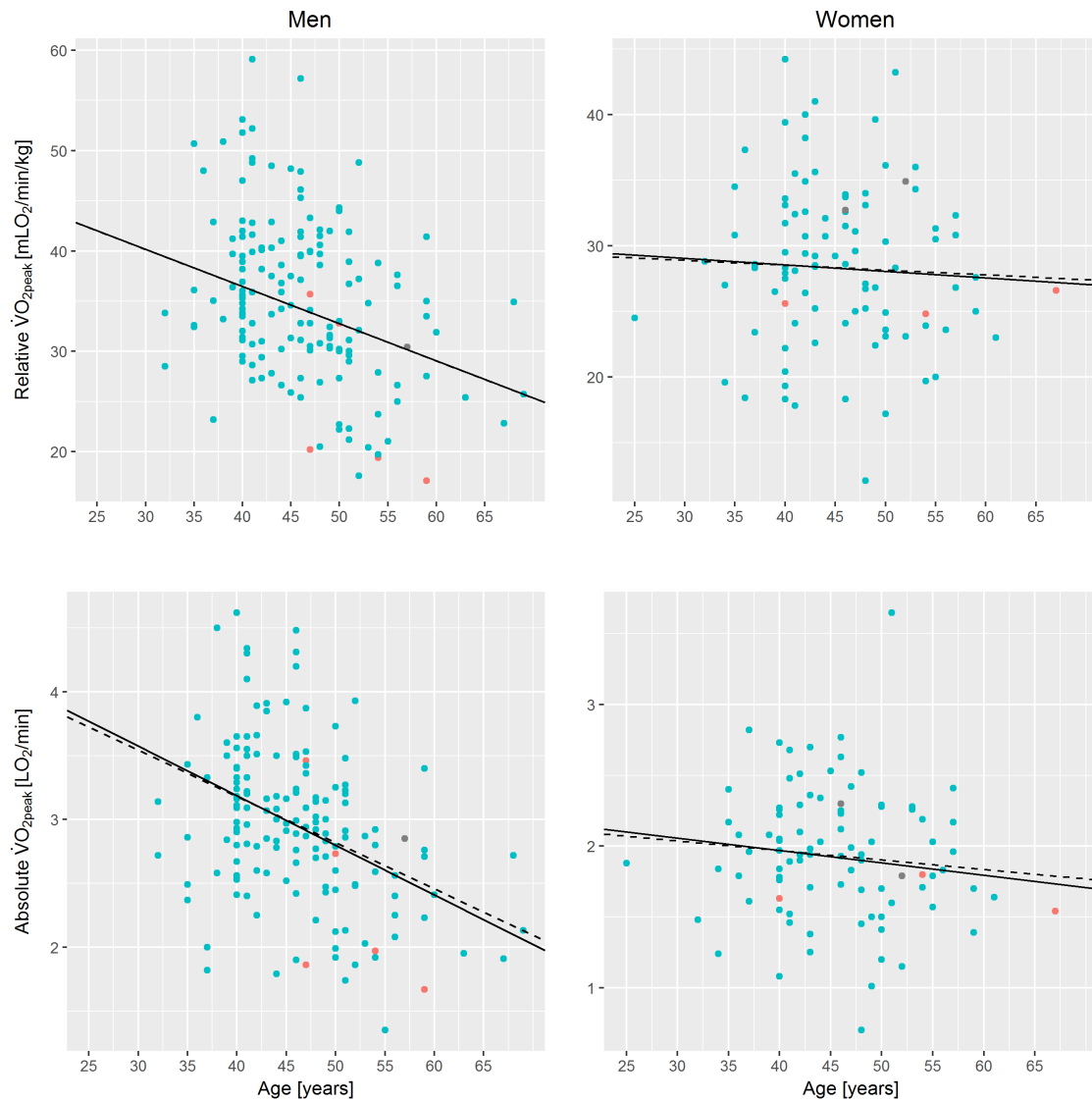
Some characteristics of CPET test modalities were not recorded in the full dataset. Therefore, 252 participants were randomly selected from the entire main study dataset, and CPET modalities were manually added from medical records for this random sample (chapter 3.3). Some characteristics were compared between full dataset and the random sample (table A.1), to evaluate wheather the random sample was representative for the whole dataset. However, none of the characteristics were significantly different in the two datasets. Therefore, the representativity of the random sample was assumed.

Three measures of the participants' effort were analysed in the random sample. Adequate effort was defined when either  $RER \geq 1.1$  or lactate  $\geq 8$  mmol/L or the heart rate was  $\geq 90\%$  of the age-predicted maximal heart rate [66, 45]. Overall, 239/247 (96.8% 95% CI 94.4% to 98.8%) of the observations achieved exertion as defined above. This proportion was 150/155 (96.8% 95% CI 93.6% to 99.4%) in men and 89/92 (96.7% 95% CI 92.5% to 100%) in women. In five participants who were selected for the random sample, none of these three variables was recorded.

Multistage protocols were used in 130/243 (54%, 95% CI 47% to 60%) exercise tests of the random sample, and ramp protocols in 113/243 (47%, 95% CI 40% to 53%) exercise tests of the random sample. The overall median maximal heart rate was 174/min [164;182]

(men: 174/min [165;182], women: 175/min [163;182]).

To depict how reference values were affected when subjects with poor effort were excluded, median regressions were fitted for i) the entire random sample and ii) individuals with no adequate effort excluded (fig. 4.11). The regression lines were approximately congruent. In addition to the graphical approach, median regressions were also calculated for the sample data, including an interaction term of age and exertion as the independent variable (4.9). None of the regression coefficients were statistically significant.



**Figure 4.11:** Median regressions for inclusion versus exclusion of participants with no maximal effort.

**Note:** Median regressions were calculated using i) all subjects without missing values in the random sample ( $n = 247$ ) and ii) only subjects that showed maximal effort ( $n = 239$ , dashed line, table 2.1). Participants with no adequate effort are printed red.

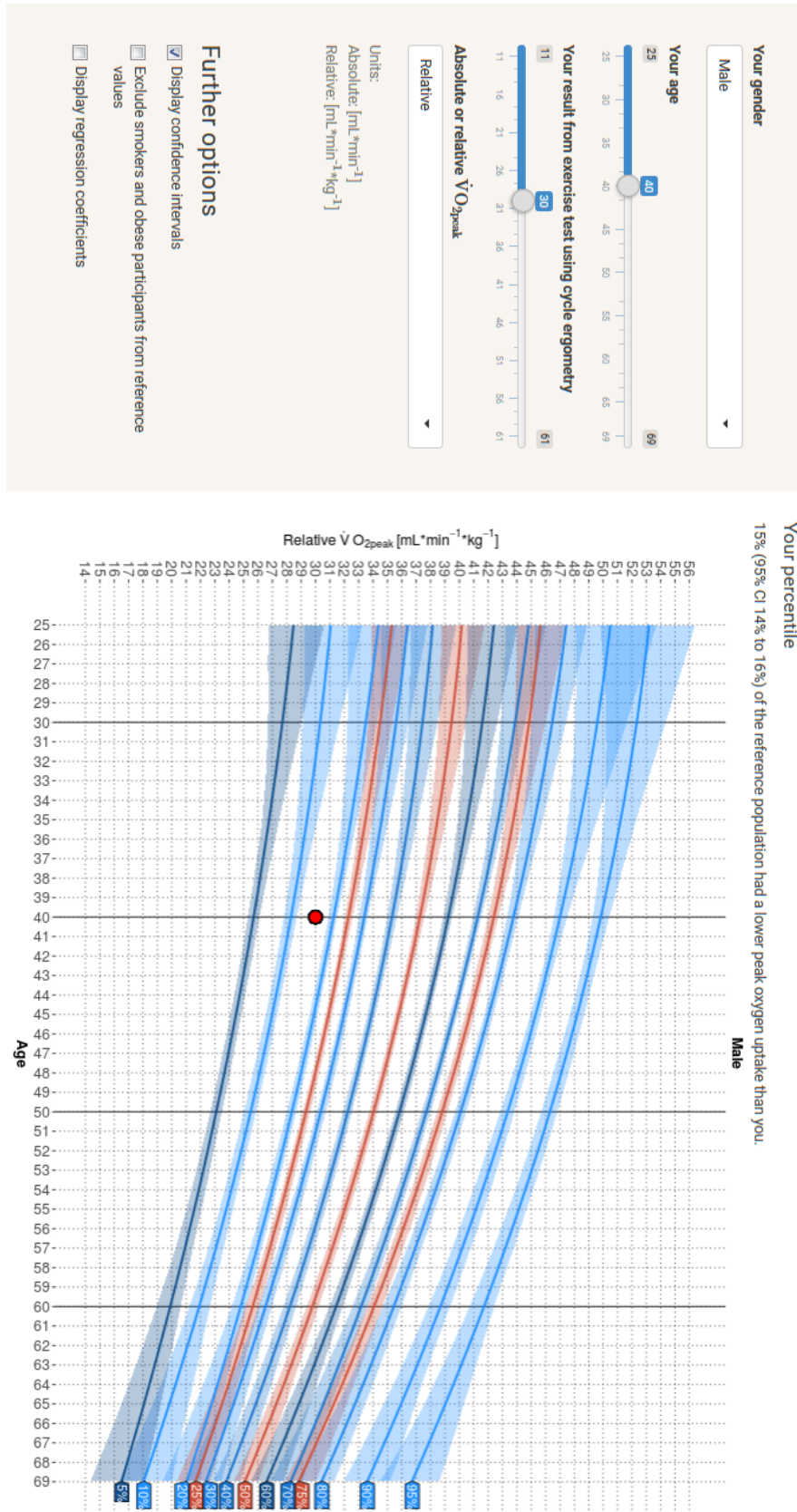
**Table 4.9:** Median regression using exertion and age as interaction terms.

	Coefficients	95% CI		P value
		Lower bound	Upper bound	
<b>Men, absolute <math>\dot{V}O_{2peak}</math></b>				
Intercept	12.61	-27.95	25.08	0.262
Age	-0.20	-0.45	0.61	0.352
Exertion(yes)	-3.99	-10.18	16.18	0.477
Age * Exertion(yes)	0.08	-0.32	0.21	0.447
<b>Men, relative <math>\dot{V}O_{2peak}</math></b>				
Intercept	165.81	-402.94	345.16	0.232
Age	-2.73	-6.24	8.72	0.309
Exertion(yes)	-57.26	-144.86	225.74	0.408
Age * Exertion(yes)	1.18	-4.52	2.89	0.380
<b>Women, absolute <math>\dot{V}O_{2peak}</math></b>				
Intercept	1.29	-0.39	3.95	0.322
Age	-0.00	-0.04	0.04	1.000
Exertion(yes)	0.47	-1.07	1.54	0.519
Age * Exertion(yes)	-0.00	-0.03	0.02	0.817
<b>Women, relative <math>\dot{V}O_{2peak}</math></b>				
Intercept	18.27	-1.48	31.99	0.108
Age	0.11	-0.24	0.57	0.633
Exertion(yes)	5.85	-4.54	20.94	0.414
Age * Exertion(yes)	-0.07	-0.38	0.18	0.636

**Note:** Exertion was coded as “yes” =2 and “no” = 1. Only 5 men and 3 women were recorded in the group exertion = “no”. Therefore, the results have to be interpreted with caution.

## 4.5 Web application

An interactive web application was created to facilitate the interpretation of the present study’s results. This application can be accessed at [www.uks.eu/vo2peak](http://www.uks.eu/vo2peak). A screenshot is displayed in fig. 4.12.



**Figure 4.12:** Screenshot of the  $\dot{V}O_{2peak}$  calculator web application which was deployed as part of the present study ([www.uks.eu/vo2peak](http://www.uks.eu/vo2peak), section 3.6, [55])



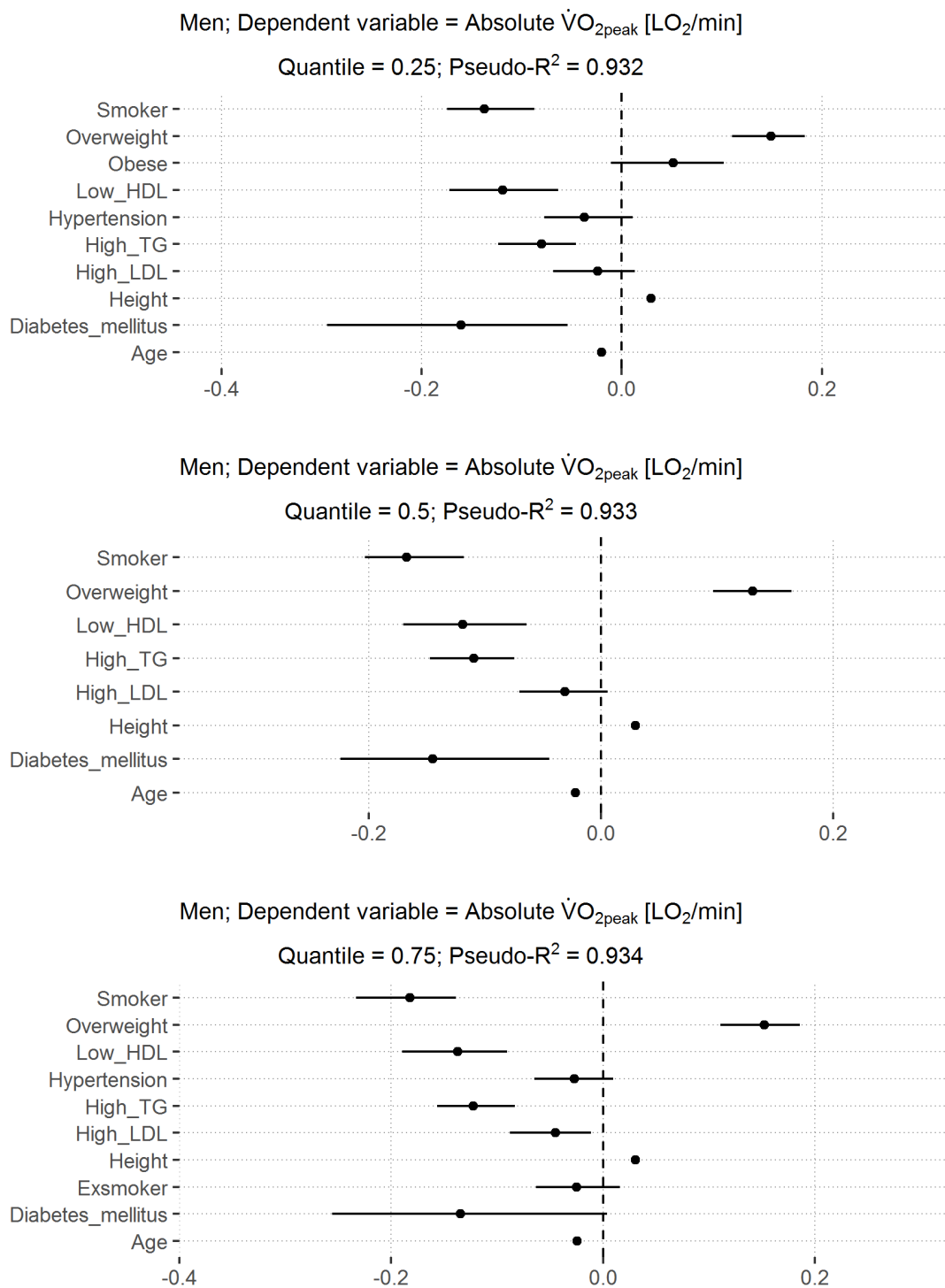
## 4.6 Multivariable analyses

An explorative multivariable regression modelling was performed to find associations of peak oxygen uptake and characteristics of the present study sample. Quantile regressions were calculated using the quantiles 0.25, 0.5, and 0.75. The regression coefficients are displayed in tables 4.10 to 4.13. Figs. 4.13, 4.14, 4.15, and 4.16 are coefficient plots of the regression coefficients.

Correlation matrices were calculated for all eligible quantitative variables that were considered for multivariable regression (figs. A.3 and A.4) to avoid collinearity. Some variables were excluded because they were shown to be highly correlated. A high correlation was observed between BMI, body weight, waist circumference and body fat. Therefore, only BMI was considered for multivariable analyses and body weight, waist circumference and body fat were excluded. Another high correlation was present between total cholesterol and LDL cholesterol. As total cholesterol includes HDL and LDL cholesterol, and HDL cholesterol was also considered for the analysis, total cholesterol was excluded. LDL and HDL were recoded into binary variables (section 3.2) because some skewed distributions were observed.

The regressions' goodness of fit showed to be quite high in absolute  $\dot{V}O_{2\text{peak}}$  and lower in relative  $\dot{V}O_{2\text{peak}}$ . In absolute  $\dot{V}O_{2\text{peak}}$ , all R squared were above 90%. In relative  $\dot{V}O_{2\text{peak}}$ , R squared were observed to be around 20% (figs. 4.13 to 4.16).

Some characteristics were associated with peak oxygen uptake in the multivariable analyses. Age showed statistically significant associations in both sexes and in all regressions. Peak oxygen uptake was lower in older ages. In both men and women, cigarette smoking was found to be strongly associated with absolute and relative  $\dot{V}O_{2\text{peak}}$ . This was also observed across all regression quantiles. In men, for example, the estimated 0.75 quantile of relative  $\dot{V}O_{2\text{peak}}$  was 2.2 mL $O_2$ /min/kg (95% CI 2.9 to 1.3) lower compared to non-smokers. Furthermore, triglycerides were shown to be negatively associated with peak oxygen uptake in all regressions. In women with elevated triglycerides, the estimated 0.75 quantile of relative  $\dot{V}O_{2\text{peak}}$  was 2.2 mL $O_2$ /min/kg (95% CI 2.9 to 1.1) lower compared to women with triglyceride levels below 150 mg/dL. Low HDL cholesterol levels were also strongly associated with lower peak oxygen uptake across all regressions of men but in none of the models of women. The estimated 0.75 quantile of relative  $\dot{V}O_{2\text{peak}}$  was 2.1 mL $O_2$ /min/kg (95% CI 2.9 to 1.4) lower compared to men with HDL cholesterol levels above 40 mg/dL. The body composition, represented by overweight and obesity, was also associated with decreased relative and increased absolute  $\dot{V}O_{2\text{peak}}$ . On the other hand, body height was shown to be positively associated with absolute peak oxygen uptake. Diabetes mellitus and hypertension were associated with lower peak oxygen uptake especially in men.

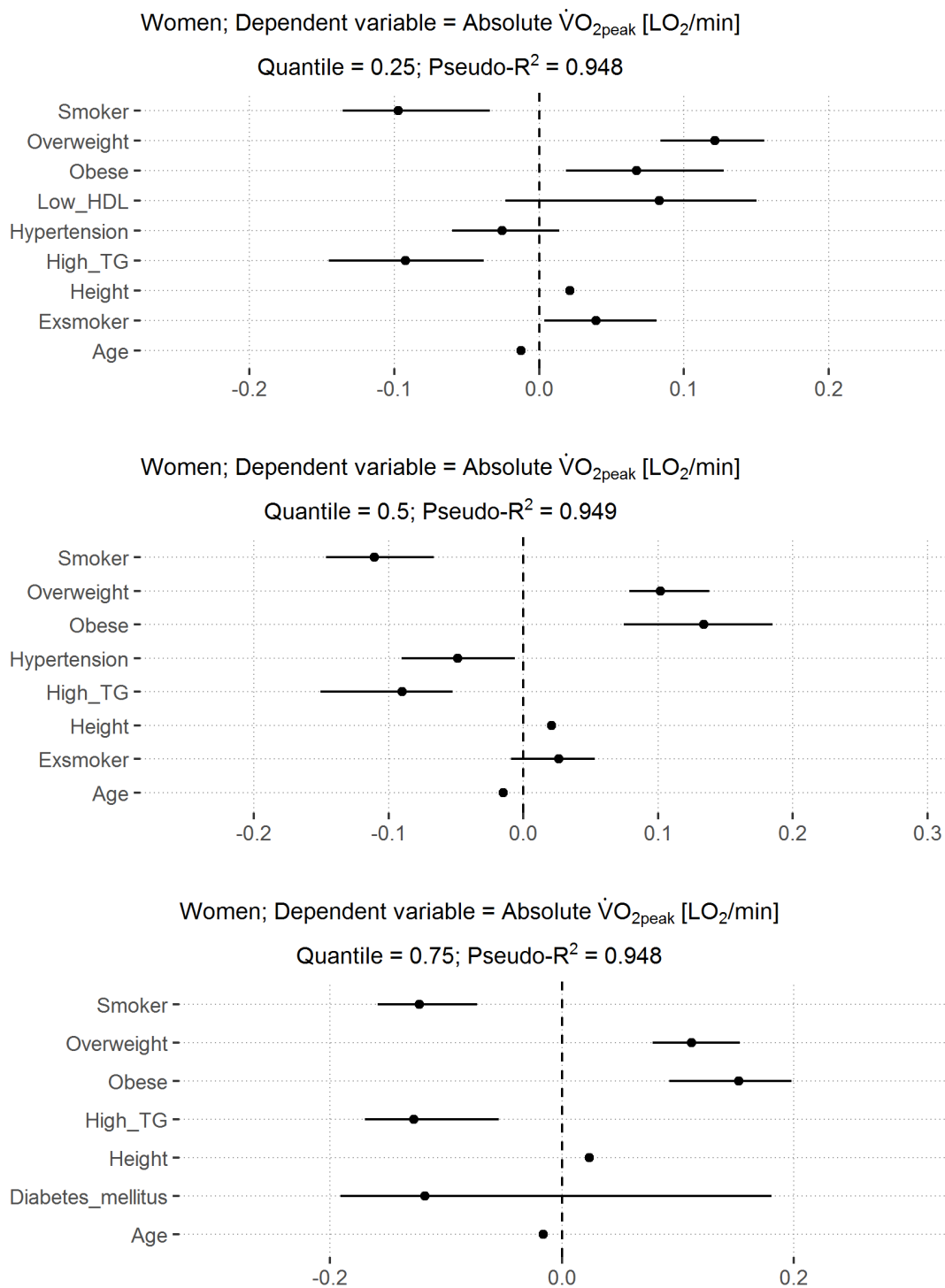


**Figure 4.13:** Coefficient plots of multivariable quantile regression models for absolute  $\dot{V}O_{2peak}$  in men.

**Note:** Dependent variable = absolute  $\dot{V}O_{2peak}$  [ $LO_2/min$ ].

Quantile regressions for the quantiles 0.25, 0.5, and 0.75 were fitted. Independent variables were selected using stepwise regression. AIC was used as criterion statistic for model comparison. Backward and forward variable elimination were selected as direction. Pseudo  $R^2$  was calculated according to [22, p. 52].

Units of measurement: age [years], height [cm], overweight:  $BMI \geq 25 \text{ kg/m}^2$ , obesity:  $BMI \geq 30 \text{ kg/m}^2$ , hypertension: systolic blood pressure  $\geq 140 \text{ mmHg}$  or diastolic blood pressure  $\geq 90 \text{ mmHg}$ , low HDL: HDL cholesterol  $\leq 40 \text{ mg/dL}$ , high LDL: LDL cholesterol  $\geq 115 \text{ mg/dL}$ , high TG: triglycerides  $\geq 150 \text{ mg/dL}$ , diabetes mellitus: fasting glucose  $\geq 126 \text{ mg/dL}$  or  $HbA_{1c} \geq 6.5\%$  (section 3.2).

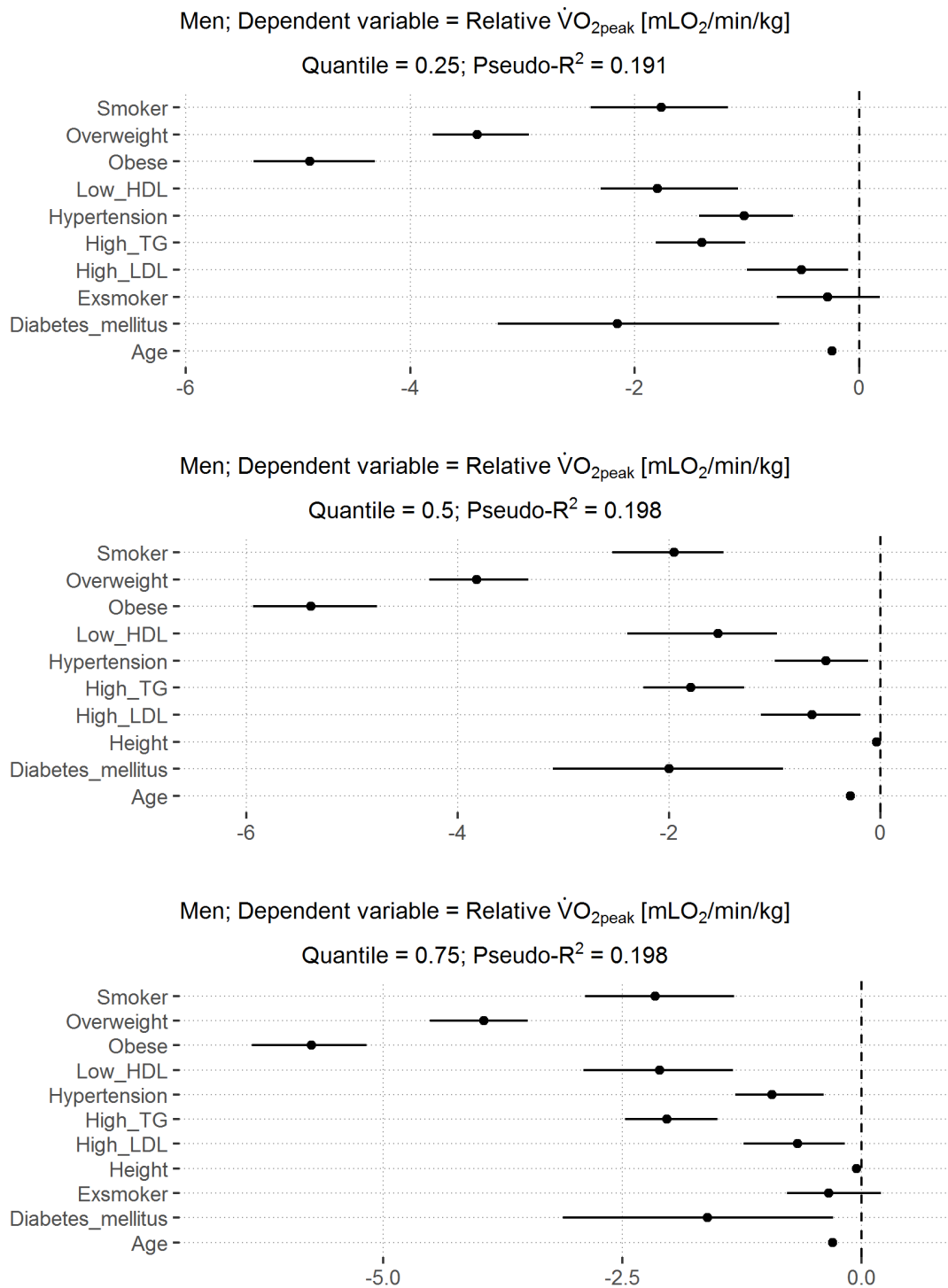


**Figure 4.14:** Coefficient plots of multivariable quantile regression models for absolute  $\dot{V}O_{2peak}$  in women.

**Note:** Dependent variable = absolute  $\dot{V}O_{2peak}$  [ $LO_2/min$ ].

Quantile regressions for the quantiles 0.25, 0.5, and 0.75 were fitted. Independent variables were selected using stepwise regression. AIC was used as criterion statistic for model comparison. Backward variable elimination was selected as direction. Pseudo  $R^2$  was calculated according to [22, p. 52].

Units of measurement: age [years], height [cm], overweight:  $BMI \geq 25 \text{ kg/m}^2$ , obesity:  $BMI \geq 30 \text{ kg/m}^2$ , hypertension: systolic blood pressure  $\geq 140 \text{ mmHg}$  or diastolic blood pressure  $\geq 90 \text{ mmHg}$ , low HDL: HDL cholesterol  $\leq 40 \text{ mg/dL}$ , high LDL: LDL cholesterol  $\geq 115 \text{ mg/dL}$ , high TG: triglycerides  $\geq 150 \text{ mg/dL}$ , diabetes mellitus: fasting glucose  $\geq 126 \text{ mg/dL}$  or  $HbA_{1c} \geq 6.5\%$  (section 3.2).

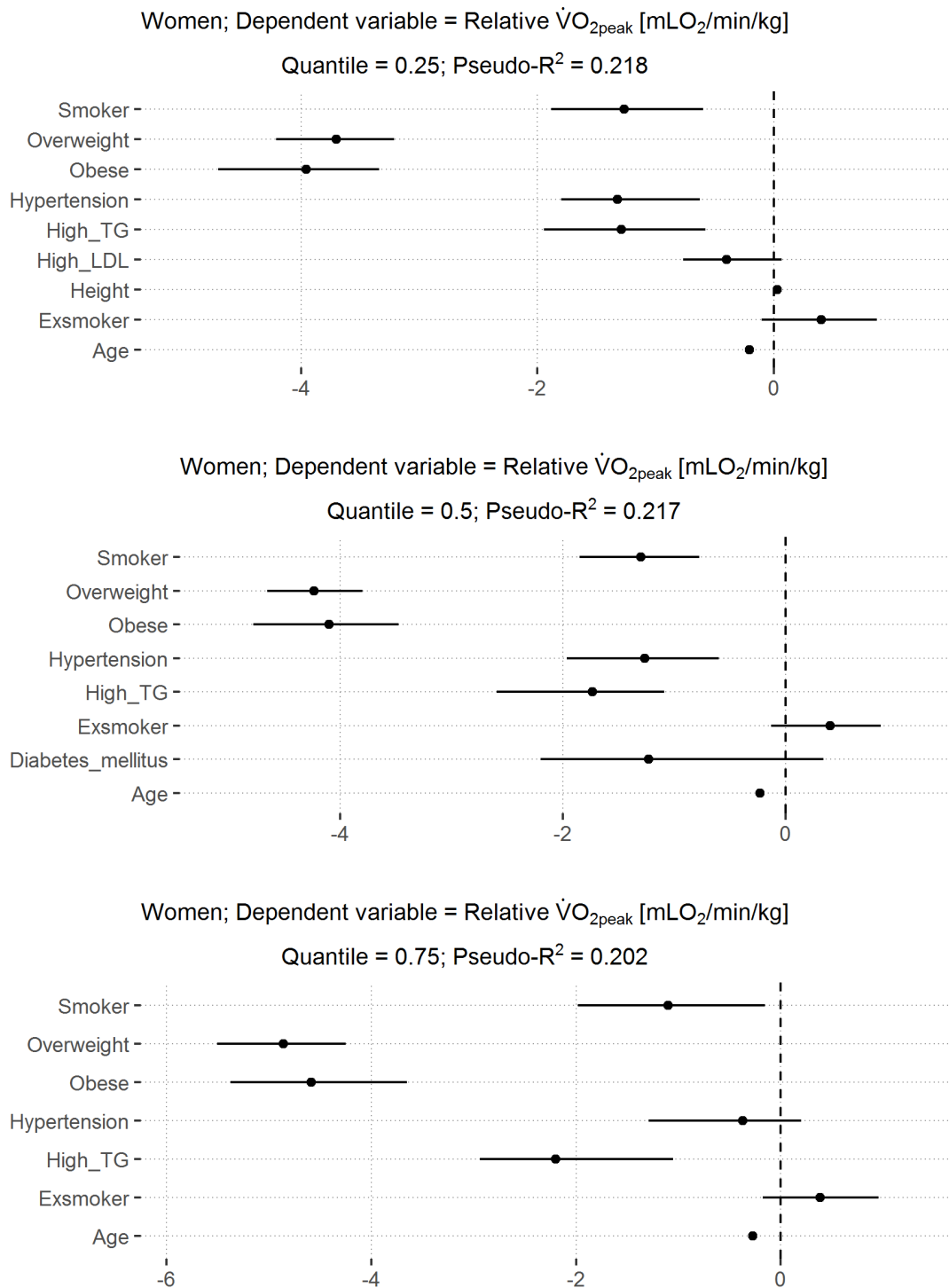


**Figure 4.15:** Coefficient plots of multivariable quantile regression models for relative  $\dot{V}O_{2peak}$  in men.

**Note:** Dependent variable = relative  $\dot{V}O_{2peak}$  [mLO<sub>2</sub>/min/kg].

Quantile regressions for the quantiles 0.25, 0.5, and 0.75 were fitted. Independent variables were selected using stepwise regression. AIC was used as criterion statistic for model comparison. Backward variable elimination was selected as direction. Pseudo R<sup>2</sup> was calculated according to [22, p. 52].

Units of measurement: age [years], height [cm], overweight: BMI ≥ 25 kg/m<sup>2</sup>, obesity: BMI ≥ 30 kg/m<sup>2</sup>, hypertension: systolic blood pressure ≥ 140 mmHg or diastolic blood pressure ≥ 90 mmHg, low HDL: HDL cholesterol ≤ 40 mg/dL, high LDL: LDL cholesterol ≥ 115 mg/dL, high TG: triglycerides ≥ 150 mg/dL, diabetes mellitus: fasting glucose ≥ 126 mg/dL or HbA<sub>1c</sub> ≥ 6.5% (section 3.2).



**Figure 4.16:** Coefficient plots of multivariable quantile regression models for relative  $\dot{V}O_{2peak}$  in women.

**Note:** Dependent variable = relative  $\dot{V}O_{2peak}$  [mL $O_2$ /min/kg].

Quantile regressions for the quantiles 0.25, 0.5, and 0.75 were fitted. Independent variables were selected using stepwise regression. AIC was used as criterion statistic for model comparison. Backward variable elimination was selected as direction. Pseudo  $R^2$  was calculated according to [22, p. 52].

Units of measurement: age [years], height [cm], overweight: BMI  $\geq 25$  kg/m $^2$ , obesity: BMI  $\geq 30$  kg/m $^2$ , hypertension: systolic blood pressure  $\geq 140$  mmHg or diastolic blood pressure  $\geq 90$  mmHg, low HDL: HDL cholesterol  $\leq 40$  mg/dL, high LDL: LDL cholesterol  $\geq 115$  mg/dL, high TG: triglycerides  $\geq 150$  mg/dL, diabetes mellitus: fasting glucose  $\geq 126$  mg/dL or HbA $1c$   $\geq 6.5\%$  (section 3.2).

**Table 4.10:** Regression coefficients of multivariable quantile regressions for relative  $\dot{V}O_{2\text{peak}}$  in men.

	Coefficient	95% CI		P value
		Lower bound	Upper bound	
<b>Quantile 0.25</b>				
Intercept	46.80	45.38	48.10	<0.001
Age	-0.24	-0.27	-0.21	<0.001
Overweight	-3.41	-3.80	-2.94	<0.001
Obese	-4.89	-5.40	-4.32	<0.001
Hypertension	-1.03	-1.43	-0.59	<0.001
Low HDL	-1.80	-2.31	-1.08	<0.001
High LDL	-0.52	-1.00	-0.10	0.025
High TG	-1.40	-1.81	-1.02	<0.001
Diabetes mellitus	-2.16	-3.22	-0.71	0.001
Smoker	-1.76	-2.39	-1.17	<0.001
Ex-smoker	-0.28	-0.73	0.18	0.219
<b>Quantile 0.50</b>				
Intercept	59.61	52.47	65.60	<0.001
Age	-0.28	-0.31	-0.25	<0.001
Height	-0.04	-0.07	<-0.01	0.029
Overweight	-3.82	-4.27	-3.33	<0.001
Obese	-5.39	-5.94	-4.76	<0.001
Hypertension	-0.52	-1.00	-0.11	0.023
Low HDL	-1.53	-2.39	-0.98	<0.001
High LDL	-0.65	-1.13	-0.19	0.006
High TG	-1.80	-2.24	-1.29	<0.001
Diabetes mellitus	-2.00	-3.10	-0.92	<0.001
Smoker	-1.95	-2.54	-1.48	<0.001
<b>Quantile 0.75</b>				
Intercept	68.18	62.22	73.78	<0.001
Age	-0.30	-0.34	-0.26	<0.001
Height	-0.05	-0.08	-0.02	<0.001
Overweight	-3.95	-4.51	-3.49	<0.001
Obese	-5.75	-6.37	-5.17	<0.001
Hypertension	-0.94	-1.32	-0.40	<0.001
Low HDL	-2.11	-2.91	-1.35	<0.001
High LDL	-0.67	-1.23	-0.18	0.013
High TG	-2.03	-2.47	-1.51	<0.001
Diabetes mellitus	-1.61	-3.13	-0.30	0.031
Smoker	-2.16	-2.89	-1.33	<0.001
Exsmoker	-0.34	-0.78	0.20	0.178

**Note:** Dependent variable = relative  $\dot{V}O_{2\text{peak}}$  [mL $O_2$ /min/kg].

Quantile regressions for the quantiles 0.25, 0.5, and 0.75 were fitted. Independent variables were selected using stepwise regression. AIC was used as criterion statistic for model comparison. Backward variable elimination was selected as direction. Pseudo  $R^2$  was calculated according to [22, p. 52].

Units of measurement: age [years], height [cm], overweight: BMI  $\geq 25$  kg/m<sup>2</sup>, obesity: BMI  $\geq 30$  kg/m<sup>2</sup>, hypertension: systolic blood pressure  $\geq 140$  mmHg or diastolic blood pressure  $\geq 90$  mmHg, low HDL: HDL cholesterol  $\leq 40$  mg/dL, high LDL: LDL cholesterol  $\geq 115$  mg/dL, high TG: triglycerides  $\geq 150$  mg/dL, diabetes mellitus: fasting glucose  $\geq 126$  mg/dL or HbA<sub>1c</sub>  $\geq 6.5\%$  (section 3.2).

**Table 4.11:** Regression coefficients of multivariable quantile regressions for relative  $\dot{V}O_{2\text{peak}}$  in women.

	Coefficient	95% CI		P value
		Lower bound	Upper bound	
<b>Quantile 0.25</b>				
Intercept	32.26	25.74	38.11	<0.001
Age	-0.20	-0.23	-0.17	<0.001
Height	0.03	<-0.01	0.07	0.077
Overweight	-3.70	-4.21	-3.21	<0.001
Obese	-3.96	-4.70	-3.34	<0.001
Hypertension	-1.32	-1.80	-0.63	<0.001
High LDL	-0.40	-0.77	0.07	0.064
High TG	-1.29	-1.94	-0.58	<0.001
Smoker	-1.27	-1.88	-0.60	<0.001
Ex-smoker	0.40	-0.10	0.87	0.099
<b>Quantile 0.50</b>				
Intercept	42.03	40.48	43.50	<0.001
Age	-0.23	-0.26	-0.20	<0.001
Overweight	-4.23	-4.65	-3.80	<0.001
Obese	-4.10	-4.78	-3.47	<0.001
Hypertension	-1.27	-1.97	-0.60	<0.001
High TG	-1.73	-2.59	-1.09	<0.001
Diabetes mellitus	-1.23	-2.20	0.34	0.065
Smoker	-1.30	-1.85	-0.77	<0.001
Ex-smoker	0.40	-0.13	0.85	0.115
<b>Quantile 0.75</b>				
Intercept	47.36	45.34	49.11	<0.001
Age	-0.27	-0.31	-0.23	<0.001
Overweight	-4.86	-5.51	-4.25	<0.001
Obese	-4.59	-5.38	-3.65	<0.001
Hypertension	-0.37	-1.29	0.20	0.344
High TG	-2.20	-2.94	-1.05	<0.001
Smoker	-1.10	-1.98	-0.15	0.016
Ex-smoker	0.39	-0.18	0.95	0.177

**Note:** Dependent variable = relative  $\dot{V}O_{2\text{peak}}$  [mL $O_2$ /min/kg].

Quantile regressions for the quantiles 0.25, 0.5, and 0.75 were fitted. Independent variables were selected using stepwise regression. AIC was used as criterion statistic for model comparison. Backward variable elimination was selected as direction. Pseudo  $R^2$  was calculated according to [22, p. 52].

Units of measurement: age [years], height [cm], overweight: BMI  $\geq$  25 kg/m<sup>2</sup>, obesity: BMI  $\geq$  30 kg/m<sup>2</sup>, hypertension: systolic blood pressure  $\geq$  140 mmHg or diastolic blood pressure  $\geq$  90 mmHg, low HDL: HDL cholesterol  $\leq$  40 mg/dL, high LDL: LDL cholesterol  $\geq$  115 mg/dL, high TG: triglycerides  $\geq$  150 mg/dL, diabetes mellitus: fasting glucose  $\geq$  126 mg/dL or HbA<sub>1c</sub>  $\geq$  6.5% (section 3.2).

**Table 4.12:** Regression coefficients of multivariable quantile regressions for absolute  $\dot{V}O_{2\text{peak}}$  in men.

	Coefficient	95% CI		P value
		Lower bound	Upper bound	
<b>Quantile 0.25</b>				
Intercept	-1.70	-2.25	-1.20	<0.001
Age	-0.02	-0.02	-0.02	<0.001
Height	0.03	0.03	0.03	<0.001
Overweight	0.15	0.11	0.18	<0.001
Obese	0.05	-0.01	0.10	0.072
Hypertension	-0.04	-0.08	0.01	0.098
Low HDL	-0.12	-0.17	-0.06	<0.001
High LDL	-0.02	-0.07	0.01	0.251
High TG	-0.08	-0.12	-0.05	<0.001
Diabetes mellitus	-0.16	-0.29	-0.05	0.006
Smoker	-0.14	-0.17	-0.09	<0.001
<b>Quantile 0.50</b>				
Intercept	-1.31	-1.73	-0.79	<0.001
Age	-0.02	-0.02	-0.02	<0.001
Height	0.03	0.03	0.03	<0.001
Overweight	0.13	0.10	0.16	<0.001
Low HDL	-0.12	-0.17	-0.06	<0.001
High LDL	-0.03	-0.07	0.01	0.104
High TG	-0.11	-0.15	-0.07	<0.001
Diabetes mellitus	-0.15	-0.22	-0.04	0.002
Smoker	-0.17	-0.20	-0.12	<0.001
<b>Quantile 0.75</b>				
Intercept	-1.02	-1.54	-0.50	<0.001
Age	-0.02	-0.03	-0.02	<0.001
Height	0.03	0.03	0.03	<0.001
Overweight	0.15	0.11	0.19	<0.001
Hypertension	-0.03	-0.07	0.01	0.154
Low HDL	-0.14	-0.19	-0.09	<0.001
High LDL	-0.04	-0.09	-0.01	0.021
High TG	-0.12	-0.16	-0.08	<0.001
Diabetes mellitus	-0.13	-0.26	0.00	0.053
Smoker	-0.18	-0.23	-0.14	<0.001
Ex-smoker	-0.02	-0.06	0.02	0.211

**Note:** Dependent variable = absolute  $\dot{V}O_{2\text{peak}}$  [ $\text{L}O_2/\text{min}$ ].

Quantile regressions for the quantiles 0.25, 0.5, and 0.75 were fitted. Independent variables were selected using stepwise regression. AIC was used as criterion statistic for model comparison. Backward variable elimination was selected as direction. Pseudo  $R^2$  was calculated according to [22, p. 52].

Units of measurement: age [years], height [cm], overweight:  $\text{BMI} \geq 25 \text{ kg/m}^2$ , obesity:  $\text{BMI} \geq 30 \text{ kg/m}^2$ , hypertension: systolic blood pressure  $\geq 140 \text{ mmHg}$  or diastolic blood pressure  $\geq 90 \text{ mmHg}$ , low HDL: HDL cholesterol  $\leq 40 \text{ mg/dL}$ , high LDL: LDL cholesterol  $\geq 115 \text{ mg/dL}$ , high TG: triglycerides  $\geq 150 \text{ mg/dL}$ , diabetes mellitus: fasting glucose  $\geq 126 \text{ mg/dL}$  or  $\text{HbA}_{1c} \geq 6.5\%$  (section 3.2).



**Table 4.13:** Regression coefficients of multivariable quantile regressions for absolute  $\dot{V}O_{2\text{peak}}$  in women.

	Coefficient	95% CI		P value
		Lower bound	Upper bound	
<b>Quantile 0.25</b>				
Intercept	-1.32	-1.77	-0.89	<0.001
Age	-0.01	-0.02	-0.01	<0.001
Height	0.02	0.02	0.02	<0.001
Overweight	0.12	0.08	0.16	<0.001
Obese	0.07	0.02	0.13	0.013
Hypertension	-0.03	-0.06	0.01	0.173
Low HDL	0.08	-0.02	0.15	0.068
High TG	-0.09	-0.15	-0.04	0.001
Smoker	-0.10	-0.14	-0.03	<0.001
Ex-smoker	0.04	<0.01	0.08	0.046
<b>Quantile 0.50</b>				
Intercept	-0.92	-1.37	-0.54	<0.001
Age	-0.01	-0.02	-0.01	<0.001
Height	0.02	0.02	0.02	<0.001
Overweight	0.10	0.08	0.14	<0.001
Obese	0.13	0.07	0.18	<0.001
Hypertension	-0.05	-0.09	-0.01	0.023
High TG	-0.09	-0.15	-0.05	<0.001
Smoker	-0.11	-0.15	-0.07	<0.001
Ex-smoker	0.03	-0.01	0.05	0.097
<b>Quantile 0.75</b>				
Intercept	-1.10	-1.58	-0.62	<0.001
Age	-0.02	-0.02	-0.01	<0.001
Height	0.02	0.02	0.03	<0.001
Overweight	0.11	0.08	0.15	<0.001
Obese	0.15	0.09	0.20	<0.001
High TG	-0.13	-0.17	-0.05	<0.001
Diabetes mellitus	-0.12	-0.19	0.18	0.218
Smoker	-0.12	-0.16	-0.07	<0.001

**Note:** Dependent variable = absolute  $\dot{V}O_{2\text{peak}}$  [ $\text{LO}_2/\text{min}$ ].

Quantile regressions for the quantiles 0.25, 0.5, and 0.75 were fitted. Independent variables were selected using stepwise regression. AIC was used as criterion statistic for model comparison. Backward variable elimination was selected as direction. Pseudo  $R^2$  was calculated according to [22, p. 52].

Units of measurement: age [years], height [cm], overweight:  $\text{BMI} \geq 25 \text{ kg/m}^2$ , obesity:  $\text{BMI} \geq 30 \text{ kg/m}^2$ , hypertension: systolic blood pressure  $\geq 140 \text{ mmHg}$  or diastolic blood pressure  $\geq 90 \text{ mmHg}$ , low HDL: HDL cholesterol  $\leq 40 \text{ mg/dL}$ , high LDL: LDL cholesterol  $\geq 115 \text{ mg/dL}$ , high TG: triglycerides  $\geq 150 \text{ mg/dL}$ , diabetes mellitus: fasting glucose  $\geq 126 \text{ mg/dL}$  or  $\text{HbA}_{1c} \geq 6.5\%$  (section 3.2).

## 5 Discussion

Low CRF is a crucial predictor for a plethora of health threats as well as for premature all-cause mortality. The measurement of CRF in the course of preventive health screenings is valuable because regular exercise can increase CRF. CRF is accepted as one of the most important modifiable risk factors due to strong associations with cardiovascular disease and premature all-cause mortality (section 2.1). Therefore, incremental cardiopulmonary exercise testing is a frequently applied procedure in clinical practice to estimate cardiorespiratory fitness by determining peak oxygen uptake. However, as  $\dot{V}O_{2\text{peak}}$  is associated with sex and age [23], it is only conclusive to interpret the  $\dot{V}O_{2\text{peak}}$  of an individual in the light of sex-specific and age-specific reference values. The presented study provides sex-specific and age-specific reference values for peak oxygen uptake that are based on 25- to 69-year-old participants who underwent primary preventive health screenings and CPET using cycle ergometry in one of three German cities. By acquiring 9,354 participants, this study is one of the largest that has been published so far in this field [50] [53, p. 88].

This chapter discusses the results of the present study. First, the presented reference values are compared with the literature in section 5.1. The representativity of this study population and the generalisability of the results is described in section 5.2. The sections 5.3, 5.4, and 5.5 discuss the quality of  $\dot{V}O_{2\text{peak}}$  measurement, the statistical methods as well as the results of the multivariable quantile regressions, respectively. Finally the strengths and limitations of this study are outlined in section 5.6.

## 5.1 Reference values

The presented values for  $\dot{V}O_{2\text{peak}}$  showed a strong association with sex and age. Women had lower peak oxygen uptake compared to men and participants with higher ages had a lower peak oxygen uptake than younger participants. Furthermore, there was a wide range of  $\dot{V}O_{2\text{peak}}$  for a given sex and age, which was likely to be - at least partially - the result of physical activity and endurance exercise. The decline over age was present in both sexes and at all levels of fitness. Hawkins & Wiswell (2003) [23] described that the age-dependent decline of  $\dot{V}O_{2\text{peak}}$  was approximately 10% per decade regardless of the individual level of fitness. In the present study, this decline seems to be lower in both sexes. Median relative  $\dot{V}O_{2\text{peak}}$  in men at the ages of 30, 40, 50 and 60 years were 38.9, 37.2, 34.0 and 29.3 mL $O_2$ /min/kg, respectively. In women, it was 33.3, 30.6, 27.4 and 23.8 mL $O_2$ /min/kg, respectively (table 4.6\*). Some reasons for the decline might be difficulties maintaining training in higher ages, a decline of maximal heart rate and lean body mass. According to ACSM guidelines [53], physical activity and exercise should be performed to maintain CRF and lean body mass. Maximal heart rate, on the other hand, seems not to be strongly associated with physical activity [23].

Although CRF was lower in women and older individuals, the results also demonstrated that women with high fitness were fitter than men with poor fitness and also that elders with good fitness were fitter than young with low fitness. When relative  $\dot{V}O_{2\text{peak}}$  in men was compared, a 60-year-old at the 95% quantile showed a higher fitness level than the median at 30-years. Also, women at the 95% quantile were fitter than the median in men for the entire age-span.

The present reference values compared with past studies are plotted in fig. 5.1. For this comparison, the SHIP study by Koch et al. (2009) [34] was selected because the study was conducted in Germany. The FRIEND study by Kaminsky (2017) [31] was also chosen due to its recency and large sample size. The reference values of the present study and the SHIP study were comparable. The median of 40-year-old men was 37.2

---

\*

men:

$$\text{Relative } \dot{V}O_{2\text{peak}} = (0.355 * 30) + (-0.0075 * 30 * 30) + 35.04 = 38.9\text{mL}O_2/\text{min}/\text{kg}$$

$$\text{Relative } \dot{V}O_{2\text{peak}} = (0.355 * 40) + (-0.0075 * 40 * 40) + 35.04 = 37.2\text{mL}O_2/\text{min}/\text{kg}$$

$$\text{Relative } \dot{V}O_{2\text{peak}} = (0.355 * 50) + (-0.0075 * 50 * 50) + 35.04 = 34.0\text{mL}O_2/\text{min}/\text{kg}$$

$$\text{Relative } \dot{V}O_{2\text{peak}} = (0.355 * 60) + (-0.0075 * 60 * 60) + 35.04 = 29.3\text{mL}O_2/\text{min}/\text{kg}$$

women:

$$\text{Relative } \dot{V}O_{2\text{peak}} = (-0.131566 * 30) + (-0.002043 * 30 * 30) + 39.085612 = 33.3\text{mL}O_2/\text{min}/\text{kg}$$

$$\text{Relative } \dot{V}O_{2\text{peak}} = (-0.131566 * 40) + (-0.002043 * 40 * 40) + 39.085612 = 30.6\text{mL}O_2/\text{min}/\text{kg}$$

$$\text{Relative } \dot{V}O_{2\text{peak}} = (-0.131566 * 50) + (-0.002043 * 50 * 50) + 39.085612 = 27.4\text{mL}O_2/\text{min}/\text{kg}$$

$$\text{Relative } \dot{V}O_{2\text{peak}} = (-0.131566 * 60) + (-0.002043 * 60 * 60) + 39.085612 = 23.8\text{mL}O_2/\text{min}/\text{kg}$$

mLO<sub>2</sub>/min/kg (table 4.6 \*) in the present study and 36.5 mLO<sub>2</sub>/min/kg [34]<sup>†</sup> in the SHIP study. Reference values from FRIEND study showed higher values in the youngest age group but also a rapid decline and lower values in all other age groups. The median for 40-year-old men was only 27.1 mLO<sub>2</sub>/min/kg. Some reasons for the differences between the reference values are discussed below.

A number of past reference values excluded smokers and obese subjects from reference values [34, 26]. The exclusion of smokers was also recommended in the systematic review by Paap & Takken (2014) [50]. Therefore, the present study included subgroup analyses by excluding smokers and obese subjects from the analyses (supplementary tables and figures [49]). The reference values were considerably higher when smokers were excluded from the sample. Median relative  $\dot{V}O_{2\text{peak}}$  in 45-year-old men was 37.3 mLO<sub>2</sub>/min/kg when smokers and obese subjects were excluded and 35.8 mLO<sub>2</sub>/min/kg when smokers and obese subjects remained in the reference data (supplementary tables and figures [49]<sup>‡</sup>).

Aside from the exclusion of smokers and obese subjects, subgroup analyses were also performed for the inclusion of all subjects from Rüdeshheim, Frankfurt and Munich. In that case, the reference values were consistent with the primary results (supplementary tables and figures [49]).

Past studies have shown that  $\dot{V}O_{2\text{peak}}$  can be increased by physical activity and exercise. The average effect was estimated at 5.4 mLO<sub>2</sub>/min/kg in men and 3.2 mLO<sub>2</sub>/min/kg in women by Lin et al. (2015) [39]. When those estimations are compared to the present nomograms, the mean effect of structured exercise on  $\dot{V}O_{2\text{peak}}$  of 40-year old men and women may be equal to the difference of median and 0.7 quantile. This can be interpreted as a 20% increase relative to the reference population.

---

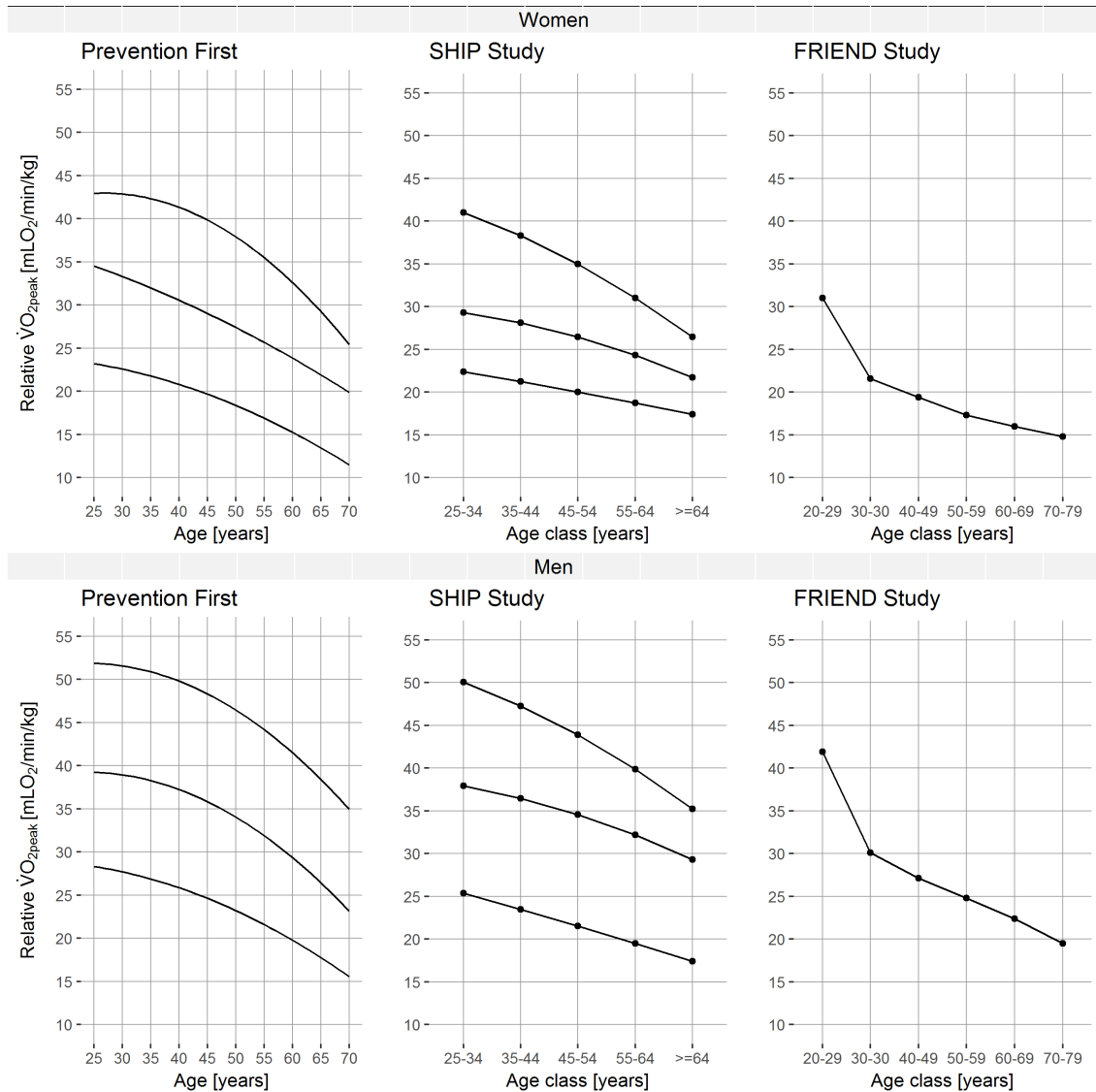

$$*\text{Relative}\dot{V}O_{2\text{peak}} = (0.355 * 40) + (-0.0075 * 40 * 40) + 35.04 = 37.2\text{mLO}_2/\text{min}/\text{kg}$$

$$^\dagger\text{Relative}\dot{V}O_{2\text{peak}} = 47.7565 + (-0.9880 * 2) + (-0.2356 * 2 * 2) + (-8.8697 * 1) + (2.3597 * 0) + (-2.0308 * 2 * 0) + (-3.7405 * 1 * 0) + (0.2512 * 2 * 1) + (1.3797 * 2 * 1 * 0) = 36.5, \text{ where male sex was denoted by 1, the age class of 35 to 44 years as 2 and BMI} < 25 \text{ kg/m}^2 \text{ as 0.}$$

<sup>‡</sup>men:

$$\text{Relative}\dot{V}O_{2\text{peak}} = (0.37 * 45) + (-0.0075 * 45 * 45) + 35.8375 = 37.3\text{mLO}_2/\text{min}/\text{kg}$$

$$\text{Relative}\dot{V}O_{2\text{peak}} = (0.355 * 45) + (-0.0075 * 45 * 45) + 35.04 = 35.8\text{mLO}_2/\text{min}/\text{kg}$$



**Figure 5.1:** Comparison of reference values for relative  $\dot{V}O_{2peak}$ .

**Note:** SHIP (cycle ergometry-based): 95%, 50% and 5% percentiles plotted; [34]; FRIEND (cycle ergometry-based) 50% percentile plotted (the study did not provide 5% and 95% percentiles) [31]; Prevention First (cycle ergometry-based) 95%, 50% and 5% percentiles plotted.

## 5.2 Study population and representativity

### Study population and data acquisition

Several reference values for peak oxygen uptake have been published using different study designs and populations [50, 30, 31, 53]. In the rating system of the systematic review by Paap & Takken (2014) [50], four items dealt with study design and sampling. The authors assigned one point each for i) community-based sampling (27/35 of the reviewed studies, 77%), ii) including a randomization process (1/35 of the reviewed studies, 3%), iii) a large sample size with uniform distribution of age and sex (2/35 of the reviewed

studies, 6%), as well as iv) a prospective study design (23/35 of the reviewed studies, 66%). The present study was based on register data, and no population-based sampling process was conducted to select the participants of the study. The numbers of women ( $n = 3,291$ ) were lower compared to men ( $n = 6,063$ ) and the distribution of age showed low numbers of cases in marginal age groups (figs. A.1 and A.2). This effect of sparse age groups was less prevalent in population-based studies. In the underlying training dataset (Frankfurt/Rüdesheim), only nine women were observed in the age class 25 to 29 years. In the study of Koch et al. (2009) [34] the age class with the least number of cases was the class of women above 75 years. 23 participants were recorded in this class. In other well-designed population-based studies, the observations were also rather uniformly distributed over the age groups [17, 26]. Furthermore, the problem of sparsity in marginal age groups is smaller in recent register-based studies [30, 31]. Due to the relatively low number of cases in marginal age groups, estimations for these groups were less precise in the present study. This was represented by wide confidence intervals at both ends of the regression curves (figs. 4.3 and 4.4). However, age groups from 30 to 64 years showed higher numbers of observations compared to most other studies leading to more precise estimations and narrow confidence intervals. The reference values for these age groups are particularly precise compared to other studies.

### Reference population

When the data collection in studies was not based on a population-based sampling, it is important to know the general setup of the data acquisition. For this reason, an outline is provided to explain why CPET was performed in the first place. The reference values are not appropriate for an individual when the reference population is distinctly different from the individual's population. This becomes particularly obvious when reference values of the present study are compared with data from Kaminski et al. (2015, 2017) [30, 31] or long-established reference values from Wasserman (cited by [70]). The participants of Kaminski et al. conducted CPET prior to exercise programmes or research studies, and Wasserman et al. acquired former shipyard workers. Both populations were different from the sample of the underlying study which was primarily composed of German workers with a predominantly sedentary working environment. Men at the age of 45 years had a median relative  $\dot{V}O_{2\text{peak}}$  of 35.8 mL $O_2$ /min/kg in the present study (table 4.6\*), 27.1 mL $O_2$ /min/kg in the study of Kaminski et al. (2015) [30] and a mean of 34.0 mL $O_2$ /min/kg<sup>†</sup> for a male at 180 cm and ideal weight using Wasserman's equation [70]. Reference values from Wasserman were higher compared to Kaminski et al. The reason for this might be that shipyard workers had a physically active working environment.

---

\*Relative $\dot{V}O_{2\text{peak}} = (0.355 * 45\text{years}) + (-0.0075 * 45\text{years} * 45\text{years}) + 35.04 = 35.8\text{mL}O_2/\text{min}/\text{kg}$

<sup>†</sup>Ideal weight =  $0.79 * 180\text{cm} - 60.7 = 81.5\text{kg}$

Peak  $\dot{V}O_2 = 0.0337 * 180\text{cm} - 0.000165 * 45\text{years} * 180\text{cm} - 1.963 + 0.006 * (81.5\text{kg} - 81.5\text{kg}) = 2.77\text{LO}_2/\text{min}$

Relative $\dot{V}O_{2\text{peak}} = \frac{2.77 * 1000}{81.5\text{kg}} = 34.0\text{mL}O_2/\text{min}/\text{kg}$ , see [70, p. 158]

In Kaminski et al., the participants conducted CPET prior to exercise programmes. This may suggest that there was an indication for exercise prescriptions to counteract sedentary lifestyles or other risk factors such as smoking or obesity. Furthermore, there were more than 20 years between the data collection of both reference values, and both Kaminski's and Wasserman's reference values were based on a US-American sample, which might impact the results. As the present study was based on German workers with a primarily sedentary working environment, distinct differences from the reference values mentioned above were observed. Hence, reference values based on US-American samples should be applied with caution to German participants.

### Representativity

The sample of the present study was compared to 'Studie zur Gesundheit Erwachsener in Deutschland (DEGS1)' [19], which was based on a population-based sampling process and was representative for the German population. Five binary characteristics were selected that were published in DEGS1 and were measured equally in the underlying data. All five characteristics can be characterised as cardiovascular risk factors. DEGS1 results were age-standardised using the German population of 2011 as standard [38, 43, 48] and the presented results were standardised according to DEGS1 (section 3.4). Men and women showed significant differences from DEGS1 results (table 4.4). In men, 3/5 characteristics were significantly different in the present study compared to DEGS1. Smoking, overweight and obese participants were less prevalent in the present study. In women, 4/5 characteristics were significantly different, and only ex-smokers did not differ significantly. The differences were particularly large in regard to obesity, where 23.9% of the women in DEGS1 but only 12.1% in the underlying data met the criterion. A large difference was also present in the smoking status of men. 26.1% of men in DEGS1 were smokers but only 14.7% of the underlying data. The risk factors were less prevalent in the present study compared to DEGS1. This might indicate selection bias. It is likely that the present study sample consisted of persons with a particularly healthy lifestyle. The reason for this could be that the participation in the workplace health promotion programmes was voluntary and persons who participated were more prone to live a healthy lifestyle. The presented reference values might be assumed to be higher compared to the German population and also probably higher than in the population of German workers with a sedentary working environment because it is likely that there was a selection of particularly healthy participants.

Koch et al. (2009) [34] published reference values using a population-based sampling but assumed that selection bias was present in their study, nevertheless. This would also have led to higher reference values compared to the whole population. 45-year-old women had a median relative  $\dot{V}O_{2\text{peak}}$  of 29.0 mL $O_2$ /min/kg in the present study (table 4.6\*) and

---

\*Relative $\dot{V}O_{2\text{peak}}$  =  $(-0.131566 * 45\text{years}) + (-0.002043 * 45\text{years} * 45\text{years}) + 39.085612 = 29.0\text{mL}O_2/\text{min}/\text{kg}$

26.4 mL $O_2$ /min/kg\* in the SHIP study by Koch et al. However, it has to be noted that age was modelled in 10-year age classes by Koch et al. Another reason for the differences was that smokers and obese subjects were excluded by Koch et al. As peak oxygen uptake is associated with smoking status and obesity, the exclusion of such participants was likely to increase the reference values. However, reference values by Koch et al. were slightly lower, overall. This might be due to a more severe selection of study subjects in the present study or also regional differences. Koch et al. collected the data in north-eastern Germany, the present study in regions of southern Germany. The time between publication by Koch et al. (2009) and the present study was not assumed to have a strong impact as data collection at Prevention First<sup>®</sup> started in 2001.

Itoh et al. (2013) [26] produced reference values based on a Japanese population. Compared to the present study, 45-year-old women showed an even higher  $\dot{V}O_{2peak}$  of 31.3 mL $O_2$ /min/kg<sup>†</sup>. However, this study performed even more rigorous exclusion than Koch et al. [34]. In summary, the results of the present study were more comparable to the results based on the German population [34], rather than the US-American or Japanese reference values [31, 70]. Reference values from different nations and settings showed large differences. This has to be considered when reference values are used to interpret the  $\dot{V}O_{2peak}$  of a CPET participant.

### Exclusion criteria

The exclusion criteria were different in past studies, and there is no clear guideline regarding which participants should be excluded or not. Paap & Takken (2014) suggested to exclude smokers and obese participants as these cases do not represent a healthy population. However, if this rationale is used, participants with other conditions such as arterial hypertension or dyslipoproteinemia should also be excluded. This leads to an artificial study population and is likely to increase the reference values because healthy persons probably have a higher  $\dot{V}O_{2peak}$ . Some of the past studies performed rigorous exclusion of smokers, obese individuals and such with other risk factors [34, 50]. Itoh et al. (2013) [26] also excluded individuals with abnormal blood pressure at rest and those who exercised more than twice a week or were not 145 to 190 cm tall. Other studies [30, 31] did not provide information on how they treated smokers and obese subjects.

The present study did not exclude individuals with risk factors from the sample for the primary results. However, subgroup analyses were performed with exclusion of smokers and obese subjects (supplementary tables and figures, [49]). As expected, reference values were slightly higher when smokers were excluded (section 5.1). However, by keeping participants with risk factors in the sample, this study aimed to represent the entire Ger-

---

\*Relative $\dot{V}O_{2peak} = 47.7565 + (-0.9880 * 3) + (-0.2356 * 3 * 3) + (-8.8697 * 2) + (2.3597 * 0) + (-2.0308 * 3 * 0) + (-3.7405 * 2 * 0) + (0.2512 * 3 * 2) + (1.3797 * 3 * 2 * 0) = 26.4\text{mL}O_2/\text{min}/\text{kg}$ , where women were coded as 2, age group of 45 to 54 as 3 and BMI  $\geq$  as 0 [34].

†Relative $\dot{V}O_{2peak} = 61.06 + (-0.510 * 45\text{years}) + (-20.4 * 1) + (0.301 * 1 * 45\text{years}) = 31.3\text{mL}O_2/\text{min}/\text{kg}$  (29 mL $O_2$ /min/kg in the present study), where female was coded as 1 [26].



man population. This might increase the generalisability of the present reference values for the German population. Furthermore, interactive subgroup analyses are possible at [www.uks.eu/vo2peak](http://www.uks.eu/vo2peak).

### Further population information

As mentioned above, differences in peak oxygen uptake between countries were observed. Furthermore, it is valuable to report the level of physical activity of the study population. In the systematic review by Paap & Takken (2014) [50], physical activity was reported by 20/35 (57%) of the reviewed studies. Itoh (2013) [26], for example, excluded participants who stated that they exercised more than twice a week. Edvardsen et al. (2013) [17] recorded physical activity using activity accelerometers and questionnaires. In the present study, no information on physical activity level was recorded, which is why the level of physical activity of the present sample remained unknown.

Paap & Takken (2014) [50] emphasised that the ethnicity of the sample should be described. Few past studies have provided information on the ethnic composition of their sample or have taken this information into account by stratification or exclusion of subjects. Caucasian, Japanese and Scandinavian populations were mostly used as reference population. According to Paap & Takken, Asian, Middle-Eastern, African and South-American populations have been underrepresented so far. The underlying data for the present study did not provide information on ethnicity but it can be assumed that largely Caucasian subjects were recorded.

To sum up, the present registry-based study provided one of the largest study samples in its field. Most of the participants were white-collar workers with a primarily sedentary working environment. As in other registry-based studies and also in population-based studies, it can be assumed that a selection of particularly healthy participants was present. Compared to other studies, the present reference values were relatively high. Furthermore, there was no exclusion of participants with cardiovascular risk factors, but an exclusion can be performed interactively by using the online web application ([www.uks.eu/vo2peak](http://www.uks.eu/vo2peak)) or by using the appropriate nomograms from the subgroup analyses (supplementary tables and figures, [49]).

### 5.3 Exercise test modalities

Some CPET test modalities such as the choice of ergometer (cycle or treadmill) and the time of termination of CPET have a substantial impact on the  $\dot{V}O_{2\text{peak}}$ . The CPET modalities of the present study and the impact on the test results are summarised below.

### Type of ergometer

Most exercise test protocols are based on an incremental work rate (section 2.3). The most common devices for conducting test protocols are cycle ergometers or treadmills. However, the choice of ergometer type has a strong impact on the  $\dot{V}O_{2\text{peak}}$ , which is measured in the exercise test. The differences between cycle ergometry-based and treadmill-based  $\dot{V}O_{2\text{peak}}$  can be observed by comparing the studies of Kaminski et al. (2015 and 2017) [30, 31]. Both studies were based on the same registry (Fitness Registry and the Importance of Exercise National Database, “FRIEND”) but used either cycle ergometers or treadmills to assess  $\dot{V}O_{2\text{peak}}$ . 35-year-old men showed a median relative  $\dot{V}O_{2\text{peak}}$  of 42.4 mL $O_2$ /min/kg using treadmill ergometer [30] and 30.1 mL $O_2$ /min/kg using cycle ergometer [31]. This difference is even stronger than the difference between sexes and also considerably higher than the 5 to 10% increase stated by ATA/ACCP guidelines [1, p. 218]. Medians in 35-year-old women were 30.2 mL $O_2$ /min/kg using treadmill and 21.6 mL $O_2$ /min/kg using cycle ergometer. A similar effect was observed by Itoh et al. (2013) [26]. The relative  $\dot{V}O_{2\text{peak}}$  for a 35-year old man using treadmill and cycle ergometer was expected to be 43.2 mL $O_2$ /min/kg\* and 32.7 mL $O_2$ /min/kg†, respectively. In 35-year-old women, the expected values for treadmill and cycle ergometers were 33.4 mL $O_2$ /min/kg‡ and 28.3 mL $O_2$ /min/kg§. In the present study, the type of ergometer was not recorded in the entire dataset. Prevention First® provided both options - cycle or treadmill ergometer - but it was assumed that all tests that were recorded for the present study were performed using cycle ergometer. To validate this assumption, the type of ergometer was recorded retrospectively in the random sample of 252 observations. 249/249 tests were recorded using cycle ergometer, and in three tests, no information on the type of ergometer has been recorded. This provided further evidence that exclusively cycle ergometers were used in this study. However, this information is based on a random sample and was not recorded for the entire dataset. If tests using treadmill ergometers were present, it would be likely that the presented reference values were falsely high.

### Measures of maximal effort

Another CPET test modality that impacts the measured  $\dot{V}O_{2\text{peak}}$  is the time of termination of the exercise test. Ideally, the participant has to perform until the true peak volitional work rate is achieved to capture a  $\dot{V}O_{2\text{peak}}$  that is close to the true peak volitional oxygen uptake. However, this is dependent on the participant’s motivation and effort. To assess the participants’ effort, several techniques were suggested: i) the relationship between increasing work rate and increasing oxygen uptake could be considered

---


$$*\text{Relative}\dot{V}O_{2\text{peak}} = 61.07 + (-0.510 * 35) + (-20.4 * 0) + (0.301 * 0 * 35) = 43.2\text{mL}O_2/\text{min}/\text{kg}$$

$$\dagger\text{Relative}\dot{V}O_{2\text{peak}} = 42.05 + (-0.268 * 35) + (-7.22 * 0) + (0.0811 * 0 * 35) = 32.7\text{mL}O_2/\text{min}/\text{kg}$$

$$\ddagger\text{Relative}\dot{V}O_{2\text{peak}} = 61.07 + (-0.510 * 35) + (-20.4 * 1) + (0.301 * 1 * 35) = 33.4\text{mL}O_2/\text{min}/\text{kg}$$

$$\S\text{Relative}\dot{V}O_{2\text{peak}} = 42.05 + (-0.268 * 35) + (-7.22 * 1) + (0.0811 * 1 * 35) = 28.3\text{mL}O_2/\text{min}/\text{kg} [26]$$

(section 2.3). If a levelling-off (that is, smaller increases of oxygen uptake per increase of work) appears, the recorded  $\dot{V}O_2$  can be assumed to be close to the true physiological peak oxygen uptake. This relationship was not used extensively in past studies as a  $\dot{V}O_2$  increase of  $\leq 150$  mL $O_2$ /min is rarely observed [59]. Other measures of the participant's effort were used more frequently including ii) respiratory exchange ratio ( $RER = \frac{\dot{V}CO_2}{\dot{V}O_2}$ ), iii) capillary lactate levels and iv) maximal heart rate. An RER of  $\geq 1.1$ , capillary lactate levels of  $\geq 8$  mmol/L or heart rate of  $\geq 90\%$  of the age-predicted maximal heart (APMHR) rate should be achieved [59, 45] to ascertain adequate effort. The criteria for maximal effort were different in past studies. For example, Kaminsky et al. used an RER of  $\geq 1.1$  for CPET using cycle ergometers [31] and RER of  $\geq 1.0$  for CPET using treadmill ergometers. Itoh et al. (2013) [26] defined RER of  $< 1.0$  as poor effort.

In most prospectively-designed studies, participants with poor effort were excluded from the sample and were not used for calculation of reference values. In the present study, the measures of maximal effort were not recorded in the entire database but were acquired manually for the random sample (section 4.4). Measures of good exertion were defined as mentioned above (RER  $\geq 1.1$  or capillary lactate levels  $\geq 8$  mmol/L or maximal heart rate  $\geq 90\%$  of APMHR). 97% (95% CI 94% to 99%) of the participants in the random sample showed adequate effort. However, as measures of effort were not recorded in the database, participants with poor effort could not be excluded from the analysis. Therefore an estimated 3% of the exercise tests were not valid. Assuming that patients with poor effort showed lower  $\dot{V}O_2$  at the time of termination, the here presented reference values were reduced by keeping these participants in the sample. To see if this effect was present in the random sample, median regressions were calculated in the random sample for i) inclusion and ii) exclusion of subjects with poor effort. Furthermore, a multivariable median regression including a binary predictor variable for exertion (yes/no) and an interaction term (age\*exertion) was calculated to see if the effect was significant. However, regression lines did not show a large deviation in the sample (fig. 4.11) and exertion, as well as interaction terms in men, were not significant (table 4.9). In women, only three observations with poor effort were available leading to imprecise regression estimates. To conclude, it can be stated that the numbers of invalid exercise test results were low in the present study. Based on the random sample, it is likely that the effect on the presented reference values of those tests was low.

### Exercise test protocol

In incremental exercise tests, the work rate is low at the beginning and increases over time. There are plenty of protocols for incremental exercise tests including ramp protocols or multistage protocols (section 2.3). In past studies, mostly one single protocol was used throughout the data acquisition. Itoh et al. (2013) [26] used ramp protocols and different increments of work per time. They described that work rate at peak exercise was higher if the increments were higher. However, there was no impact of the different increments

on the assessed  $\dot{V}O_{2\text{peak}}$ . This finding is similar to earlier studies which did not find an association between protocol and peak oxygen uptake [6, 47]. In the present study, the type of exercise test protocol was not recorded in the entire database but was assessed retrospectively in the random sample. Multistage protocols (130/243 times) were used more often compared to ramp protocols (113/243 times). However, based on the literature, it can be assumed that there is no substantial effect of the exercise test protocol on the measured value of  $\dot{V}O_{2\text{peak}}$ .

To conclude, it can be stated that the test modalities were not recorded in the entire study dataset and had to be estimated based on a random sample. The choice of ergometer has a substantial impact on reference values. In the random sample, all exercise tests were performed using cycle ergometer. Participants with poor effort remained in the dataset and contributed to the reference values. However, as there was an estimated proportion of 97% participants with adequate effort, the impact of this shortcoming was low.

**Table 5.1:** Quality assessment of the measurement of  $\dot{V}O_{2\text{peak}}$ .

Characteristic	Value
<b>Study design</b>	
- Prospective design	0
- Proper randomization	0
- Community-based sampling of the study population	0
- The number of study subjects is at least as high as calculated in sample size estimation	1
<b>Characteristics of the exercise tests</b>	
- Measurement of gas exchange data and $\dot{V}O_{2\text{peak}}$ is averaged over time to avoid noise (preferably $\geq 30$ seconds intervals)	0
- CPET was performed using breath-by-breath or mixing chamber analysis according to ATS/ACCP guidelines	1
- Quality control was performed according to ATS/ACCP guidelines	1
<b>Important background reported</b>	
- Level of physical activity reported	0
- Exercise testing protocol described	1
<b>Data analysis and reporting</b>	
- External validation of the statistical model	1
- Adequate fitting of the regression model was performed	1
- Analysis was done stratified by racial group	0
- Smokers were excluded	1
- Confidence limits were given for descriptive statistics	1
<b>Sum</b>	<b>8/14</b>

**Note:** This rating score was modified according to Paap & Takken (2014) [50]. The overall quality score of the present study's measurement was 8/14. The quality of measurement can therefore be rated as "medium quality" (section 2.4).

## 5.4 Statistical methods

### Modelling maximal oxygen uptake

Past studies of reference values for  $\dot{V}O_{2\text{peak}}$  used different statistical approaches. The most common choices were i) to use age as categorical variable and to calculate empirical

quantiles within each age class [53, 30, 31] or ii) to perform linear regression with age as predictor variable [26, 17]. iii) Koch et al. [34] used second-degree polynomial quantile regression. Further well-accepted reference values which are, however, not related with  $\dot{V}O_{2\text{peak}}$  are the WHO's nomograms for child growth. The authors of these reference values used spline smoothing [75].

Those four approaches were also considered for the present analysis. However, it was decided that some approaches were less adequate for reference values. First, linear regression would not be sufficient to represent the distribution of  $\dot{V}O_{2\text{peak}}$  for a given age because only the conditional mean would be estimated and the large variation of CRF due to physical activity and interpersonal factors would not be represented. Secondly, in order to calculate quantiles, age had to be measured in age classes and could not be treated as a continuous variable. Past studies often treated age as a categorical variable using ten-year age classes [30, 31, 34, 53]. In the present study, substantial differences were observed within a 10-year age class. For women, median relative  $\dot{V}O_{2\text{peak}}$  was 27.4 mL $O_2$ /min/kg\* and 24.2 mL $O_2$ /min/kg† for 50 and 59 years, respectively. This leads to a difference of 3.2 mL $O_2$ /min/kg, which is approximately equivalent to the difference between median and 0.7 quantile in women of that age group the present study.

As the approaches mentioned above would have yielded a large imprecision, this study estimated quantiles based on age as a continuous predictor. To achieve this, quantile regression was selected. The decision of how to treat age as a predictor (linear, polynomial or using spline smoothing) was based on AIC as well as apparent and external validation. Linear quantile regression performed worst and was hence excluded. The most accurate regression fits were observed for spline regression. However, this study used polynomial quantile regression instead of spline models because polynomial quantile regression models are less complex than spline models. The number of regression coefficients was equal in all quantile regressions of this study. Therefore, it was possible to display the regression coefficients in tables (section 4.3).

### Validation of reference values

Paap & Takken (2014) [50] suggested to do external validation of reference values for  $\dot{V}O_{2\text{peak}}$ . However, none of the reviewed studies since 2010 conducted validation of the reference values in an external population.

The present study aimed to perform external validation by splitting the dataset by study centre. Data from Rüdeseim and Frankfurt were used as training data and data from Munich were used as external validation data. Validation was performed as i) apparent validation using the training data to fit regression models and also for validation and ii) external validation using the training data to fit regression models and external validation data for validation.

---

\*Relative $\dot{V}O_{2\text{peak}} = (-0.131566 * 50) + (-0.002043 * 50 * 50) + 39.085612 = 27.4\text{mL}O_2/\text{min}/\text{kg}$

†Relative $\dot{V}O_{2\text{peak}} = (-0.131566 * 59) + (-0.002043 * 59 * 59) + 39.085612 = 24.2\text{mL}O_2/\text{min}/\text{kg}$ , table 4.6

Validation according to established approaches was not possible for several reasons. This should be outlined based on the example of validation of a linear regression model. In linear regression, the conditional mean of the continuous dependent variable  $y$  is estimated as the expected value (e.g. the best prediction) for an individual  $i$  conditional to the values in a set of predictor variables of that individual  $i$ . Validation can easily be performed using the predicted value  $\hat{y}_i$  and the observed value  $y_i$  to plot a scatter plot. In this calibration data of  $\hat{y}_i$  and  $y_i$ , a linear regression can be performed and coefficients for the intercept close to 0 and for the slope close to 1 are desirable. This technique applies to apparent as well as external validation using the training data or external validation data to obtain  $y_i$ , respectively. In the present study, however, it was not desired to get the best prediction of  $\dot{V}O_{2\text{peak}}$  for an individual but to derive quantiles based on age as a continuous predictor. As quantiles could not be calculated for an individual, aggregated empirical data had to be used. Therefore, the observed age was aggregated into age classes. To be able to compare the predicted quantiles with aggregated empirical quantiles, age has also been aggregated as the predictor in quantile regression models to obtain estimated quantiles. This aggregation was only done for model validation but not for the calculation of nomograms and reference values. Therefore, there was a limitation of not using the same models in nomograms and for model validation. The true validity, hence, could only be estimated based on this approach.

This validation process showed that the regression models fitted the empirical data well. As expected, better validity was observed in apparent validation compared to external validation. In men aged 25 to 29 years, the predicted  $\dot{V}O_{2\text{peak}}$  was higher than the observed  $\dot{V}O_{2\text{peak}}$ . This was observed in apparent and external validation. However, this age group was sparse, including only 22 observations (figs. 4.7 and 4.8). Regression coefficients also showed adequate validity with most intercepts close to zero and slope as well as R squared close to one. A larger deviation from ideal values was observed for  $\dot{V}O_{2\text{peak}}$  of men. Here, intercept was 5.9 (95% CI 1.4 to 9.9), slope 0.8 (95% CI 0.7 to 1.0) and R squared 0.9 (95% CI 0.9 to 1.0). Significant differences from intercept = 0 and slope = 1 were especially observed in men. All coefficients of men in apparent validation showed significant differences from the ideal coefficients. Apparent validation of women, on the other hand, showed accurate results. In external validation, the results were more accurate in men but less accurate in women.

Nevertheless, although there were significant differences in calibration plots, the regression fits seemed quite accurate, especially in the age classes from 30 to 64 years (figs. 4.7 to 4.10). The approach for model validation of the present study can be assumed to be particularly rigorous and might be accountable for some of the deviation from optimal values. As mentioned above, observed quantiles were based on  $\dot{V}O_{2\text{peak}}$  which was calculated within age classes. However, the numbers of observations in age classes were different across the age classes. The age classes 25 to 29 years and 65 to 69 years were relatively sparse. The estimation of quantiles was less precise in these age classes. Therefore, the differences between predicted and observed values were relatively large. As each

age class weighted the same in the calibration plot, regardless of the number of cases within the age class, the presented validation underestimated the validity of the present quantile regressions.

To sum up, apparent as well as external validation were rarely performed in past studies in this field. Still, validation adds some valuable information on how adequate and valid reference values are when they are compared to other individuals besides the study population. The reference values presented by this study showed adequate regression fits but also significant differences from optimal values in some cases. As different approaches were used for the calculation of reference values and model validation, it was only possible to get an estimation of how valid the presented quantile curves were.

### Number of observations

For the development of reference values, high numbers of cases and a uniform distribution over all age classes are desirable to achieve precise estimations and narrow confidence intervals [50]. Overall, the present study was based on one of the largest datasets that were reported in this field, so far. Using the search term from Paap & Takken et al. (2014) [50] in the search engine Pubmed (<https://www.ncbi.nlm.nih.gov/pubmed/>) did not yield any study with higher numbers at the time of preparation of this manuscript. Only the sample of reference values by the Cooper Institute (cited by [53, p. 88]) was larger. However, it has to be noted that the total number of cases in the present study were large, but the cases were not uniformly distributed across the age classes. This led to more precise estimations in age classes 30 to 64 years, but also to less precise estimations in marginal classes (< 25 years and  $\geq 65$  years). The overlapping confidence intervals visualised the less precise estimations in the nomograms (figs. 4.3 and 4.4).

## 5.5 Multivariable analyses

In addition to reference values, exploratory analyses were performed using multivariable quantile regression for the quantiles 0.25, 0.5, and 0.75. AIC-based forward and backward variable selection was used to select predictors for  $\dot{V}O_{2\text{peak}}$ . The goodness of fit, which was assessed using pseudo R squared (section 3.6), showed high values of  $\approx 90\%$  for absolute and lower values of  $\approx 20\%$  for relative  $\dot{V}O_{2\text{peak}}$ . This large deviation of R squared in absolute or relative  $\dot{V}O_{2\text{peak}}$  was likely to be explained by the presence of the variable “overweight” as an independent variable. Being overweight showed a strong positive association in each regression where absolute  $\dot{V}O_{2\text{peak}}$  was used and a strong negative association in each regression where relative  $\dot{V}O_{2\text{peak}}$  was used. This was likely due to the fact that relative  $\dot{V}O_{2\text{peak}}$  was defined as absolute  $\dot{V}O_{2\text{peak}}$  relative to body weight\*. The findings in the present study showed that  $\dot{V}O_{2\text{peak}}$  increases with body weight, that is, a

---

\*Absolute  $\dot{V}O_{2\text{peak}}$  measured in  $LO_2/\text{min}$   
Relative  $\dot{V}O_{2\text{peak}}$  measured in  $mLO_2/\text{min}/\text{kg}$ .

heavier person is able to consume more oxygen. On the other hand, obese subjects seem to have lower CRF relative to their body weight compared to non-obese subjects. This effect has already been described by Edvardsen et al. (2013) [17]. It was the largest effect over all predictors and was even stronger than the effect of tobacco smoking. In addition to body weight, triglycerides and in some cases higher blood glucose levels or diabetes mellitus were associated with lower peak oxygen uptake. HDL on the other hand, was higher in men with a good CRF. However, this association was not present in women. A comprehensive meta-analysis by Lin et al. (2015) [39] found associations between physical activity, CRF and lipids as well as lipoprotein markers. The authors suggested that physical activity increased CRF and HDL cholesterol and decreased triglycerides. In this meta-analysis, physical activity led to decreased triglycerides by  $-5.3$  mg/dL (95% CI  $-10.6$  to  $-0.9$ ). Accordingly, in the present study men with triglycerides  $\geq 150$  mg/dL showed a decreased median of relative  $\dot{V}O_{2\text{peak}}$  by  $-1.8$  mL $O_2$ /min/kg (95% CI  $-2.2$  to  $-1.3$ ) compared to men with lower triglyceride levels (table 4.10).

Another strong association with  $\dot{V}O_{2\text{peak}}$  was observed for tobacco smoking. Median relative  $\dot{V}O_{2\text{peak}}$  in women was decreased by  $1.3$  mL $O_2$ /min/kg (95% CI  $1.9$  to  $0.8$ , table 4.11). This effect was even stronger in men ( $-2.0$  mL $O_2$ /min/kg, 95% CI  $-2.5$  to  $-1.5$ , table 4.10) and was observed across all regression quantiles. Ex-smokers, on the other hand, did not show significantly decreased CRF. However, there was no measure of how long non-smoking had been sustained. Tobacco smoking has been widely considered in past studies. Smokers were excluded from some analyses because they were assumed to distort the results and because the samples were intended to represent a healthy population [34, 26]. Therefore, smoking and obesity were considered in multivariable analyses and in subgroup analyses of this study ([www.uks.eu/vo2peak](http://www.uks.eu/vo2peak), supplementary tables and figures [49]).

Diabetes mellitus was associated with lower  $\dot{V}O_{2\text{peak}}$  especially in men. This effect was also well-described in past studies such as the meta-analysis by Zaccardi et al. (2015) [77]. The hazard ratio to develop diabetes mellitus was 0.93 per 1-MET increase of CRF. In contrast to this result, diabetes mellitus was used as an independent variable in the present study. Men with diabetes showed a lower  $\dot{V}O_{2\text{peak}}$  ( $-2.0$  mL $O_2$ /min/kg, 95% CI  $-3.1$  to  $-0.9$ ). However, it was interesting to see that this association was not present in women of the present study. This might be partially due to the lower numbers of women in the dataset but future studies might consider sex as effect modifier or report stratified results.

Furthermore, as mentioned before, an inverse association of  $\dot{V}O_{2\text{peak}}$  and age was observed. This association was well-described before [23, 50] and was present in this analysis even when the effect was adjusted for other predictors.

A positive association was observed between  $\dot{V}O_{2\text{peak}}$  and body height. However, as body height cannot be influenced, this result seemed to be of minor interest.

Weaker associations were observed for blood pressure and LDL cholesterol.

To conclude, explorative multivariable results yielded some strong inverse associations



between  $\dot{V}O_{2\text{peak}}$  and body composition as well as tobacco smoking. Smaller effects were observed for other blood lipids, hypertension and diabetes mellitus. Even though those associations were based on explorative analyses, the results confirmed results from past studies [39, 50].

## 5.6 Summary of strengths and limitations

This section summarises the strengths and limitations of the present study compared to other studies.

One important strength was the large amount of data, which included a sample size of more than 9,000 individuals. Based on this data, precise reference values could be calculated, especially for age groups from 30 to 64 years. Secondly, the data analysis was done using age as a continuous predictor and using quantile regression to plot nomograms for a number of quantiles. The nomograms were desired to be helpful in clinical practice when the results of CPET are interpreted. Thirdly, in contrast to past studies, validation of the reference values was performed in an external population in order to get information on the external validity of the reference values. The validity was shown to be adequate. Finally, CPET was performed by experienced exercise test instructors of a quality network for primary preventive health screening institutions. Finally, an interactive web application was created to facilitate the usage of the reference values in clinical practice ([www.uks.eu/vo2peak](http://www.uks.eu/vo2peak)).

On the other hand, the present study also showed some limitations. Firstly, the numbers of observations were low in the age classes of  $< 30$  years and  $\geq 64$  years. Reference values for those ages were less precise. Furthermore, the study population was not drawn by a prospective population-based sampling process including a randomisation process. There were also large differences from the German population, indicating a selection of healthy participants. The quality of the data acquisition according to Paap & Takken (2014) [50] was only “moderate” (table 5.1) and the exercise test modalities were not documented in the main data file. Therefore, a small proportion of participants who did not continue the exercise test until exertion remained in the data. Another limitation was the validation of regression models as it was performed in the present study. The approach of using quantile regressions necessitated aggregation of the data, and therefore, age was treated as a continuous predictor for reference values and as a categorical predictor for model validation.

## 6 Conclusions and clinical implications

The present study provides reference values for peak oxygen uptake. They can be used for participants of cycle ergometry-based cardiopulmonary exercise tests who are 25 to 69 years old and who are part of a population that is similar to this study.

Reference values are essential for cardiopulmonary exercise testing because they allow the classification of individual results based on the reference population. It is only possible to classify the individual cardiorespiratory fitness as “high” or “low” if the result is compared with sex-specific and age-specific reference values from a comparable population. If the cardiorespiratory fitness of a person is low compared to other persons of the same sex and age, this person has an increased risk to die prematurely. From a public health perspective, it is therefore particularly valuable to increase the cardiorespiratory fitness of the population. Preventive health screenings are an appropriate setting to achieve this goal. However, it is essential that physicians consider cardiorespiratory fitness as a risk factor in addition to well-known risk factors such as tobacco smoking or arterial hypertension.

The present study adds important value to this field. The study population yielded one of the most extensive samples that has been published so far. The participants of this study were predominantly German workers with a sedentary lifestyle. A large proportion of the populations of industrialised countries have similar working environments.

This study aimed to facilitate the interpretation and access of the presented reference values by constructing nomograms and an interactive web application ([www.uks.eu/vo2peak](http://www.uks.eu/vo2peak)). Depending on which population is desired as the reference population, subgroup analyses can be performed interactively. Furthermore, this study allows physicians to evaluate the cardiorespiratory fitness of an exercise test participant precisely. Past studies often have used 10-year age classes or linear regression, but these approaches yielded substantial imprecision. This study presents the reference values as percentile values, which is an accurate representation of the inter-individual variability.

The evaluation of the validity and the generalisability is crucial for the interpretation of reference values. External validation was suggested by Paap & Takken (2014) [50], but the validity and the representativity were rarely assessed in previous studies. Although these two analyses revealed some shortcomings of the present sample, they are valuable for the interpretation of the presented reference values. It is important to keep in mind that this sample was probably a selection of healthy participants. This might have led

to optimistic reference values. Hence, the cardiorespiratory fitness of an exercise test participant who is compared with the present reference values might be evaluated too pessimistically. This has to be considered by the exercise test instructor.

It is important that reference values for cardiorespiratory fitness are presented and updated regularly. Further accurate studies are needed to create valid and generalisable reference values. These studies should aim to evaluate the generalisability and to perform external validation.

## 7 Statistical code

### 7.1 Web application

#### 7.1.1 ui.R

```
library(shiny)
library(shinythemes)

# Define UI for application that draws a histogram
shinyUI(fluidPage(theme = shinytheme("sandstone"),

# Application title
titlePanel(h1("Peak oxygen uptake calculator for cycle ergometry")),

navbarPage(id="selectedTab",

",

# how to put the title on top and the panels below?
# br() # does not work!

#####
# First NavbarPanel
#####
tabPanel(

  "Quantile reference values",

  # Math equations
  withMathJax(),

  # Sidebar with a slider input for number of bins
  sidebarLayout(

    sidebarPanel(

      #####
      # Main
      #####
      wellPanel(

        selectInput("gender", "Your gender",
                    choices = c("Male", "Female")),
        sliderInput("age", "Your age",
                    min = 25, max = 69, value = 40, step = 1),
        uiOutput("slider"),
        selectInput("relabs", "Absolute or relative  $\dot{V}O_2$ - $\dot{V}O_{2peak}$ ",
                    choices = c("Relative", "Absolute")),
        helpText("Units:", br(),
                 "Absolute:  $[mL \cdot \min^{-1}]$ ", br(),
                 "Relative:  $[mL \cdot \min^{-1}] \cdot kg^{-1} \cdot [m^2 \cdot s^{-1}]$ ")

      ), # end of well panel

      #####
      # Display further options
      #####
      wellPanel(
        h3("Further options"),
        checkboxInput("opts_ci", "Display confidence intervals", F),
        checkboxInput("opts_ex", "Exclude smokers and obese participants from reference values", F),
        checkboxInput("opts_tab", "Display regression coefficients", F)
      )# end of well panel

    ),

    # Show a plot of the generated distribution
```

```

mainPanel(

  h4("Your percentile"),
  textOutput("txt"),
  plotOutput("nomo", height="700"),

  conditionalPanel(
    condition = "input.opts_tab == true",
    h4("Regression coefficients"),
    tabPanel("coefficients", tableOutput("ConditionalTable"))
  )

)# End of mainPanel
)# End of this sidebarLayout
),# End of this tabPanel

#####
# Second NavbarPanel
#####

tabPanel(withMathJax("Estimate  $\dot{V} \text{ vs } \dot{V}_{0-2peak}$ "),

  sidebarLayout(

    sidebarPanel(

      sliderInput("wgt", "Your body weight (kilograms)",
        min = 50, max = 125, value = 75, step = 1),
      sliderInput("wtt", "Your maximal work rate (watts)",
        min = 90, max = 350, value = 200, step = 5)

    ),# End of SidebarPanel

    mainPanel(

      wellPanel(
        h3("Estimated  $\dot{V} \text{ vs } \dot{V}_{0-2peak}$ "),
        tableOutput("vo2.tab"),
        helpText("Units:", br(),
          "Absolute: [mL*min-1]", br(),
          "Relative: [mL*min-1]*kg-1"]
        ),# End of well panel

      wellPanel(
        h3("Rationale"),
        p("If no gas exchange measurement was conducted during cardiopulmonary exercise testing,
 $\dot{V} \text{ vs } \dot{V}_{0-2peak}$ 
can be estimated based on maximal work rate from incremental exercise tests and body weight.")
        ),# End of well panel

    ),# End of mainPanel
  ),# End of sidebarLayout

),# End of tabPanel2

#####
# Third NavbarPanel
#####
tabPanel("About",

  mainPanel(
    wellPanel(
      h3("Quantile reference values for cycle ergometry"),br(),
      p("Quantile reference values have been derived from a sample of 10,090 German white-collar workers who were recorded to the
'Prevention Fitst Registry'. For further information please check [1].")
    ),# End of well panel

    wellPanel(
      h3("Estimation of  $\dot{V} \text{ vs } \dot{V}_{0-2peak}$ "),br(),
      p("For estimation we used the ACSM equation [2]:"),br(),
      p(" $\dot{V} \text{ vs } \dot{V}_{0-2peak} = (10.8 * \dot{W} * M^{-1}) + 7$ "),
      helpText("W = maximal work rate in incremental cycle ergometry measured in watts", br(),
        "M = body mass in kilograms")
    ), # End of well panel

    wellPanel(
      h3("References"),br(),
      tableOutput("tab.ref")
    )
  )
)

```

```

    )# End of well panel

    )# End of mainPanel

    )# End of tabPanel3

)# End of this navbarPage
)# End of fluidPage
)# End of Ui

```

## 7.1.2 server.R

```

library(shiny)

# Define server logic required to draw a histogram
shinyServer(function(input, output) {

#####
# V02max slider
#####

# Make dynamic slider for absolute/relative V02max
output$slider <- renderUI({
  sliderInput("inSlider", "Your result from exercise test using cycle ergometry",
             min = ifelse(input$relabs == "Relative", 11, 0.9),
             max = ifelse(input$relabs == "Relative", 61, 5),
             value = ifelse(input$relabs == "Relative", 30, 2.7),
             step = ifelse(input$relabs == "Relative", 1, 0.1))
})

#####
# Text Output
#####

output$txt <- renderText({

# quantiles
q_m_abs <- read.csv("../www/quantiles/q_m_abs.csv")
q_f_abs <- read.csv("../www/quantiles/q_f_abs.csv")
q_m_rel <- read.csv("../www/quantiles/q_m_rel.csv")
q_f_rel <- read.csv("../www/quantiles/q_f_rel.csv")

q_m_ex_abs <- read.csv("../www/quantiles/q_m_ex_abs.csv")
q_f_ex_abs <- read.csv("../www/quantiles/q_f_ex_abs.csv")
q_m_ex_rel <- read.csv("../www/quantiles/q_m_ex_rel.csv")
q_f_ex_rel <- read.csv("../www/quantiles/q_f_ex_rel.csv")

if(input$gender == "Male" & input$relabs == "Relative" & input$opts_ex == F) mat <- q_m_rel
if(input$gender == "Female" & input$relabs == "Relative" & input$opts_ex == F) mat <- q_f_rel
if(input$gender == "Male" & input$relabs == "Absolute" & input$opts_ex == F) mat <- q_m_abs
if(input$gender == "Female" & input$relabs == "Absolute" & input$opts_ex == F) mat <- q_f_abs

if(input$gender == "Male" & input$relabs == "Relative" & input$opts_ex == T) mat <- q_m_ex_rel
if(input$gender == "Female" & input$relabs == "Relative" & input$opts_ex == T) mat <- q_f_ex_rel
if(input$gender == "Male" & input$relabs == "Absolute" & input$opts_ex == T) mat <- q_m_ex_abs
if(input$gender == "Female" & input$relabs == "Absolute" & input$opts_ex == T) mat <- q_f_ex_abs

# confidence intervals
ci_m_abs <- read.csv("../www/ci_for_quantiles/ci_m_abs.csv")
ci_f_abs <- read.csv("../www/ci_for_quantiles/ci_f_abs.csv")
ci_m_rel <- read.csv("../www/ci_for_quantiles/ci_m_rel.csv")
ci_f_rel <- read.csv("../www/ci_for_quantiles/ci_f_rel.csv")

ci_m_ex_abs <- read.csv("../www/ci_for_quantiles/ci_m_ex_abs.csv")
ci_f_ex_abs <- read.csv("../www/ci_for_quantiles/ci_f_ex_abs.csv")
ci_m_ex_rel <- read.csv("../www/ci_for_quantiles/ci_m_ex_rel.csv")
ci_f_ex_rel <- read.csv("../www/ci_for_quantiles/ci_f_ex_rel.csv")

if(input$gender == "Male" & input$relabs == "Relative" & input$opts_ex == F) ci <- ci_m_rel
if(input$gender == "Female" & input$relabs == "Relative" & input$opts_ex == F) ci <- ci_f_rel
if(input$gender == "Male" & input$relabs == "Absolute" & input$opts_ex == F) ci <- ci_m_abs
if(input$gender == "Female" & input$relabs == "Absolute" & input$opts_ex == F) ci <- ci_f_abs

```

```

if(input$gender == "Male" & input$relabs == "Relative" & input$opts_ex == T) ci <- ci_m_ex_rel
if(input$gender == "Female" & input$relabs == "Relative" & input$opts_ex == T) ci <- ci_f_ex_rel
if(input$gender == "Male" & input$relabs == "Absolute" & input$opts_ex == T) ci <- ci_m_ex_abs
if(input$gender == "Female" & input$relabs == "Absolute" & input$opts_ex == T) ci <- ci_f_ex_abs

# x and y coordinates of matrices

mat_x <- which(input$page == 25:69) + 1
if(input$relabs == "Relative") mat_y <- which(input$inSlider == 11:61)
if(input$relabs == "Absolute") mat_y <- which(input$inSlider == seq(from = 0.9, to = 5, by = 0.1))

# confidence intervals reactive to checkbox
if(input$opts_ci == F){ci_txt <- ""} else {ci_txt <- paste("(95% CI ", ci[mat_y,mat_x], ")", sep = "")}

t <- paste(mat[mat_y,mat_x]*100, "% ", ci_txt, " of the reference population had a lower peak oxygen uptake than you.", sep = "")
})

#####
# NOMOGRAM default plot
#####

output$nomo <- renderPlot({

  req(input$inSlider)

  # Reading data
  m_abs_all <- read.csv("./www/m_abs_all.csv")
  m_abs_ex <- read.csv("./www/m_abs_ex.csv")
  m_rel_all <- read.csv("./www/m_rel_all.csv")
  m_rel_ex <- read.csv("./www/m_rel_ex.csv")
  f_abs_all <- read.csv("./www/f_abs_all.csv")
  f_abs_ex <- read.csv("./www/f_abs_ex.csv")
  f_rel_all <- read.csv("./www/f_rel_all.csv")
  f_rel_ex <- read.csv("./www/f_rel_ex.csv")

  # colours
  col <- c("dodgerblue", "dodgerblue1", "tomato3", "dodgerblue2", "dodgerblue3", "dodgerblue4",
          "tomato3",
          "dodgerblue4", "dodgerblue3", "tomato3", "dodgerblue2", "dodgerblue1", "dodgerblue")

  # function for fitting of ggplots
  library(ggplot2)
  library(directlabels)

  # function for fitting of ggplots
  plt <- function(dat, sex, ymin, ymax, rel = T, ci = F){

    # y-label
    if(rel == T){
      # y label
      ylb <- expression("Relative"~dot("V")~O["2peak"]~"[L*min"^-1]*kg"^-1*")")
      # y breaks by 1
      ybr <- 1
      # digits 0
      ymin <- round(ymin, digits = 0)
    }
    if(rel == F){
      # y label
      ylb <- expression(dot("V")~O["2peak"]~"[L*min"^-1]*")")
      # y breaks by 0.1
      ybr <- 0.1
      # digits 2
      ymin <- round(ymin, digits = 1)
    }
  }

  lmt <- c(ymin, ymax)
  brk <- seq(from = ymin, to = ymax, by = ybr)

  ggp <- ggplot(data = dat, aes(x = ALTER, y = fit)) +

  # defining background of ggplot
  theme(panel.grid.major = element_line(colour = "grey30", linetype = "dotted"),
        panel.background = element_rect(fill = "white"),
        axis.text.x = element_text(angle = 90, vjust=0.5, size=14),

```

```

    axis.text.y = element_text(size=14),
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    axis.title = element_text(size=14,face="bold")) +
  scale_x_continuous(limits = c(25,69),
                    breaks = seq(from = 25, to = 69, by = 1))+
  scale_y_continuous(limits = lmt,
                    breaks = brk) +
  labs(x = "Age",
       y = ylb,
       title = sex) +
  geom_vline(xintercept = c(30, 40, 50, 60),
            colour = "gray25",
            size = 0.6)

# adding confidence bands
if(ci == T){
  ggp <- ggp +
    geom_ribbon(aes(x = ALTER, ymin = lower, ymax = higher, fill = percentile),
              alpha = 0.3, inherit.aes = F, show.legend = F, na.rm = T) +
    scale_fill_manual(values = col)
}

# adding quantile curves
ggp <- ggp +
  geom_line(aes(x = ALTER, y = fit, colour = percentile),
           show.legend = F, size = 1, na.rm = T) +
  scale_color_manual(values = col)

ggp <- direct.label(ggp, list("last.polygons", colour = "white"))

return(ggp)
}# End of plt-function

# if else etatements to select adjacent plot
# strategy to avoid error "argument is of length zero":
# decisions...
# 1st: are there additional opts? --> no? --> data set "all"
# 2nd: additional opts but no exclusion? --> data set "all"
# 3rd: additional opts and exclusion? --> data set "ex"
if(input$opts_ex == F){

  if(input$gender == "Male" & input$relabs == "Relative") ds <- m_rel_all
  if(input$gender == "Female" & input$relabs == "Relative") ds <- f_rel_all
  if(input$gender == "Male" & input$relabs == "Absolute") ds <- m_abs_all
  if(input$gender == "Female" & input$relabs == "Absolute") ds <- f_abs_all
}

if(input$opts_ex == T){

  if(input$gender == "Male" & input$relabs == "Relative") ds <- m_rel_ex
  if(input$gender == "Female" & input$relabs == "Relative") ds <- f_rel_ex
  if(input$gender == "Male" & input$relabs == "Absolute") ds <- m_abs_ex
  if(input$gender == "Female" & input$relabs == "Absolute") ds <- f_abs_ex
}

# relative or absolute
if(input$relabs == "Relative") rel.true <- T
if(input$relabs == "Absolute") rel.true <- F

yc <- input$inSlider

ggplt <- plt(ds, input$gender, min(ds[, "lower"]), max(ds[, "higher"]), rel = rel.true, ci = input$opts_ci)
ggplt + geom_point(aes(x = input$age, y = yc), shape = 21, size = 5, colour = "black", fill = "red", stroke = 1)

})

#####
# coefficient tables
#####
output$ConditionalTable <- renderTable({

# Overall
if(input$opts_ex == F & input$gender == "Female" & input$relabs == "Absolute") ds1 <- read.csv("./www/tables/adff.csv")
if(input$opts_ex == F & input$gender == "Male" & input$relabs == "Absolute") ds1 <- read.csv("./www/tables/adfm.csv")

```



```

if(input$opts_ex == F & input$gender == "Female" & input$relabs == "Relative") ds1 <- read.csv("./www/tables/rdff.csv")
if(input$opts_ex == F & input$gender == "Male" & input$relabs == "Relative") ds1 <- read.csv("./www/tables/rdfm.csv")

# Exclusion of smokers and obese
if(input$opts_ex == T & input$gender == "Female" & input$relabs == "Absolute") ds1 <- read.csv("./www/tables/exadff.csv")
if(input$opts_ex == T & input$gender == "Male" & input$relabs == "Absolute") ds1 <- read.csv("./www/tables/exadfm.csv")

if(input$opts_ex == T & input$gender == "Female" & input$relabs == "Relative") ds1 <- read.csv("./www/tables/exrdff.csv")
if(input$opts_ex == T & input$gender == "Male" & input$relabs == "Relative") ds1 <- read.csv("./www/tables/exrdfm.csv")

ds1

}, digits=7)

#output$ConditionalTable <- renderTable({

  #if(input$opts_tab == F) data.frame()
  #if(input$opts_tab == T) read.csv("./www/tables/adff.csv")
  # read.csv("./www/tables/adff.csv")

  }, digits=7)

#####
# Estimation of VO2max
#####

output$vo2.tab <- renderTable({

  # based on ACSM guidelines (Franklin, 2000)
  in.thousand <- 10.8 * input$wt + 7 * input$wt
  abs.vo2.est <- in.thousand / 1000
  rel.vo2.est <- in.thousand / input$wt

  data.frame(Absolute = abs.vo2.est, Relative = rel.vo2.est)

})

#####
# References
#####

output$tab.ref <- renderTable({

  data.frame(
    number <- c("1", "2"),
    ref <- c("Rapp D, Scharhag J, Wagenpfeil S, Scholl J. Reference values for peak oxygen uptake: cross-sectional analysis of cycle ergometry-
    based cardiopulmonary exercise tests of 10,090 adult German volunteers from the Prevention First Registry. BMJ Open. 2018 Mar 5;8(3):e018697.
    doi: 10.1136/bmjopen-2017-018697",
    "Franklin BA, Balady GJ, American College of Sports Medicine. ACSM's guidelines for exercise testing and prescription. 6. ed.
    Philadelphia: Lippincott Williams & Wilkins 2000, ISBN: 978-0-7817-2735-8, p. 303.")
  )

}, include.colnames = F)

})

```

## 7.2 Calculations for web application

```
#####
# DATA AND PACKAGES
#####

library(quantreg)
library(parallel)

# data sets overall and exclusion of obese and smokers
df <- readRDS("./170525_n10090_bmj")
df <- df[,c("Geschlecht", "ALTER", "REL_VO2_MAX", "VO2_MAX", "obese", "ZIGARETTEN")]
dfm <- subset(df, Geschlecht == "Male")
dff <- subset(df, Geschlecht == "Female")
dfm_ex <- subset(dfm, obese == "no" & ZIGARETTEN == "no")
dff_ex <- subset(dff, obese == "no" & ZIGARETTEN == "no")

# quantiles
taus <- seq(from = 0.01, to = 0.99, by = 0.01)

# range of REL vo2max (0.1 to 99.9 percentiles of data)
rvo <- seq(from = 11, to = 61, by = 1)
avo <- seq(from = 0.9, to = 5, by = 0.1)

# range of age
ag <- 25:69

rfrm <- as.formula("REL_VO2_MAX ~ ALTER + I(ALTER^2)")
afrm <- as.formula("VO2_MAX ~ ALTER + I(ALTER^2)")

nboot <- 10000

#####
# FUNCTIONS FOR MATRIX "QUANTILE"
#####

# equivalent to predict.rq
# predict.rq cannot be applied
# as 10,000*100 quantile regressions (coefficients, residuals etc) are stored in RAM (too large)
# this function is able to predict with only coefficients and not whole rq-object
predict_rq <- function(fit){

  Intercept = fit[1,]
  Slope1 = fit[2,]
  Slope2 = fit[3,]

  newdata = matrix(rep(25:69, 99), nrow = 45)
  colnames(newdata) <- as.character(taus)

  prd <- function(i) Intercept[i] + Slope1[i] * newdata[,i] + Slope2[i] * newdata[,i] * newdata[,i]

  for(i in 1:99) newdata[,i] <- prd(i)

  newdata <- as.data.frame(newdata)
  newdata$ALTER <- 25:69 # adding ALTER

  return(newdata)
}

# which predicted value from which tau is the closest for given age and vo2max?
# CAVE: data set has to be called pred!
q.check.fun <- function(age, vo2){

  sset <- subset(pred, ALTER == age)
  sset <- sset[,!(names(sset)) == "ALTER"] # drop ALTER

  row <- which.min(abs(vo2 - sset))

  return(names(row))
}

# INNER FUNCTION
# for a given age: taus are calculated for each vo2max
q.inner <- function(x) {
```

```

col <- lapply(vo, function(y) q.check.fun(age = x, vo2 = y))
col <- do.call(data.frame, col)
col <- t(col)
row.names(col) <- vo

return(col)
}

# OUTER FUNCTION
q.outer <- function() {

matrix <- lapply(ag, q.inner)

matrix <- do.call(data.frame, matrix)
colnames(matrix) <- as.character(ag)
row.names(matrix) <- as.character(vo)

return(matrix)
}

## FUNCTION OVERALL
#fit <- rq(frm, data = in_data, tau = taus)$coefficients
#pred <- predict_rq(fit)
#res <- q.outer()

## FUNCTION CALLS
t1 <- Sys.time()

#####
## relvo2max
#####
frm <- rfrm
vo <- rvo

# MALES
fit <- rq(frm, data = dfm, tau = taus)$coefficients
pred <- predict_rq(fit)
q_m_rel <- q.outer()
write.csv(q_m_rel, "q_m_rel.csv")

fit <- rq(frm, data = dfm_ex, tau = taus)$coefficients
pred <- predict_rq(fit)
q_m_ex_rel <- q.outer()
write.csv(q_m_ex_rel, "q_m_ex_rel.csv")

# FEMALES
fit <- rq(frm, data = dff, tau = taus)$coefficients
pred <- predict_rq(fit)
q_f_rel <- q.outer()
write.csv(q_f_rel, "q_f_rel.csv")

fit <- rq(frm, data = dff_ex, tau = taus)$coefficients
pred <- predict_rq(fit)
q_f_ex_rel <- q.outer()
write.csv(q_f_ex_rel, "q_f_ex_rel.csv")

#####
## absvo2max
#####
frm <- afrm
vo <- avo

# MALES
fit <- rq(frm, data = dfm, tau = taus)$coefficients
pred <- predict_rq(fit)
q_m_abs <- q.outer()
write.csv(q_m_abs, "q_m_abs.csv")

fit <- rq(frm, data = dfm_ex, tau = taus)$coefficients
pred <- predict_rq(fit)
q_m_ex_abs <- q.outer()
write.csv(q_m_ex_abs, "q_m_ex_abs.csv")

# FEMALES
fit <- rq(frm, data = dff, tau = taus)$coefficients
pred <- predict_rq(fit)

```

```

q_f_abs <- q.outer()
write.csv(q_f_abs, "q_f_abs.csv")

fit <- rq(frm, data = dff_ex, tau = taus)$coefficients
pred <- predict_rq(fit)
q_f_ex_abs<- q.outer()
write.csv(q_f_ex_abs, "q_f_ex_abs.csv")

t2 <- Sys.time()
t2-t1 # time spent

#####
# BOOTSTRAP FUNCTIONS
#####

# function to calculate 10,000 bootstrap samples
# output are 10,000 rq coefficients!
boot_all <- function(dat){

  cl <- makeCluster(mc <- getOption("cl.cores", parallel::detectCores()-1))
  clusterEvalQ(cl, library(quantreg))
  clusterExport(cl, varlist = c("taus", "dfm", "dff", "dfm_ex", "dff_ex", "frm"))

  # creating bootstrap sample (replace = T)
  gen_sample <- function() dat[sample(nrow(dat), replace = T),]

  l1 <- parLapply(cl, 1:nboot, function(z) rq(frm, data = gen_sample(), tau = taus)$coefficients )

  stopCluster(cl)
  return(l1)
}

# predicted values for all quantile regressions and all ages
# input must be called l
pred_all <- function() {

  l1 <- lapply(1:nboot, function(z) {

    d <- predict_rq(l[[z]])
    return(d)

  })

  return(l1)
}

# function for checking out nearest tau of quantile regression
# list must be called l1
check.fun <- function(age, vo2){

  l2 <- lapply(l1, function(x) subset(as.data.frame(x), ALTER == age)[,as.character(taus)])
  qt1 <- lapply(l2, function(x) {

    col <- which.min(abs(x - vo2))
    qt1 <- as.numeric(names(col))

    return(qt1)
  })# End of lapply

  qt11 <- t(do.call(data.frame, qt1))
  lo <- round(quantile(qt11, probs = c(0.025))*100, digits = 0)
  hi <- round(quantile(qt11, probs = c(0.975))*100, digits = 0)
  ci <- paste0(lo, "% to ", hi, "%")

  return(ci)
}

# apply check.fun to all achieved vo2_max (inner function) and all ages (outer function)
# INNER FUNCTION
inner <- function(x) {

  col <- lapply(vo, function(y) check.fun(age = x, vo2 = y))
  col <- do.call(data.frame, col)
  col <- t(col)
  row.names(col) <- vo

  return(col)
}

```

```

}

# OUTER FUNCTION
outer <- function() {
  cl <- makeCluster(mc <- getOption("cl.cores", parallel::detectCores()-1))
  clusterExport(cl=cl, varlist=c("vo", "ag", "l1", "inner", "check.fun", "taus"))
  clusterCall(cl, function() library("quantreg"))
  ci_matrix <- parLapply(cl, ag, inner)
  stopCluster(cl)

  ci_matrix <- do.call(data.frame, ci_matrix)
  colnames(ci_matrix) <- as.character(ag)
  row.names(ci_matrix) <- as.character(vo)

  return(ci_matrix)
}

## FUNCTION OVERALL
#l <- boot_all(dat_input)
#l1 <- pred_all()
#res <- outer()

## FUNCTION CALLS

#####
## relvo2max
#####
frm <- rfrm
vo <- rvo

# MALES
l <- boot_all(dfm)
l1 <- pred_all()
ci_m_rel <- outer()
write.csv(ci_m_rel, "ci_m_rel.csv")

l <- boot_all(dfm_ex)
l1 <- pred_all()
ci_m_ex_rel <- outer()
write.csv(ci_m_ex_rel, "ci_m_ex_rel.csv")

# FEMALES
l <- boot_all(dff)
l1 <- pred_all()
ci_f_rel <- outer()
write.csv(ci_f_rel, "ci_f_rel.csv")

l <- boot_all(dff_ex)
l1 <- pred_all()
ci_f_ex_rel <- outer()
write.csv(ci_f_ex_rel, "ci_f_ex_rel.csv")

#####
## absvo2max
#####
frm <- afrm
vo <- avo

# MALES
l <- boot_all(dfm)
l1 <- pred_all()
ci_m_abs <- outer()
write.csv(ci_m_abs, "ci_m_abs.csv")

l <- boot_all(dfm_ex)
l1 <- pred_all()
ci_m_ex_abs <- outer()
write.csv(ci_m_ex_abs, "ci_m_ex_abs.csv")

# FEMALES
l <- boot_all(dff)
l1 <- pred_all()
ci_f_abs <- outer()
write.csv(ci_f_abs, "ci_f_abs.csv")

l <- boot_all(dff_ex)

```

```

11     <- pred_all()
ci_f_ex_abs <- outer()
write.csv(ci_f_ex_abs, "ci_f_ex_abs.csv")

# END

```

## 7.3 Statistical code

### 7.3.1 Specifications

```

# loading packages
library(Hmisc)
library(compareGroups)
library(boot)
library(quantreg)
library(grid)
library(gridExtra)
library(xtable)
library(psych)
library(e1071)
library(reshape2)
library(plyr)
library(dplyr)
library(epitools)
library(directlabels)
library(splines)

# quantiles
TAUS <- c(0.05,0.1,0.2,0.25,0.3,0.4,0.5,0.6,0.7,0.75,0.8,0.9,0.95)

# defining number of bootstrap samples
NBOOT <- 10000

# labels for plots
REL_LAB <- expression("Relative"~dot("V")*0["2peak"]~"[mL"*0["2"]*/min/kg*")")
ABS_LAB <- expression("Absolute"~dot("V")*0["2peak"]~"[L"*0["2"]*/min*")")

# function for saving plots
SAVE.PLOT <- function(PLOT, FILE.NAME, HEIGHT, WIDTH){

  # path
  P1 <- paste0("./results/fig/", FILE.NAME)

  # format
  png(filename = P1,
       units = "in",
       width = WIDTH,
       height = HEIGHT,
       res = 300)

  plot(PLOT)

  dev.off()

}

```

### 7.3.2 Data

```

#####
# reading data
#####

# reading full data
df <- read.table(
  "./data/160301_Originaldaten.csv",
  header = T, sep = ";", dec = ",")

```

```

# adding missing data
df2 <- read.table(
  "../data/missings161020.csv",
  header = T, dec = ".", sep = ";")

#####
# Calculating new variables
#####

# renaming id
names(df)[which(names(df) == "Ã..ID")] <- "id"
names(df2)[which(names(df2) == "Ã..ID")] <- "id"

# adding missing data
df_miss <- subset(df, id %in% df2$id)
df_miss <- df_miss[order(df_miss$id),]
df2 <- df2[order(df2$id),]
df_miss$ALTER <- df2$ALTER # adding age from df2
df_miss$REL_VO2_MAX <- df2$REL_VO2_MAX # adding rel VO2 from df2
df_miss$VO2_MAX <- df2$VO2_MAX # adding abs VO2 from df2
df_nomiss <- subset(df, !(df$id %in% df2$id))
df <- rbind(df_miss, df_nomiss)

# define date variable
df$datum <- as.Date(df$DATUM, "%d.%m.%Y")
df <- df[order(df$datum),]
df$DATUM <- NULL

# recode binary variables with -1 / 0
# defining relevant variables
vars <- c("ZIGARETTEN", "EXRAUCHER")

# function for changing -1 to 1
f <- function(x){
  ifelse(x == -1, 1, x)}

# apply function to all relevant variables
df[, vars] <- sapply(df[, vars], f)

# computing vars
# diabetes, hypertension, overweight, obesity
df$MANIFEST_DM <- ifelse(df$BLUTZUCKER >= 126 | df$HBA1C >= 6.5, 1, 0)
df$BLUTHOCHDRUCK_WHO <- ifelse(df$SYST >= 140 | df$DIAST >= 90, 1, 0)
df$BMI_GE_25 <- ifelse(df$BMI >= 25, 1, 0)
df$obese <- ifelse(df$BMI >= 30, 1, 0)

#####
# adding study center
#####

sto <- read.csv("../data/160810_standort.csv")
df <- merge(df, sto, by = "id", all.x = T)
df$standort <- factor(df$standort,
  levels = c("F", "M"),
  labels = c("Frankfurt/Ruedesheim", "Munich"))

#####
# Implausible cases
#####

# Function for assiging NA to implausible cases
f <- function(var, mi = min(df[var]), ma = max(df[var])){
  d <- ifelse(df[var] > ma | df[var] <= mi, NA, df[[var]])
  d <- as.vector(d)
  return(d)
}

df$ALTER <- f("ALTER", 15, 100)
df$KOERPERFETT_TANITA <- f("KOERPERFETT_TANITA", 0, 100)
df$KOERPERFETT_CALIPER <- f("KOERPERFETT_CALIPER", 0, 100)
df$SYST <- f("SYST", 50, 400)
df$DIAST <- f("DIAST", 30, 200)
df$VO2_MAX <- f("VO2_MAX", 0, 7)
df$REL_VO2_MAX <- f("REL_VO2_MAX", 0, 100)
df$GESAMTCHOLESTERIN <- f("GESAMTCHOLESTERIN", 0, 600)
df$HDL_CHOLESTERIN <- f("HDL_CHOLESTERIN", 10, 150)
df$TRIGLYCERIDE <- f("TRIGLYCERIDE", 0, 2000)
df$HBA1C <- f("HBA1C", 0, 100)

```

```

# implausible dates
df <- df[order(df$datum), ]
df$datum[df$datum < as.Date("2000-01-01")] <- NA

# defining factor variables in df
f <- function(var){
  out <- factor(df[,var], levels = c(0,1), labels = c("no", "yes"))
}

df$Geschlecht <- factor(df$Geschlecht,
  levels = c(0,1),
  labels = c("Men", "Women"))
df$BMI_GE_25 <- f("BMI_GE_25")
df$obese <- f("obese")
df$ZIGARETTEN <- f("ZIGARETTEN")
df$EXRAUCHER <- f("EXRAUCHER")
df$MANIFEST_DM <- f("MANIFEST_DM")
df$BLUTHOCHDRUCK_WHO <- f("BLUTHOCHDRUCK_WHO")

# create variable ex smoker
df$EXRAUCHER[which(df$ZIGARETTEN == "yes")] <- "no"
df$EXRAUCHER[which(df$ZIGARETTEN == "no" & is.na(df$EXRAUCHER))] <- "no"

# add variable age group
df$ageclass <- cut(df$ALTER,
  breaks = c(25,30,35,40,45,50,55,60,65,69),
  include.lowest = TRUE, include.highest = F, right = F)
# creating numerical age class used as linear predictor in regression (validation)
df$ageclass_double <- as.double(df$ageclass)

#####
# cleaning data
#####

keeps <- c("id", "Geschlecht", "ALTER",
  "GEWICHT", "GROESSE",
  "BMI", "BAUCHUMFANG", "KOERPERFETT_CALIPER", "KOERPERFETT_TANITA",
  "SYST", "DIAST",
  "VO2_MAX", "REL_VO2_MAX", "MAX_LEISTUNG",
  "GESAMTCHOLESTERIN", "HDL_CHOLESTERIN", "LDL_CHOLESTERIN", "TRIGLYCERIDE",
  "ZIGARETTEN", "EXRAUCHER",
  "MANIFEST_DM", "BLUTHOCHDRUCK_WHO", "BMI_GE_25", "obese",
  "ageclass", "ageclass_double",
  "BLUTZUCKER", "HBA1C", "datum", "standort")

df <- df[,which(names(df) %in% keeps)]

#####
# Adding Hmisc variable labels
#####
label(df$Geschlecht) <- "Sex"
label(df$ALTER) <- "Age [years]"
label(df$ageclass) <- "Age class"
label(df$REL_VO2_MAX) <- "Relative  $\dot{V}O_{2peak}$  [ $\frac{mL}{min \cdot kg}$ ]"
label(df$VO2_MAX) <- "Absolute  $\dot{V}O_{2peak}$  [ $\frac{L}{min}$ ]"
label(df$GEWICHT) <- "Weight [kg]"
label(df$GROESSE) <- "Height [cm]"
label(df$BMI) <- "BMI [ $\frac{kg}{m^2}$ ]"
label(df$KOERPERFETT_CALIPER) <- "Body fat Caliper [%]"
label(df$BMI_GE_25) <- "Overweight"
label(df$obese) <- "Obese"
label(df$SYST) <- "Systolic [mmHg]"
label(df$DIAST) <- "Diastolic [mmHg]"
label(df$BLUTHOCHDRUCK_WHO) <- "Hypertension"
label(df$BLUTZUCKER) <- "Blood glucose [ $\frac{mg}{dL}$ ]"
label(df$HBA1C) <- "HbA1c [%]"
label(df$MANIFEST_DM) <- "Diabetes mellitus"
label(df$GESAMTCHOLESTERIN) <- "Total cholesterol [ $\frac{mg}{dL}$ ]"
label(df$HDL_CHOLESTERIN) <- "HDL Cholesterol [ $\frac{mg}{dL}$ ]"
label(df$LDL_CHOLESTERIN) <- "LDL Cholesterol [ $\frac{mg}{dL}$ ]"
label(df$TRIGLYCERIDE) <- "Triglycerides [ $\frac{mg}{dL}$ ]"
label(df$ZIGARETTEN) <- "Smoker"
label(df$EXRAUCHER) <- "Ex smoker"
label(df$standort) <- "Study center"

```



## 7.3.3 Random sample

```
#####
# draw random sample
#####

set.seed(1)
spl<-df[sample(nrow(df),252),]

#####
# compare full data vs. sample
#####
# df for comparison of df and spl
spl$spl <- 1
df_no_spl <- subset(df, !(df$id %in% spl$id))
df_no_spl$spl <- 0
test_spl <- rbind(df_no_spl, spl)
test_spl$spl <- factor(test_spl$spl,
                      levels = c(0,1),
                      labels = c("Full data", "Random sample"))

# defining quantitative and categorical variables
qvars<-c("ALTER", "REL_VO2_MAX", "VO2_MAX", "GEWICHT", "GROESSE", "BMI")
cvars<-c("Geschlecht", "ZIGARETTEN", "BMI_GE_25", "standort")

# creating formula
frm <- as.formula(paste0("spl ~", paste(c(qvars, cvars), collapse = " + ")))

# creating bivariate table for sample vs. full data
tab <- compareGroups(frm, data = test_spl, method = 2)
tab <- createTable(tab)
export2latex(tab)

#####
# reading file with manually added information
#####

# read sample Ruedesheim and create two data sets
# sr: only sample characteristics
# sf: merged with original data
# (without excluded cases as random IDs were drawn prior to exclusion)
sr <- read.csv("../data/sample_Rued_ergaenz_160622.csv",
              sep=";", dec=",", header=T)
sr <- merge(sr, df, by.x = "ID", by.y = "id", all.x = T)

#####
# measures of exertion
#####

# calculating age-predicted maximal heart rate (APMHR, pmid: 11153730)
# the criterion for maximal effort is 90% of APMHR (pmid: 18027991)
sr$th_hr_max <- (208 - 0.7 * sr$ALTER) * 0.9

# measure of exertion: lac>8 OR RER >1.1 OR HR >90% of APMHR
sr$lac_ge8 <- ifelse(sr$laktat >= 8, 1, 0)
sr$rer_ge1 <- ifelse(sr$RER >= 1.1, 1, 0)
sr$hr_ge <- ifelse(sr$HF.max >= sr$th_hr_max, 1, 0)

# specifying rows where lac, rer and hr are all NA
sr$na <- rowSums(is.na(cbind(sr$rer_ge1, sr$lac_ge8, sr$hr_ge)))
sr$Ex <- rowSums(cbind(sr$lac_ge8, sr$rer_ge1, sr$hr_ge), na.rm=T)
sr$EX <- ifelse(sr$na == 3, NA, sr$Ex)
sr$Exertion <- ifelse(sr$EX >= 1, 1, sr$EX)

# Bootstrapped 95% confidence intervals for proportions
SAMPLE.PROPORTION <- function(x, d) {
  z = sum(x[d]==1, na.rm = T)
  n = sum(x[d]!=0, na.rm = T)
  p = z/(z+n)
  return(p)
}

# function to list N(%) [95% CI] of exertion
EXERTION <- function(SUBGROUP = ""){

  x <- sr[SUBGROUP, "Exertion"]
  YES <- sum(x == 1, na.rm = T)

```

```

NO <- sum(x == 0, na.rm = T)
N <- paste0(YES, "/", YES + NO)
PERCENT <- round((YES / (YES + NO))*100, digits = 0)

set.seed(1)
b <- boot(x, SAMPLE.PROPORTION, R = NBOOT)
CI <- paste(round(quantile(b$t, probs = c(0.025, 0.975))*100, digits = 0))
return(paste0(N, " (", PERCENT, "%)", " [", CI[1], "% to ", CI[2], "%]"))
}

# overall, males, females
EXERTION(sr$sex == "m" | sr$sex == "w")
EXERTION(sr$sex == "m")
EXERTION(sr$sex == "w")

#####
# does inclusion of cases with exertion == "no" change results?
#####

sr$Exertion <- factor(sr$Exertion, levels = c(0, 1), labels = c("no", "yes"))

# checking visually using scatter plots
# producing scatter plot with two regression lines:
# 1) all cases of the sample 2) only cases with Exertion == "yes"
# for males and females / absolute and relative vo2max (fig 4.13)

SC.PLOT <- function(SEX = "m", OUTCOME, XLAB, YLAB, TITLE = ""){

  DAT <- subset(sr, sex == SEX)
  DAT.NOEX <- subset(DAT, Exertion == "yes")

  FRM <- as.formula(paste(OUTCOME, "~ ALTER"))

  FIT <- rq(FRM, data = DAT, tau = 0.5)$coefficients
  FIT.NOEX <- rq(FRM, data = DAT.NOEX, tau = 0.5)$coefficients

  gp <- ggplot(data = DAT, aes_string(x = "ALTER", y = OUTCOME, colour = "Exertion")) +
    geom_point(na.rm = T) +
    geom_abline(intercept = FIT[1], slope = FIT[2]) +
    geom_abline(intercept = FIT.NOEX[1], slope = FIT.NOEX[2], lty = 2) +
    scale_x_continuous(limits = c(25, 69), breaks = seq(from = 25, to = 69, by = 5)) +
    scale_y_continuous(limits = c(min(DAT[,OUTCOME]), max(DAT[,OUTCOME]))) +
    labs(x = XLAB, y = YLAB, title = TITLE) +
    theme(plot.title = element_text(hjust = 0.5), legend.position="none")

  return(gp)
}

SAVE.PLOT(
  PLOT = grid.arrange(
    SC.PLOT("m", "REL_VO2_MAX", "", REL_LAB, "Men"),
    SC.PLOT("w", "REL_VO2_MAX", "", "", "Women"),
    SC.PLOT("m", "VO2_MAX", "Age [years]", ABS_LAB, ""),
    SC.PLOT("w", "VO2_MAX", "Age [years]", "", ""),
    ncol = 2),
  "sample.png", 7, 7
)

# function to create an xtable with ANCOVA- results
# Question: does Exertion significantly alter Results?
# Median regression:
# DV = VO2peak
# IV = ALTER + Exertion (yes/no) + interaction term ALTER * Exertion
ANCOVA <- function(dat = srf, outcome = "REL_VO2_MAX"){

  # 1) data without missings (avoid error by boot.rq)
  dat <- dat[!is.na(dat[,outcome ]),]
  dat <- dat[!is.na(dat[, "ALTER" ]),]
  dat <- dat[!is.na(dat[, "Exertion"]),]

  # 2) how many cases with "exertion = no"?
  print(table(dat[, "Exertion"]))

  # 2) extract variables
  A <- dat[, "ALTER"]
  E <- as.numeric(dat[, "Exertion"])
  DV <- dat[, outcome]

```

```

# 3) create formula
frm <- paste(outcome, "~ A + E + I(A * E)")
frm <- as.formula(frm)

# 4) estimate for coefficient and bootstrapped P values
fit <- rq(frm, tau = 0.5, data = dat)
estpv <- summary(fit, se = "boot")$coefficients
est <- estpv[,1]
pval <- estpv[,4]

# 5) bootstrapped confidence intervals
set.seed(1)
fit.b <- boot.rq(cbind(1, A, E, I(A * E)), DV, tau = 0.5, R = NBOOT)
ci <- t(apply(fit.b$B, 2, quantile, c(0.025, 0.975)))

# 6) cbind to create a table
tab <- cbind(est, ci, pval)
tab <- as.data.frame(tab)
row.names(tab) <- c("Intercept", "Age", "Exertion(yes)", "Age * Exertion(yes)")

print(xtable(tab, digits=c(2,2,2,2,3)), include.rownames = T)
}

ANCOVA(dat = subset(sr, sex == "m"), outcome = "REL_V02_MAX")
ANCOVA(dat = subset(sr, sex == "m"), outcome = "V02_MAX")
ANCOVA(dat = subset(sr, sex == "w"), outcome = "REL_V02_MAX")
ANCOVA(dat = subset(sr, sex == "w"), outcome = "V02_MAX")

## PLAUSIBILITY CHECK:
## visualise the regression lines for exertion = yes/no seperately
#pdat <- data.frame(age = seq(from = 20, to = 80))
#pdat$pred_exertion <- 18.27 + (0.11*pdat$age) + (5.85*2) + (-0.07*2*pdat$age)
#pdat$pred_no_exertion <- 18.27 + (0.11*pdat$age) + (5.85*1) + (-0.07*1*pdat$age)

#ggplot() +
# geom_point(aes(x = ALTER, y = REL_V02_MAX), data = srf, na.rm = T) +
# geom_line(aes(x = age, y = pred_exertion), data = pdat, colour = "red") +
# geom_line(aes(x = age, y = pred_no_exertion), data = pdat, colour = "blue")

```

### 7.3.4 Descriptive statistics

```

#####
# Numbers of cases for flow chart
#####

# function for counting non-missings
n <- function(x) sum(!is.na(x))
ex <- function(x) sum(is.na(x))

# crude n
nrow(df)
table(df$Geschlecht)

# excluded cases due to missing values
ex(df$REL_V02_MAX | df$V02_MAX)
ex(df[!is.na(df$REL_V02_MAX), "standort"])
ex(df[!is.na(df$REL_V02_MAX) &
      !is.na(df$standort), "ALTER"])

# n after exclusion
n(df[!is.na(df$REL_V02_MAX) &
      !is.na(df$standort) &
      !is.na(df$ALTER), "REL_V02_MAX"])
tapply(df[!is.na(df$REL_V02_MAX) &
          !is.na(df$standort) &
          !is.na(df$ALTER), "REL_V02_MAX",
        df[!is.na(df$REL_V02_MAX) &
            !is.na(df$standort) &
            !is.na(df$ALTER), "Geschlecht"], n)

# exclusion of age >= 70 or < 25
ex(df[!is.na(df$REL_V02_MAX) &

```

```

!is.na(df$standort) &
!is.na(df$ALTER), "ageclass"])

n(df[!is.na(df$REL_V02_MAX) &
!is.na(df$standort) &
!is.na(df$ALTER) &
!is.na(df$ageclass), "REL_V02_MAX"])

# Selective dropout analysis
df$dropout <- ifelse(
  is.na(df$REL_V02_MAX) |
  is.na(df$standort) |
  is.na(df$ALTER) |
  is.na(df$ageclass),
  1,0
)

# defining quantitative and categorical variables
qvars<-c("REL_V02_MAX", "VO2_MAX", "ALTER", "GEWICHT", "GROESSE", "BMI")
cvars<-c("Geschlecht", "ZIGARETTEN", "BMI_GE_25", "standort")

# creating formula
frm <- as.formula(paste0("dropout ~", paste(c(qvars, cvars), collapse = " + ")))

# creating bivariate table for sample vs. full data
tab <- compareGroups(frm, data = df, method = 2)
tab <- createTable(tab)
export2latex(tab)

# exclude cases
df <- df[df$dropout == 0,]
tapply(df$VO2_MAX, df$Geschlecht, n)

# split data by location
df_FR <- subset(df, standort == "Frankfurt/Ruedesheim")
df_M <- subset(df, standort == "Munich")

# split data by sex
dfm <- subset(df, Geschlecht == "Men")
dff <- subset(df, Geschlecht == "Women")

# split data by sex and location
dfm_FR <- subset(df_FR, Geschlecht == "Men")
dff_FR <- subset(df_FR, Geschlecht == "Women")
dfm_M <- subset(df_M, Geschlecht == "Men")
dff_M <- subset(df_M, Geschlecht == "Women")

#####
# Date and location
#####

SAVE.PLOT(

# density function of date of examination
ggplot(df, aes(x = datum, colour = Geschlecht)) +
  stat_bin(data=subset(df, Geschlecht=="Men"), aes(y=cumsum(..count..)), geom="step", na.rm = T) +
  stat_bin(data=subset(df, Geschlecht=="Women"), aes(y=cumsum(..count..)), geom="step", na.rm = T) +
  scale_x_date(breaks = seq(as.Date("2001-01-01"), as.Date("2015-12-31"), by="2 years"), date_labels = "%Y") +
  scale_y_continuous(breaks = seq(from = 0, to = 6000, by = 1000)) +
  scale_colour_discrete(name = "Sex",
    breaks=c("Women", "Men")) +
  geom_vline(xintercept = as.Date("2001-01-01", "%Y-%m-%d"), colour = "grey60", size = 0.3) +
  geom_vline(xintercept = as.Date("2015-12-31", "%Y-%m-%d"), colour = "grey60", size = 0.3) +
  labs(y = "Cumulative number of participants", x = "Date")

, "time.png", 5, 5
)

#####
# Bivariate descriptive Table
#####

# bivariate descriptive table

# exclude missing values from
# age, relative and absolute V02peak
# use only subset of variables
vars <- c(

```

```

# id, independent and dependent variables
"Geschlecht", "REL_VO2_MAX", "VO2_MAX", "ALTER", "ageclass",
# descriptive statistics and multivariable regression
"GEWICHT", "GROESSE", "BMI", "KOERPERFETT_CALIPER", "BMI_GE_25", "obese",
"SYST", "DIAST", "BLUTHOCHDRUCK_WHO", "BLUTZUCKER", "HBA1C", "MANIFEST_DM",
"GESAMTCHOLESTERIN", "HDL_CHOLESTERIN", "LDL_CHOLESTERIN", "TRIGLYCERIDE",
"ZIGARETTEN", "EXRAUCHER", "standort")

df1 <- df[, vars]

# description of full data
tab <- compareGroups(
  Geschlecht ~ .,
  data = df1,
  method = 2)
export2latex(createTable(tab),
  header=c(p.overall='P value'),
  caption = "Descriptive statistics by sex.")

# description of data for reference values
df1m <- subset(df1, Geschlecht == "Men")
df1f <- subset(df1, Geschlecht == "Women")

males <- compareGroups(
  standort ~ . -Geschlecht,
  data = df1m,
  method = 2)

females <- compareGroups(
  standort ~ . -Geschlecht,
  data = df1f,
  method = 2)

males <- createTable(males)
females <- createTable(females)

export2latex(cbind(Males = males, Females = females),
  header=c(p.overall='P value'),
  caption = "Descriptive statistics by sex and study center.")

#####
# Description of all variables in data set
#####

# description of all variables in original data set

# quantitative
qvars <- c(
  # target variables
  "ALTER", "REL_VO2_MAX", "VO2_MAX",
  # other quantitative variables
  "GEWICHT", "GROESSE", "BMI", "KOERPERFETT_CALIPER", "KOERPERFETT_TANITA", "BAUCHUMFANG",
  "SYST", "DIAST", "BLUTZUCKER", "HBA1C",
  "GESAMTCHOLESTERIN", "HDL_CHOLESTERIN", "LDL_CHOLESTERIN", "TRIGLYCERIDE")

qvars_labels <- c(
  # target variables
  "Age", "Relative \n V02peak", "Absolute \n V02peak",
  # other quantitative variables
  "Weight", "Height", "BMI", "Body fat \n (Caliper)", "Body fat \n (Tanita)", "Waist \n circumference",
  "Systolic", "Diastolic", "Blood \n glucose", "HbA1c",
  "Total \n cholesterol", "HDL \n cholesterol", "LDL \n cholesterol", "Tri- \n glycerides")

# equations as labels in histograms
qvars_labels2 <- qvars_labels
qvars_labels2[2] <- expression("Relative"*dot("V")*0["2peak"])
qvars_labels2[3] <- expression("Absolute"*dot("V")*0["2peak"])
qvars_labels2[13] <- expression("HbA1c["c"])

# histograms for all qvars using labels
hst <- function(dat){

  # empty list to save in
  lst <- list()

  for(i in qvars){
    # calculate means, sd and the index of label

```

```

m <- mean(dat[, i], na.rm = T)
s <- sd(dat[, i], na.rm = T)
num <- qvars_labels2[which(qvars %in% i)]

# plot
g <- ggplot(data = dat, aes_string(x = i)) +
  geom_histogram(aes(y=.density.), color="black", bins = 15, na.rm= T) +
  scale_x_continuous(limits = c(m - 3*s, m + 3*s)) +
  stat_function(fun = dnorm, args = list(mean = m, sd = s), color = "red", na.rm = TRUE) +
  labs(x="", y="") +
  ggtitle(num)

# save plot in list
lst[[i]] <- g }

return(lst)
}

# Arranging all plots to one

# males
lst_m <- hst(dfm)
lst_m[[18]] <- blankPanel<-grid.rect(gp=gpar(col="white")) # adding white space
gm <- do.call(grid.arrange, c(lst_m, list(ncol = 6, top = "Men")))
SAVE.PLOT(gm, "histograms_male.png", 10,15)

# females
lst_f <- hst(dff)
lst_f[[18]] <- blankPanel<-grid.rect(gp=gpar(col="white")) # adding white space
gf <- do.call(grid.arrange, c(lst_f, list(ncol = 6, top = "Women")))
SAVE.PLOT(gf, "histograms_female.png", 10,15)

# correlation matrices (SPLOM)
png(height = 4500, width = 6000, pointsize = 20, res = 300,
     file="./results/fig/corplot_males.png")

pairs.panels(
  setNames(dfm[, qvars], qvars_labels),
  pch = ".",
  gap = 0,
  method = "spearman",
  density = F,
  hist.col = "white",
  rug = F)
dev.off()

png(height = 4500, width = 6000, pointsize = 20, res = 300,
     file="./results/fig/corplot_females.png")

pairs.panels(
  setNames(dff[, qvars], qvars_labels),
  pch = ".",
  gap = 0,
  method = "spearman",
  density = F,
  hist.col = "white",
  rug = F)
dev.off()

# defining quantile-based skewness (Hao, 2007: 14)
qsk <- function(x){
  qsk_numerator <-
    quantile(x, probs = 0.90, na.rm = T) -
    quantile(x, probs = 0.5, na.rm = T)
  qsk_denominator <-
    quantile(x, probs = 0.5, na.rm = T) -
    quantile(x, probs = 0.1, na.rm = T)
  qsk <- (qsk_numerator / qsk_denominator) - 1
  return(qsk)}

# Skewness and qsk for all quantitative variables
melted <- melt(df[, c(qvars, "Geschlecht")], id.vars = "Geschlecht")

grouped <- group_by(melted, Geschlecht, variable)
tab <- summarise(grouped, skewness = skewness(value, na.rm = T), qsk=qsk(value))
tab <- cbind(tab[tab$Geschlecht == "Men", 2:ncol(tab)],
            tab[tab$Geschlecht == "Women", 3:ncol(tab)])

```

```

print(xtable(tab), include.rownames = F)

#####
# Box plots Vo2 by age class
#####

# Function because BPLs are also needed for validation
PLT.BPL <- function(DAT, REL = T, LAB = REL_LAB){

  # ageclass is needed as "factor"
  DF1 <- dply(DAT, c("ageclass", "Geschlecht"), transform, N = length(ageclass))
  DF1$label <- paste0(DF1$ageclass, "\n", "(n=", DF1$N, ")")

  # labels and axes
  if(REL == T){
    OUTCOME = "REL_VO2_MAX"
    LMT = c(0,60)
    BRK = seq(from=0,to=60,by=5)
  }else{
    OUTCOME = "VO2_MAX"
    LMT = c(0,6)
    BRK = seq(from = 0,to = 6,by = 0.5)
  }

  # absolute VO2peak
  BPL <- ggplot(DF1, aes_string(x = "label", y = OUTCOME)) +
    geom_boxplot(na.rm = T) +
    scale_y_continuous(limits = LMT, breaks = BRK) +
    facet_grid(.~Geschlecht, scales = "free")+
    labs(x = "",
         y = LAB)

  return(BPL)
}

SAVE.PLOT(grid.arrange(PLT.BPL(df, REL = F, LAB = ABS_LAB),
                      PLT.BPL(df, REL = T, LAB = REL_LAB),
                      ncol=1),
          "bpl.png", 7, 10)

```

### 7.3.5 Representativity

```

# read standard population
sb <- read.table("../data/Altersverteilung_Zensus_2011.csv",
                 header = T, sep = ";", dec = ",")

sb$Men <- sb$Male
sb$Women <- sb$Female
sb$Male <- sb$Female <- NULL

# same age groups for standard population and PF sample
df$a.standard <- cut(df$ALTER,
                    breaks = c(25,30,35,40,45,50,55,60,65,69),
                    include.lowest = TRUE, include.highest = F, right = F)

sb$a.standard <- cut(sb$Alter,
                    breaks = c(25,30,35,40,45,50,55,60,65,69),
                    include.lowest = TRUE, include.highest = F, right = F)

# restricting census age groups to the margins of the present sample
sb <- subset(sb, !is.na(a.standard))

# aggregating standard population by age group
sb <- aggregate(cbind(Men, Women) ~ a.standard, data = sb, sum)

# bootstrap function
ias <- function(dat, sex, var, b, standard = sb){

  set.seed(1)

  bt <- dat[b,]

```

```

# variables
repvars <- c("ZIGARETTEN",
            "EXRAUCHER",
            "BMI_GE_25",
            "obese",
            "BLUTHOCHDRUCK_WHO")

# aggregating PF sample by age group
events <- aggregate(
  x = bt[,repvars],
  by = bt[,c("a.standard", "Geschlecht")],
  function(x) sum(x=="yes", na.rm =T))

n_per_AG <- aggregate(
  x = bt[,repvars],
  by = bt[,c("a.standard", "Geschlecht")],
  function(x) sum(!is.na(x)))

e <- subset(events, Geschlecht == sex)
n <- subset(n_per_AG, Geschlecht == sex)

# event rate
er <- epitools::ageadjust.direct(
  e[,var], n[,var], stdpop = standard[,sex])

return(er*100)
}

# function to execute bootstrap for sexes and variables
bt <- function(sx, vr){

  b = boot::boot(
    data = df,
    sex = sx, var = vr, statistic = ias,
    R = NBOOT
  )

  return(
    c(b$t[2],
      quantile(b$t[,2], probs = c(0.025,0.975))
    )
  )
}

# Creating overall table
PF_MALE <- data.frame(

  Smoker      = bt("Men", "ZIGARETTEN"),
  Former_Smoker = bt("Men", "EXRAUCHER"),
  Overweight  = bt("Men", "BMI_GE_25"),
  Obesity     = bt("Men", "obese"),
  Hypertension = bt("Men", "BLUTHOCHDRUCK_WHO")

)

PF_FEMALE <- data.frame(

  Smoker      = bt("Women", "ZIGARETTEN"),
  Former_Smoker = bt("Women", "EXRAUCHER"),
  Overweight  = bt("Women", "BMI_GE_25"),
  Obesity     = bt("Women", "obese"),
  Hypertension = bt("Women", "BLUTHOCHDRUCK_WHO")

)

DEGS_MALE <- data.frame(

  Smoker      = c(26.1, 24.0, 28.2),
  Former_Smoker = c(33.7, 31.9, 35.5),
  Overweight  = c(67.1, 65.0, 69.2),
  Obesity     = c(23.3, 21.2, 25.4),
  Hypertension = c(33.3, 31.1, 35.6)

)

DEGS_FEMALE <- data.frame(

```



```

Smoker      = c(21.4, 19.7, 23.1),
Former_Smoker = c(22.8, 21.4, 24.2),
Overweight  = c(53.0, 50.8, 55.1),
Obesity     = c(23.9, 22.0, 25.9),
Hypertension = c(29.9, 28.1, 31.9)

)

TAB <- round(as.data.frame(cbind(

  t(PF_MALE),
  t(DEGS_MALE),
  t(PF_FEMALE),
  t(DEGS_FEMALE)

)), digits = 1)

LTAB <- cbind(
  paste0(TAB[,1], " (", TAB[,2], " to ", TAB[,3], ")"),
  paste0(TAB[,4], " (", TAB[,5], " to ", TAB[,6], ")"),
  paste0(TAB[,7], " (", TAB[,8], " to ", TAB[,9], ")"),
  paste0(TAB[,10], " (", TAB[,11], " to ", TAB[,12], ")")
)

LTAB <- as.data.frame(LTAB)
names(LTAB) <- c("PF_M", "DEGS_M", "PF_F", "DEGS_F")
row.names(LTAB) <- row.names(TAB)

# saving results as a list on hard drive
TAB_REPRESENTATIVITY <- list(

  DATA = TAB,
  LATEX = xtable(LTAB)

)

saveRDS(TAB_REPRESENTATIVITY, "./results/dat/tab_representativity.rds")
save.image()

# load(".RData")

```

### 7.3.6 Reference values

```

#####
# Quantiles
#####

# relative V02peak

# aggregating by TAUS and age class
tab1 <- aggregate(REL_V02_MAX ~ ageclass + Geschlecht,
  data = df,
  FUN = function(x) quantile(x, probs = TAUS))

tab2 <- aggregate(REL_V02_MAX ~ ageclass + Geschlecht,
  data = df,
  FUN = function(x) length(x))

tab <- do.call(data.frame, merge(tab1, tab2, by = c("Geschlecht", "ageclass")))
names(tab) <- c("Sex", "Age class", TAUS, "N")

print(xtable(tab,
  digits = c(0, 0, rep(1, 15))),
  include.rownames = F)

## PLAUSIBILITY CHECK:
# sum(df_FR$Geschlecht == "Female" & df_FR$ageclass == "[25,30)")
# quantile(subset(dfm_FR, ageclass == "[65,69]")$REL_V02_MAX, probs = TAUS)
# correct!

# absolute V02peak
tab1 <- aggregate(V02_MAX ~ ageclass + Geschlecht,
  data = df,

```

```

FUN = function(x) quantile(x, probs = TAUS))

tab2 <- aggregate(WO2_MAX ~ ageclass + Geschlecht,
  data = df,
  FUN = function(x) length(x))

tab <- do.call(data.frame, merge(tab1, tab2, by = c("Geschlecht", "ageclass")))
names(tab) <- c("Sex", "Age class", TAUS, "N")

print(xtable(tab,
  digits = c(0, 0, rep(2, 15))),
  include.rownames = F)

## PLAUSIBILITY CHECK:
# quantile(
# subset(df,
#   Geschlecht == "Male" &
#   ALTER >= 25 &
#   ALTER < 30)$REL_VO2_MAX,
# probs = c(0.05, 0.5, 0.95)
# )
## --> correct!

#####
# Nomograms: calculation data
#####

PLT.DAT <- function(dat, var){

  # empty list to save results
  lst <- list()

  # loop over all taus and calculate predicted values
  for(i in TAUS){

    # fit quantile regression
    fit <- rq(dat[,var] ~ ALTER + I(ALTER^2), data = dat, tau = i)

    # create empty prediction data set
    pdat <- data.frame(ALTER = seq(from = 25, to = 70, length.out = 1000))

    # calculate predicted values and confidence intervals, and adding tau
    set.seed(1)
    pdat <-
      cbind(pdat, tau = i,
            predict.rq(fit, newdata = pdat, interval = "confidence", level = .95,
                      se = "boot", type = "percentile", R = NBOOT)
            )

    # creating list with all results
    nam <- paste("pdat", i, sep = "")
    lst[[nam]] <- pdat
  }

  #####
  ## >pdat
  ## head(lst$pdat0.95)

  # ALTER tau fit lower higher
  #1 25.00000 0.95 52.41516 49.79706 54.98178
  #2 25.04505 0.95 52.40571 49.79676 54.96343
  #3 25.09009 0.95 52.39624 49.79642 54.94508

  ## --> all entries (taus) should be row-binded
  #####

  # row-binding for all prediction data frames (pdat)
  ds_overall <- do.call(rbind, lst)

  # express tau as perventile
  ds_overall$tau <- ds_overall$tau*100
  ds_overall$percentile <- paste(ds_overall$tau, "%", sep = "")

  return(ds_overall)
}# plt.dat function end

```

```

# Apply PLT.DAT to desired data sets

# list of all desired data sets
# --> 8 data sets --> perfect for parallelisation with 4 cpus
DSETS <- list(

# males and females with all observations
dfm = dfm,
dff = dff,

# males and female, smokers and obese excluded
# exm = exclusion of smokers&obese, male
exm = subset(dfm, ZIGARETTEN == "no" & obese == "no"),
exf = subset(dff, ZIGARETTEN == "no" & obese == "no"),

# including only participants from Frankfurt/Ruedesheim
# dfrm = df Ruedesheim, male
dfm_FR = dfm_FR,
dff_FR = dff_FR,

# including only participants from F/R and no smokers/obese
# exrm = exclusion of smoker&obese, Ruedesheim, male
exm_FR = subset(dfm_FR, ZIGARETTEN == "no" & obese == "no"),
exf_FR = subset(dff_FR, ZIGARETTEN == "no" & obese == "no")
)

# initiate clusters for parallel computing
cl <- makeCluster(mc <- getOption("cl.cores", 4))
clusterExport(cl=c1, varlist=ls())
clusterCall(cl, function() library("quantreg"))

ABS <- parLapply(cl, DSETS, function(x) PLT.DAT(x, "V02_MAX"))
REL <- parLapply(cl, DSETS, function(x) PLT.DAT(x, "REL_V02_MAX"))

stopCluster(cl)

# saving data sets (long computing time)
saveRDS(ABS, "./results/dat/nomodata_abs.rds")
saveRDS(REL, "./results/dat/nomodata_rel.rds")

save.image()
# load(".RData")

#####
# Nomograms: plotting
#####

# defining colors
COL <- c("dodgerblue", "dodgerblue1", "tomato3", "dodgerblue2", "dodgerblue3", "dodgerblue4",
        "tomato3",
        "dodgerblue4", "dodgerblue3", "tomato3", "dodgerblue2", "dodgerblue1", "dodgerblue")

# function for fitting of ggplots
PLT <- function(data_male, data_female, rel = T, ex = F){

# y-label
if(rel == T){
  ylb <- REL_LAB
  yby <- 1
  dig <- 0
} else {
  ylb <- ABS_LAB
  yby <- 0.1
  dig <- 1
}

# breaks are not dependent on a "from =" and "to =" argument
brk_fun <- function(k) {
  step <- k
  function(y) seq(floor(min(y)), ceiling(max(y)), by = step)
}

dm <- data_male
df <- data_female
dm$sex <- "Men"
df$sex <- "Women"
d <- rbind(dm, df)

```

```

ggp <- ggplot(data = d, aes(x = ALTER, y = fit)) +

# defining background of ggplot
theme(panel.grid.major = element_line(colour = "grey60", size = 0.3),
      panel.background = element_rect(fill = "white"),
      axis.text.x       = element_text(angle = 90, vjust=0.5),
      plot.title        = element_text(hjust = 0.5)) +
scale_x_continuous(limits = c(25,69),
                  breaks = seq(from = 25, to = 69, by = 1))+
scale_y_continuous(breaks = brk_fun(yby)) +
labs(x = "Age [years]",
     y = ylb) +
geom_vline(xintercept = c(30, 40, 50, 60),
          colour = "gray25",
          size = 0.6) +
facet_wrap(~sex, scales = "free_y", ncol = 1)

# adding confidence bands
ggp <- ggp +
  geom_ribbon(aes(x = ALTER, ymin = lower, ymax = higher, fill = percentile),
            alpha = 0.3, inherit.aes = F, show.legend = F, na.rm = T) +
  scale_fill_manual(values = COL)

# adding quantile curves
ggp <- ggp +
  geom_line(aes(x = ALTER, y = fit, colour = percentile),
           show.legend = F, size = 1, na.rm = T) +
  scale_color_manual(values = COL)

# adding percentile labels to curves
ggp <- direct.label(ggp, list("last.polygons", colour = "white", cex=0.65))

return(ggp)
}

# saving plots

# Only Frankfurt and Ruedesheim (Munich is for external validation)
SAVE.PLOT(PLT(REL$dfrm_FR, REL$dff_FR, rel = T, ex = F), "nomo_rel_include_FR.png", 10.5, 7)
SAVE.PLOT(PLT(ABS$dfrm_FR, ABS$dff_FR, rel = F, ex = F), "nomo_abs_include_FR.png", 10.5, 7)
SAVE.PLOT(PLT(REL$exm_FR, REL$exf_FR, rel = T, ex = T), "nomo_rel_exclude_FR.png", 10.5, 7)
SAVE.PLOT(PLT(ABS$exm_FR, ABS$exf_FR, rel = F, ex = T), "nomo_abs_exclude_FR.png", 10.5, 7)

# Frankfurt, Ruedesheim and Munich included
SAVE.PLOT(PLT(REL$dfrm, REL$dff, rel = T, ex = F), "nomo_rel_include_OVERALL.png", 10.5, 7)
SAVE.PLOT(PLT(ABS$dfrm, ABS$dff, rel = F, ex = F), "nomo_abs_include_OVERALL.png", 10.5, 7)
SAVE.PLOT(PLT(REL$exm, REL$exf, rel = T, ex = T), "nomo_rel_exclude_OVERALL.png", 10.5, 7)
SAVE.PLOT(PLT(ABS$exm, ABS$exf, rel = F, ex = T), "nomo_abs_exclude_OVERALL.png", 10.5, 7)
dev.off()

#####
# Nomograms: Coefficient tables
#####

# calculating all coefficients and all 95% CIs for coefficients
COEF <- function(DAT, OUTCOME){

# two empty lists to save coefficients and ci (lower and higher CI)
COEF <- list()
CIL <- list()
CIH <- list()

# loop over all taus and calculate predicted values
for(i in TAUS){

# name
NAM <- paste0("fit", i)

# fit quantile regression
FIT <- rq(DAT[,OUTCOME] ~ ALTER + I(ALTER^2), data = DAT, tau = i)

# bootstrapped confidence intervals for coefficients
BT <- boot.rq(cbind(1, DAT[, "ALTER"], I(DAT[, "ALTER"]^2)),
             DAT[, OUTCOME],
             tau = i, R = NBOOT)

```

```

COEF[[NAM]] <- FIT$coefficients
CIL[[NAM]] <- apply(BT$B, 2, quantile, c(0.025))
CIH[[NAM]] <- apply(BT$B, 2, quantile, c(0.975))

}# End of loop

return(list(COEF = COEF, CIL = CIL, CIH = CIH))
}

# calculating coefficients for all data sets

# initiate clusters for parallel computing
c1 <- makeCluster(mc <- getOption("cl.cores", 4))
clusterExport(c1=c1, varlist=ls())
clusterCall(c1, function() library("quantreg"))

COEF.ABS <- parLapply(c1, DSETS, function(x) COEF(x, "VO2_MAX"))
COEF.REL <- parLapply(c1, DSETS, function(x) COEF(x, "REL_VO2_MAX"))

stopCluster(c1)

# saving data sets (long computing time)
saveRDS(COEF.ABS, "./results/dat/coef_abs.rds")
saveRDS(COEF.REL, "./results/dat/coef_rel.rds")

save.image()

# creating tables for all data sets
COEF.TAB <- function(DAT){

  C <- round(t(do.call(data.frame, DAT$COEF)), digits = 8)
  L <- round(t(do.call(data.frame, DAT$CIL)), digits = 8)
  H <- round(t(do.call(data.frame, DAT$CIH)), digits = 8)

  # for coefficient plots
  TAB <- cbind(C[,1], L[,1], H[,1],
              C[,2], L[,2], H[,2],
              C[,3], L[,3], H[,3])

  TAB <- as.data.frame(TAB)

  # Coefficients + CI, I = Intercept, a = age, a2 = age^2
  # names are needed in this format for reshape using varying (coef plots)
  names(TAB) <- c(
    "c.1", "cil.1", "cih.1", "c.2", "cil.2", "cih.2", "c.3", "cil.3", "cih.3")

  TAB$FIT <- row.names(C)
  row.names(TAB) <- NULL

  # for Latex, FIT is first col
  LTAB <- cbind(TAB$FIT, TAB[, 1:9])
  names(LTAB) <- c(
    "Fit",
    "intercept", "cil.intercept", "cih.intercept",
    "age", "cil.age", "cih.age",
    "age2", "cil.age2", "cih.age2")
  LTAB <- xtable(LTAB, digits = c(0,0,rep(8,9)))

  return(list(LTAB = LTAB, DATA = TAB))
}

# Create tables

# Absolute VO2max

# All cases
print(COEF.TAB(COEF.ABS$dfm)$LTAB, include.rownames = FALSE)
print(COEF.TAB(COEF.ABS$dff)$LTAB, include.rownames = FALSE)

# exclusion of smokers and obese
print(COEF.TAB(COEF.ABS$exm)$LTAB, include.rownames = FALSE)
print(COEF.TAB(COEF.ABS$exf)$LTAB, include.rownames = FALSE)

# Frankfurt/Ruedesheim (FR)

```

```

print(COEF.TAB(COEF.ABS$dfm_FR)$LTAB, include.rownames = FALSE)
print(COEF.TAB(COEF.ABS$dff_FR)$LTAB, include.rownames = FALSE)

# Frankfurt/Ruedesheim (FR)
print(COEF.TAB(COEF.ABS$exm_FR)$LTAB, include.rownames = FALSE)
print(COEF.TAB(COEF.ABS$exf_FR)$LTAB, include.rownames = FALSE)

# Relative V02max

# All cases
print(COEF.TAB(COEF.REL$dfm)$LTAB, include.rownames = FALSE)
print(COEF.TAB(COEF.REL$dff)$LTAB, include.rownames = FALSE)

# exclusion of smokers and obese
print(COEF.TAB(COEF.REL$exm)$LTAB, include.rownames = FALSE)
print(COEF.TAB(COEF.REL$exf)$LTAB, include.rownames = FALSE)

# Frankfurt/Ruedesheim (FR)
print(COEF.TAB(COEF.REL$dfm_FR)$LTAB, include.rownames = FALSE)
print(COEF.TAB(COEF.REL$dff_FR)$LTAB, include.rownames = FALSE)

# Frankfurt/Ruedesheim (FR)
print(COEF.TAB(COEF.REL$exm_FR)$LTAB, include.rownames = FALSE)
print(COEF.TAB(COEF.REL$exf_FR)$LTAB, include.rownames = FALSE)

#####
# coefficient plots
#####

PLT.COEF <- function(DAT_MALE = COEF.ABS$dfm, DAT_FEMALE = COEF.ABS$dff){

  DM <- COEF.TAB(DAT_MALE)$DATA
  DF <- COEF.TAB(DAT_FEMALE)$DATA

  # adding sex-variable
  DM$Sex <- "Men"
  DF$Sex <- "Women"

  # Rbindung both sexes
  COEF.DATA <- rbind(DM, DF)

  # >COEF.DATA
  # colnames are: c1 = Intercept, c2 = age, c3 = age^2
  #
  #           c.1    cil.1    cih.1    c.2    cil.2
  # fit0.05 2.148064 1.2871575 2.833846 0.0220968 -0.0071967 ...
  # fit0.1  2.283736 1.3166107 2.540710 0.0250659 0.0004518 ...
  # fit0.2  2.407500 1.2070498 2.863562 0.0309868 -0.0089181 ...
  # ...

  LONG <- reshape(COEF.DATA, dir="long", varying = 1:9, idvar = c("FIT", "Sex"))
  LONG$time <- factor(LONG$time, levels = c(1,2,3),
                     labels = c("Intercept", "Age", expression(Age^"2")))

  # >LONG
  # FIT    Sex    time c      cil      cih
  # fit0.05 Male  Age  2.1480645  1.1823432  3.1150844
  # fit0.1  Male  Age  2.2837363  1.4854133  2.9921140
  # fit0.2  Male  Age  2.4075000  1.9149587  2.6529048
  # ...

  PLT <- ggplot(LONG) +
    geom_point(aes(x = FIT, y = c)) +
    geom_linerange(aes(x = FIT, ymin = cil, ymax = cih)) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1),
          plot.title = element_text(hjust = 0.5)) +
    facet_wrap(Sex ~ time, scales = "free", labeller = label_parsed) +
    geom_hline(yintercept = 0, lty = 2) +
    labs(y = "Coefficient (95% CI)", x = "Quantile regression")

  return(PLT)
}

# Create plots

# Absolute V02max

```

```

# All cases
SAVE.PLOT(PLT.COEF(COEF.ABS$dfm,COEF.ABS$dff), "coefplot_abs_OVERALL.png", 4, 8)
# exclusion of smokers and obese
SAVE.PLOT(PLT.COEF(COEF.ABS$exm,COEF.ABS$exf), "coefplot_abs_EX.png", 4, 8)
# Frankfurt/Ruedesheim (FR)
SAVE.PLOT(PLT.COEF(COEF.ABS$dfm_FR,COEF.ABS$dff_FR), "coefplot_abs_FR.png", 4, 8)

# Relative V02max

# All cases
SAVE.PLOT(PLT.COEF(COEF.REL$dfm,COEF.REL$dff), "coefplot_rel_OVERALL.png", 4, 8)
# exclusion of smokers and obese
SAVE.PLOT(PLT.COEF(COEF.REL$exm,COEF.REL$exf), "coefplot_rel_EX.png", 4, 8)
# Frankfurt/Ruedesheim (FR)
SAVE.PLOT(PLT.COEF(COEF.REL$dfm_FR,COEF.REL$dff_FR), "coefplot_rel_FR.png", 4, 8)

```

### 7.3.7 Validation

```

#####
# calculating validation data
#####

VALI.DAT <- function(TRN.DAT = dfm_FR, VAL.DAT = dfm_FR, OUTCOME = "REL_V02_MAX",
                    FORMULA = "ageclass + I(ageclass^2)") {

  # GOALS: producing all data for validation

  # G1) empirical quantiles (EMP.QTLS) from validation data (VAL.DAT)
  # G2) quantile regression in training data (TRN.DAT)
  # G3) predicted values from quantile regressions
  # G4) 95% CI for predicted values
  # G5) AIC table for each TAUS2
  # G6) comparing empirical and estimated values in separate function

  # quantiles for validation
  TAUS2 <- c(0.25,0.5,0.75)

  # GOAL 1

  # empirical quantiles: tau = {0.25,0.5,0.75}
  FRM1 <- as.formula(paste(OUTCOME, "~ ageclass + Geschlecht"))
  EMP.QTLS <- aggregate(FRM1, data = VAL.DAT,
                      FUN = function(x) quantile(x, probs = TAUS2))
  EMP.QTLS <- do.call(data.frame, EMP.QTLS)
  names(EMP.QTLS) <- c("ageclass", "Geschlecht", "tau0.25", "tau0.5", "tau0.75")
  EMP.QTLS <- melt(EMP.QTLS, id.vars = c("ageclass", "Geschlecht"))

  # > EMP.QTLS
  # ageclass Geschlecht variable value
  # [25,30) Male tau0.25 33.525
  # [30,35) Male tau0.25 33.400
  # [35,40) Male tau0.25 32.700
  # ...

  # GOAL 2

  # ageclass variable (from training data) for regression
  TRN.DAT$ageclass <- as.numeric(TRN.DAT$ageclass)

  # formula for quantile regressions
  FRM2 <- as.formula(paste(OUTCOME, "~", FORMULA))

  PDAT <- list()
  AIC.DAT <- list()

  # loop over all taus and calculate predicted values
  for(i in TAUS2){

    FIT <- rq(FRM2, data = TRN.DAT, tau = i)

    # create empty prediction data set
    pdat <- data.frame(ageclass = 1:9)

    # function to predict value and 95% CI

```

```

PRED <- function(){

  set.seed(1)
  PV <- predict.rq(FIT,
                  newdata = pdat,
                  interval = "confidence",
                  se = "boot", type = "percentile",
                  level = .95, R = NBOOT)

  return(PV)
}

# calculate predicted values and confidence intervals, and adding tau
pdat <- cbind(pdat,
             tau = paste0("tau",i),
             PRED())
)
# > pdat
# ageclass tau fit lower higher
# 1 tau0.50 39.30000 38.31400 41.09600
# 2 tau0.50 38.83333 38.24494 39.88267
# 3 tau0.50 38.00000 37.70000 38.49750
# ...

# column must be the same name as outcome
# in order to add the dots in the box plot
pdat[,OUTCOME] <- pdat$fit

# saving prediction data
nam1 <- paste0("pdat", i)
PDAT[[nam1]] <- pdat

# saving aic data
nam2 <- paste0("aic", i)
AIC.DAT[[nam2]] <- AIC.rq(FIT)

}

PDAT <- do.call(rbind, PDAT)
row.names(PDAT) <- NULL

AIC.DAT <- do.call(rbind, AIC.DAT)
# > test$AIC.DAT
# [ ,1]
# aic0.25 32672.15
# aic0.5 32406.82
# aic0.75 33242.66

return(list(EMP.QTLS = EMP.QTLS, PDAT = PDAT, AIC.DAT = AIC.DAT))
}

# saving all data
# Nomenclature:
# A = a = apparent, e = external
# M = male, F = female
# R = REL_V02_MAX, A = ABS_V02_MAX

# -----
# Main models: Quadratic
# -----

# apparent validation
AMR <- VALI.DAT(dfm_FR, dfm_FR, "REL_V02_MAX", "ageclass + I(ageclass^2)")
AFR <- VALI.DAT(dff_FR, dff_FR, "REL_V02_MAX", "ageclass + I(ageclass^2)")
AMA <- VALI.DAT(dfm_FR, dfm_FR, "V02_MAX", "ageclass + I(ageclass^2)")
AFA <- VALI.DAT(dff_FR, dff_FR, "V02_MAX", "ageclass + I(ageclass^2)")

# external validation
EMR <- VALI.DAT(dfm_FR, dfm_M, "REL_V02_MAX", "ageclass + I(ageclass^2)")
EFR <- VALI.DAT(dff_FR, dff_M, "REL_V02_MAX", "ageclass + I(ageclass^2)")
EMA <- VALI.DAT(dfm_FR, dfm_M, "V02_MAX", "ageclass + I(ageclass^2)")
EFA <- VALI.DAT(dff_FR, dff_M, "V02_MAX", "ageclass + I(ageclass^2)")

# -----
# Other models: Linear
# -----

# apparent validation
AMR_lin <- VALI.DAT(dfm_FR, dfm_FR, "REL_V02_MAX", "ageclass")

```



```

AFR_lin <- VALI.DAT(dff_FR, dff_FR, "REL_VO2_MAX", "ageclass")
AMA_lin <- VALI.DAT(dfm_FR, dfm_FR, "VO2_MAX", "ageclass")
AFA_lin <- VALI.DAT(dff_FR, dff_FR, "VO2_MAX", "ageclass")

# -----
# Other models: spline
# -----

# apparent validation
AMR_spl <- VALI.DAT(dfm_FR, dfm_FR, "REL_VO2_MAX", "bs(ageclass)")
AFR_spl <- VALI.DAT(dff_FR, dff_FR, "REL_VO2_MAX", "bs(ageclass)")
AMA_spl <- VALI.DAT(dfm_FR, dfm_FR, "VO2_MAX", "bs(ageclass)")
AFA_spl <- VALI.DAT(dff_FR, dff_FR, "VO2_MAX", "bs(ageclass)")

save.image()
# load(".RData")

#####
# validation box plots
#####

# function for adding the estimates or CI to the box plot
ADD.LINES <- function(DAT = DAT, COLOUR = "blue", CI = F, ALPHA = 1){

  # adding confidence bands
  if(CI == T){
    BPL <- BPL +
      geom_linerange(data = DAT,
                    aes(x = ageclass, ymin = lower, ymax = higher, group = tau),
                    size = 7, colour = "red", alpha = 0.5)
  }

  BPL <- BPL +
    geom_line(data = DAT,
             stat = "smooth", method = "loess",
             aes(x = ageclass, y = fit, group = tau),
             size = 0.5, colour = COLOUR, alpha = ALPHA,
             na.rm = T) +
    geom_point(data = DAT, aes(x = ageclass, y = fit, shape = factor(tau)),
              size = 2, colour = COLOUR, alpha = ALPHA) +
    theme(legend.position="none")

  return(BPL)
}

# -----
# Apparent validation
# -----

# rel, males
BPL <- PLT.BPL(dfm_FR, REL = T, LAB = REL_LAB)
BPL1 <- ADD.LINES(DAT = AMR$PDAT, COLOUR = "blue", CI = T)

# rel, females
BPL <- PLT.BPL(dff_FR, REL = T, LAB = "")
BPL2 <- ADD.LINES(DAT = AFR$PDAT, COLOUR = "blue", CI = T)

# abs, males
BPL <- PLT.BPL(dfm_FR, REL = F, LAB = ABS_LAB)
BPL3 <- ADD.LINES(DAT = AMA$PDAT, COLOUR = "blue", CI = T)

# abs, females
BPL <- PLT.BPL(dff_FR, REL = F, LAB = "")
BPL4 <- ADD.LINES(DAT = AFA$PDAT, COLOUR = "blue", CI = T)

SAVE.PLOT(grid.arrange(BPL1,BPL2,BPL3,BPL4, ncol = 2),
          "vali_apparent.png", 8, 10.5)

# -----
# External validation
# -----

# rel, males
BPL <- PLT.BPL(dfm_M, REL = T, LAB = REL_LAB)
BPL1 <- ADD.LINES(DAT = EMR$PDAT, COLOUR = "blue", CI = T)

# rel, females
BPL <- PLT.BPL(dff_M, REL = T, LAB = "")

```

```

BPL2 <- ADD.LINES(DAT = EFR$PDAT, COLOUR = "blue", CI = T)

# abs, males
BPL <- PLT.BPL(dfm_M, REL = F, LAB = ABS_LAB)
BPL3 <- ADD.LINES(DAT = EMA$PDAT, COLOUR = "blue", CI = T)

# abs, females
BPL <- PLT.BPL(dff_M, REL = F, LAB = "")
BPL4 <- ADD.LINES(DAT = EFA$PDAT, COLOUR = "blue", CI = T)

SAVE.PLOT(grid.arrange(BPL1,BPL2,BPL3,BPL4, ncol = 2),
           "vali_external.png", 8, 10.5)

# -----
# Apparent validation including linear and spline models
# -----

# rel, males
BPL <- PLT.BPL(dfm_FR, REL = T, LAB = REL_LAB)
BPL <- ADD.LINES(DAT = AMR_lin$PDAT, COLOUR = "red", CI = F, ALPHA = 0.5)
BPL <- ADD.LINES(DAT = AMR$PDAT, COLOUR = "blue", CI = F, ALPHA = 0.5)
BPL1 <- ADD.LINES(DAT = AMR_spl$PDAT, COLOUR = "green", CI = F, ALPHA = 0.5)

# rel, females
BPL <- PLT.BPL(dff_FR, REL = T, LAB = "")
BPL <- ADD.LINES(DAT = AFR_lin$PDAT, COLOUR = "red", CI = F, ALPHA = 0.5)
BPL <- ADD.LINES(DAT = AFR$PDAT, COLOUR = "blue", CI = F, ALPHA = 0.5)
BPL2 <- ADD.LINES(DAT = AFR_spl$PDAT, COLOUR = "green", CI = F, ALPHA = 0.5)

# abs, males
BPL <- PLT.BPL(dfm_FR, REL = F, ABS_LAB)
BPL <- ADD.LINES(DAT = AMA_lin$PDAT, COLOUR = "red", CI = F, ALPHA = 0.5)
BPL <- ADD.LINES(DAT = AMA$PDAT, COLOUR = "blue", CI = F, ALPHA = 0.5)
BPL3 <- ADD.LINES(DAT = AMA_spl$PDAT, COLOUR = "green", CI = F, ALPHA = 0.5)

# abs, females
BPL <- PLT.BPL(dff_FR, REL = F, LAB = "")
BPL <- ADD.LINES(DAT = AFA_lin$PDAT, COLOUR = "red", CI = F, ALPHA = 0.5)
BPL <- ADD.LINES(DAT = AFA$PDAT, COLOUR = "blue", CI = F, ALPHA = 0.5)
BPL4 <- ADD.LINES(DAT = AFA_spl$PDAT, COLOUR = "green", CI = F, ALPHA = 0.5)

SAVE.PLOT(grid.arrange(BPL1,BPL2,BPL3,BPL4, ncol = 2),
           "vali_apparent_comparison.png", 8, 10.5)

#####
# validation calibration plots
#####

# build basic calibration plot without lines
PLT.CALI <- function(REL = T, CITY = "Frankfurt/RÃdesheim"){

  if(REL == T){ MIN = 20; MAX = 50; BY = 5 }else{
    MIN = 1 ; MAX = 4 ; BY = 0.5}

  lab <- paste("Observed \n in", CITY)

  P <- ggplot(data = D, aes(x = fit, y = value)) +
    geom_abline(intercept = 0, slope = 1, lty = 3) +
    theme(legend.position = "none") +
    scale_x_continuous(limits = c(MIN,MAX), breaks = seq(MIN, MAX, by = BY)) +
    scale_y_continuous(limits = c(MIN,MAX), breaks = seq(MIN, MAX, by = BY)) +
    labs(x = "Prediction \n based on Frankfurt/RÃdesheim", y = lab)

  return(P)
}

# add regression lines and dots
LINE.CALI <- function(X = AMR$PDAT, Y = AMR$EMP.QTLS, PLOT = CP, COLOUR = "blue", ALPHA = 1){

  Y$ageclass <- as.numeric(Y$ageclass)

  D <- merge(X, Y,
            by.x = c("ageclass", "tau"),
            by.y = c("ageclass", "variable"))

  CP <- CP +
    geom_line(stat = "smooth", method = "lm", se = F,

```

```

    data = D,
    aes(x = fit, y = value),
    colour = COLOUR, alpha = ALPHA, na.rm = T) +
  geom_point(data = D, aes(x = fit, y = value, shape = tau),
    colour = COLOUR, alpha = ALPHA, na.rm = T) +
  facet_grid(~Geschlecht)

return(CP)
}

# -----
# Apparent validation
# -----

CP <- PLT.CALI(REL = T)
CP1 <- LINE.CALI(AMR$PDAT, AMR$EMP.QTLS)
CP2 <- LINE.CALI(AFR$PDAT, AFR$EMP.QTLS)
CP <- PLT.CALI(REL = F)
CP3 <- LINE.CALI(AMA$PDAT, AMA$EMP.QTLS)
CP4 <- LINE.CALI(AFA$PDAT, AFA$EMP.QTLS)
SAVE.PLOT(grid.arrange(
  arrangeGrob(CP1,CP2, top = textGrob(REL_LAB)),
  arrangeGrob(CP3,CP4, top = textGrob(ABS_LAB)),
  ncol = 2), "cali_apparent.png", 10,8)

# -----
# External validation
# -----

CP <- PLT.CALI(REL = T, CITY = "Munich")
CP1 <- LINE.CALI(EMR$PDAT, EMR$EMP.QTLS)
CP2 <- LINE.CALI(EFR$PDAT, EFR$EMP.QTLS)
CP <- PLT.CALI(REL = F, CITY = "Munich")
CP3 <- LINE.CALI(EMA$PDAT, EMA$EMP.QTLS)
CP4 <- LINE.CALI(EFA$PDAT, EFA$EMP.QTLS)
SAVE.PLOT(grid.arrange(
  arrangeGrob(CP1,CP2, top = textGrob(REL_LAB)),
  arrangeGrob(CP3,CP4, top = textGrob(ABS_LAB)),
  ncol = 2), "cali_external.png", 10,8)

# -----
# Apparent validation comparison of regression models
# -----

CP <- PLT.CALI(REL = T) # Males, REL_VO2_MAX
CP <- LINE.CALI(AMR$PDAT, AMR$EMP.QTLS, ALPHA = 0.5, COLOUR = "blue")
CP <- LINE.CALI(AMR_lin$PDAT, AMR_lin$EMP.QTLS, ALPHA = 0.5, COLOUR = "red")
CP1 <- LINE.CALI(AMR_sp1$PDAT, AMR_sp1$EMP.QTLS, ALPHA = 0.5, COLOUR = "green")
CP <- PLT.CALI(REL = T) # Females REL_VO2_MAX
CP <- LINE.CALI(AFR$PDAT, AFR$EMP.QTLS, ALPHA = 0.5, COLOUR = "blue")
CP <- LINE.CALI(AFR_lin$PDAT, AFR_lin$EMP.QTLS, ALPHA = 0.5, COLOUR = "red")
CP2 <- LINE.CALI(AFR_sp1$PDAT, AFR_sp1$EMP.QTLS, ALPHA = 0.5, COLOUR = "green")
CP <- PLT.CALI(REL = F) # Males VO2_MAX
CP <- LINE.CALI(AMA$PDAT, AMA$EMP.QTLS, ALPHA = 0.5, COLOUR = "blue")
CP <- LINE.CALI(AMA_lin$PDAT, AMA_lin$EMP.QTLS, ALPHA = 0.5, COLOUR = "red")
CP3 <- LINE.CALI(AMA_sp1$PDAT, AMA_sp1$EMP.QTLS, ALPHA = 0.5, COLOUR = "green")
CP <- PLT.CALI(REL = F) # Females VO2_MAX
CP <- LINE.CALI(AFA$PDAT, AFA$EMP.QTLS, ALPHA = 0.5, COLOUR = "blue")
CP <- LINE.CALI(AFA_lin$PDAT, AFA_lin$EMP.QTLS, ALPHA = 0.5, COLOUR = "red")
CP4 <- LINE.CALI(AFA_sp1$PDAT, AFA_sp1$EMP.QTLS, ALPHA = 0.5, COLOUR = "green")
SAVE.PLOT(grid.arrange(
  arrangeGrob(CP1,CP2, top = textGrob(REL_LAB)),
  arrangeGrob(CP3,CP4, top = textGrob(ABS_LAB)),
  ncol = 2), "cali_apparent_comparison.png", 10,8)

#####
# validation AIC tables
#####

# Males
AIC.TAB <- rbind(

  # females
  cbind(AFR_lin$AIC.DAT, AFR_sp1$AIC.DAT, AFR$AIC.DAT,
    AFA_lin$AIC.DAT, AFA_sp1$AIC.DAT, AFA$AIC.DAT),

  # males
  cbind(AMR_lin$AIC.DAT, AMR_sp1$AIC.DAT, AMR$AIC.DAT,
    AMA_lin$AIC.DAT, AMA_sp1$AIC.DAT, AMA$AIC.DAT)
)

```

```

)

AIC.TAB <- as.data.frame(AIC.TAB)
names(AIC.TAB) <- c("AL", "AS", "AP", "RL", "RS", "RP")
row.names(AIC.TAB) <- c("F.25", "F.5", "F.75", "M.25", "M.5", "M.75")
xtable(AIC.TAB)

#####
# validation coefficient tables
#####

COEF.TAB2 <- function(DAT){

  X <- DAT["PDAT"]; X <- X$PDAT$fit
  Y <- DAT["EMP.QTLS"]; Y <- Y$EMP.QTLS$value

  D <- data.frame(X, Y)

  COEF.BT <- function(data, INDEX){
    set.seed(1)
    D1 <- data[INDEX,]
    FIT <- lm(Y ~ X, data = D1)
    ROW <- c(FIT$coefficients, summary(FIT)$r.squared)
  }

  BT <- boot(data = D, statistic = COEF.BT, R = NBOOT)
  TAB <- rbind(BT$t0, apply(BT$t, 2, quantile, c(0.025, 0.975)))
  TAB <- round(t(as.data.frame(TAB)), digits = 2)
  COL <- paste0(TAB[,1], " [", TAB[,2], " to ", TAB[,3], "]" )
  COL <- data.frame(COL = COL)
  row.names(COL) <- c("Intercept", "Slope", "Rsquared")

  #> COL
  #Intercept 5.94 [3.11 to 8.69]
  #Slope 0.83 [0.76 to 0.90]
  #Rsquared 0.92 [0.88 to 0.96]

  return(COL)
}

# -----
# Apparent validation
# -----

# CAVE: females on top, absolute on the left column

TAB <- rbind(
  cbind(COEF.TAB2(AFA), COEF.TAB2(AFR)),
  cbind(COEF.TAB2(AMA), COEF.TAB2(AMR))
)
names(TAB) <- c("Absolute", "Relative")
xtable(TAB)

# -----
# External validation
# -----

TAB <- rbind(
  cbind(COEF.TAB2(EFA), COEF.TAB2(EFR)),
  cbind(COEF.TAB2(EMA), COEF.TAB2(EMR))
)
names(TAB) <- c("Absolute", "Relative")
xtable(TAB)

# -----
# Apparent validation comparison of regression models
# -----

TAB <- rbind(
  # female
  cbind(# absolute
        COEF.TAB2(AFA_lin), COEF.TAB2(AFA_spl), COEF.TAB2(AFA),
        # relative
        COEF.TAB2(AFR_lin), COEF.TAB2(AFR_spl), COEF.TAB2(AFR)),
  # male
  cbind(# absolute
        COEF.TAB2(AMA_lin), COEF.TAB2(AMA_spl), COEF.TAB2(AMA),
        # relative

```

```

      COEF.TAB2(AMR_lin), COEF.TAB2(AMR_sp1), COEF.TAB2(AMR)
)
names(TAB) <- c("AL", "AS", "AP", "RL", "RS", "RP")
xtable(TAB)

```

### 7.3.8 Multiple quantile regression

```

# histogram robust for outliers

h <- function(var){

  # set margins of histogram
  m <- mean(df[,var], na.rm = T)
  s <- sd(df[,var], na.rm = T)

  # log transformed variable
  d <- df
  d$l <- log(df[,var])
  m1 <- mean(d$l, na.rm = T)
  s1 <- sd(d$l, na.rm = T)

  # histogram with crude variable
  c <- ggplot(df, aes_string(x = var)) +
    geom_histogram(aes(y = ..density..), bins = 20, na.rm = T) +
    scale_x_continuous(limits = c(m - 3*s, m + 3*s)) +
    facet_wrap(~Geschlecht) +
    labs(title = "crude")

  # histogram with log-transformed variable
  l <- ggplot(d, aes(x = l)) +
    geom_histogram(aes(y = ..density..), bins = 20, na.rm = T) +
    scale_x_continuous(limits = c(m1 - 3*s1, m1 + 3*s1)) +
    facet_wrap(~Geschlecht) +
    labs(title = "log-transformed")

  gg <- grid.arrange(c,l)
  return(gg)
}

h("BMI") # --> transform --> use overweight/obese
h("GEWICHT") # --> transform --> exclude due to collinearity
h("TRIGLYCERIDE") # --> transform: binary
h("BAUCHUMFANG") # transform -->
h("BLUTZUCKER") # --> transform --> binary as Diabetes_mellitus
h("HBA1C") # --> transform --> binary as Diabetes_mellitus
h("SYST") # --> no transformation
h("DIAST") # --> no transformation

# checking correlation of variables
# remove when r>=0.75 to avoid collinearity
# based on SPLOM
# --> remove:
# GEWICHT, BAUCHUMFANG, KOERPERFETT_CALIPER & TANITA,
# GESAMTCHOLESTERIN

# calculating new variables
df$Age_squared = df$ALTER^2
df$log_BMI = log(df$BMI)
df$log_triglycerides = log(df$TRIGLYCERIDE)
df$log_blood_glucose = log(df$BLUTZUCKER)
df$log_HbA1c = log(df$HBA1C)

df$High_LDL = ifelse(df$LDL_CHOLESTERIN >=115, 1, 0)
df$High_TG = ifelse(df$TRIGLYCERIDE >=150, 1, 0)
df$Low_HDL = ifelse(df$HDL_CHOLESTERIN <=40, 1, 0)
df$Hypertension = ifelse(df$SYST >= 140 | df$DIAST >= 90, 1, 0)

# renaming variables (for tick labels)
df$Height <- df$GROESSE
df$Age <- df$ALTER
df$Systolic <- df$SYST
df$Diastolic <- df$DIAST
df$HDL_cholesterol <- df$HDL_CHOLESTERIN

```

```

df$LDL_cholesterol <- df$LDL_CHOLESTERIN

# recoding factor variables
# --> necessary for following function
# --> coding as no = 1, yes = 2
BIN <- function(x){
  var <- as.numeric(df[,x])
  var <- ifelse(var == 1, 0, ifelse(var == 2, 1, NA))
  return(var)
}
df$Smoker <- BIN("ZIGARETTEN")
df$Exsmoker <- BIN("EXRAUCHER")
df$Overweight <- BIN("BMI_GE_25")
df$Obese <- BIN("obese")
df$Diabetes_mellitus <- BIN("MANIFEST_DM")

# Defining final independent variables
IV <- c(
  "Age", "Height",
  "Overweight", "Obese",
  "Hypertension", "Low_HDL", "High_LDL", "High_TG",
  "Diabetes_mellitus", "Smoker", "Exsmoker"
)

# removing all cases with NAs in data set
df$exclusion <- rowSums(is.na(df[,c(IV,"V02_MAX", "REL_V02_MAX)]))
df1 <- subset(df, exclusion == 0)

# subsetting data including new IVs
dfm <- subset(df1, Geschlecht == "Men")
dff <- subset(df1, Geschlecht == "Women")

# Regression fitting
QR <- function(dat, outcome, qtile) {

  # creating formula
  # IV have to be defined!
  frm <- as.formula(paste(outcome, "~", paste(IV, collapse = "+")))

  # fitting quantile regression
  fit <- step(rq(frm, data = dat, tau = qtile, direction = "both"))

  # bootstrapped P values
  # NBOOT has to be defined!
  set.seed(1)
  tab <- summary(fit, se = "boot", R = NBOOT)

  # extract all variables that are in the formula of final model and remove dependent var
  var <- setdiff(all.vars(tab$call$formula), outcome)

  # bootstrapping 95% confidence intervals
  set.seed(1)
  bt <- boot.rq(cbind(1, dat[, var]), dat[, outcome], tau = qtile, R = NBOOT)

  # add bootstrapped 95% CI to table
  tab <- cbind(
    tab$coefficients[, c(1,4)],
    t(apply(bt$B, 2, quantile, c(0.025,0.975))))

  # are coefficients within ci bounds?
  tab <- as.data.frame(tab)
  tab <- tab[, c("Value", "2.5%", "97.5%", "Pr(>|t|)")]

  coef_in_ci <- data.frame(
    hi = tab$Value<tab$`97.5%`,
    lo = tab$Value>tab$`2.5%`)

  # do 95% CI and P value lead to the same conclusion (sig. yes/no)?
  sig <- data.frame(
    pval = tab[,4] < 0.05,
    ci = (tab[,2] > 0) == (tab[,3] > 0)
  )
  sig$SAME_CONCLUSION <- sig$pval == sig$ci

  # calculation of R^2 according to
  # Hao, 2007, p. 52
  # https://stat.ethz.ch/pipermail/r-help/2006-August/110386.html

```

```

# http://stats.stackexchange.com/questions/129200/r-squared-in-quantile-regression
fit0 <- rq(REL_VO2_MAX ~ 1, tau = qtile, data = dat)
rho <- function(u,tau=qtile)u*(tau - (u < 0))
R1 <- round(1 - fit0$rho/fit0$rho, digits = 3)

# creating name for coefficient plot
# sex
if(dat[3,"Geschlecht"] == "Men") sex <- "Men"
if(dat[3,"Geschlecht"] == "Women") sex <- "Women"

# return the name VO2_MAX
out <- colnames(dat)[which(colnames(dat) %in% outcome)]

# build name
if(out == "REL_VO2_MAX") {
  lab <- bquote(atop(.paste0(sex))*";",
    ~Dependent~variable~"="
    ~Relative~dot('V')*0['2peak']~ "[mL" * 0["2"]*"/min/kg*x"]",
    ~Quantile~"="~.(qtile)*";",
    ~Pseudo~"~*R^2~"~. (R1)))}
if(out == "VO2_MAX") {
  lab <- bquote(atop(.paste0(sex))*";",
    ~Dependent~variable~"="
    ~Absolute~dot('V')*0['2peak']~ "[L" * 0["2"]*"/min*x"]",
    ~Quantile~"="~.(qtile)*";",
    ~Pseudo~"~*R^2~"~. (R1)))}

# add variable names
name <- rownames(tab)
tab2 <- tab
tab2$name <- name

# coefficient plot
plt <- ggplot(data = tab2[2:nrow(tab2),]) +
  geom_point(aes(x = name, y = Value)) +
  theme(panel.background = element_rect(fill = "white"),
    panel.grid.major = element_line(linetype = "dotted", colour = "grey60", size = 0.3),
    plot.title = element_text(size = 10)) +
  geom_hline(yintercept = 0, lty = 2) +
  geom_linerange(aes(x = name, ymin = `2.5%`, ymax = `97.5%`)) +
  labs(x = "", y = "") +
  ggtitle(lab) +
  scale_y_continuous(expand = c(0.1,0.1), breaks = ) +
  coord_flip()

l <- list(tab, plt, R1, coef_in_ci, sig)

return(l)}

# males
rm25 <- QR(dfm, "REL_VO2_MAX", 0.25)
rm50 <- QR(dfm, "REL_VO2_MAX", 0.50)
rm75 <- QR(dfm, "REL_VO2_MAX", 0.75)

# females
rf25 <- QR(dff, "REL_VO2_MAX", 0.25)
rf50 <- QR(dff, "REL_VO2_MAX", 0.50)
rf75 <- QR(dff, "REL_VO2_MAX", 0.75)

# arranging coefficient plots
rm <- grid.arrange(rm25[[2]], rm50[[2]], rm75[[2]], ncol = 1)
rf <- grid.arrange(rf25[[2]], rf50[[2]], rf75[[2]], ncol = 1)

# absolute VO2peak

# males
am25 <- QR(dfm, "VO2_MAX", 0.25)
am50 <- QR(dfm, "VO2_MAX", 0.50)
am75 <- QR(dfm, "VO2_MAX", 0.75)

# females
af25 <- QR(dff, "VO2_MAX", 0.25)
af50 <- QR(dff, "VO2_MAX", 0.50)
af75 <- QR(dff, "VO2_MAX", 0.75)

# arranging coefficient plots
am <- grid.arrange(am25[[2]], am50[[2]], am75[[2]], ncol = 1)
af <- grid.arrange(af25[[2]], af50[[2]], af75[[2]], ncol = 1)

```

```

# saving plots
SAVE.PLOT(rm, "multi_relative_m.png", 8,6)
SAVE.PLOT(rf, "multi_relative_f.png", 8,6)
SAVE.PLOT(am, "multi_absolute_m.png", 8,6)
SAVE.PLOT(af, "multi_absolute_f.png", 8,6)

# tables
print(xtable(rbind(rm25[[1]], rm50[[1]], rm75[[1]]), digits=c(2,2,2,2,3)))
print(xtable(rbind(rf25[[1]], rf50[[1]], rf75[[1]]), digits=c(2,2,2,2,3)))
print(xtable(rbind(am25[[1]], am50[[1]], am75[[1]]), digits=c(2,2,2,2,3)))
print(xtable(rbind(af25[[1]], af50[[1]], af75[[1]]), digits=c(2,2,2,2,3)))

```

### 7.3.9 Comparison of reference values

```

# SHIP Study (Koch, 2009)

# CODING
# m=1, f=2
# age groups: 25-34, 35-44, 45-54, 55-64, >=64
# bmi: <25 = 0, >=25 = 1

# V02peak 5%
v.05 <- function(sex, Age, bmi){
  p <- 30.9643 +
    (-2.5661*Age) +
    (-0.0263*I(Age^2)) +
    (-3.7224*sex) +
    (1.8765*bmi) +
    (0.1082*Age*bmi) +
    (-2.9703*sex*bmi) +
    (0.7361*Age*sex) +
    (0.2799*Age*sex*bmi)
  return(p)
}

v.95 <- function(sex, Age, bmi){
  p <- 61.3721 +
    (-1.9479*Age) +
    (-0.3053*I(Age^2)) +
    (-9.1229*sex) +
    (3.8892*bmi) +
    (-1.9492*Age*bmi) +
    (-6.7455*sex*bmi) +
    (0.0716*Age*sex) +
    (1.6900*Age*sex*bmi)
  return(p)
}

v.50 <- function(sex, Age, bmi){
  p <- 47.7565 +
    (-0.9880*Age) +
    (-0.2356*I(Age^2)) +
    (-8.8697*sex) +
    (2.3597*bmi) +
    (-2.0308*Age*bmi) +
    (-3.7405*sex*bmi) +
    (0.2512*Age*sex) +
    (1.3797*Age*sex*bmi)
  return(p)
}

# prediction data set
ship_m <- data.frame(Age = 1:5)
ship_f <- data.frame(Age = 1:5)
own_m <- data.frame(Age = 25:70)
own_f <- data.frame(Age = 25:70)

# males
ship_m$m.05 <- v.05(1, ship_m$Age, 0)
ship_m$m.50 <- v.50(1, ship_m$Age, 0)
ship_m$m.95 <- v.95(1, ship_m$Age, 0)

# females

```



```

ship_f$f.05 <- v.05(2, ship_f$Age, 0)
ship_f$f.50 <- v.50(2, ship_f$Age, 0)
ship_f$f.95 <- v.95(2, ship_f$Age, 0)

ship_m$Age <- ship_f$Age <- factor(
  ship_m$Age,
  levels = 1:5,
  labels = c("25-34", "35-44", "45-54", "55-64", ">=64"))

ship_m_long <- melt(ship_m, id.vars = "Age")
ship_f_long <- melt(ship_f, id.vars = "Age")

# FRIEND Study (Kaminsky, 2015)

# Age groups (6): 20-29, 30-39, 40-49, 50-59, 60-69, 70-79
kaminsky_m <- data.frame(
  Age = rep(1:6, 3),
  variable = c(rep("m.05",6), rep("m.50",6), rep("m.95",6)),
  value = c(
    # m.50
    41.9, 30.1, 27.1, 24.8, 22.4, 19.5))

kaminsky_f <- data.frame(
  Age = rep(1:6, 3),
  variable = c(rep("f.05",6), rep("f.50",6), rep("f.95",6)),
  value = c(
    # f.50
    31.0, 21.6, 19.4, 17.3, 16.0, 14.8))

# recoding age
kaminsky_m$Age <- kaminsky_f$Age <- factor(
  kaminsky_m$Age,
  levels = 1:6,
  labels = c("20-29", "30-30", "40-49", "50-59", "60-69", "70-79"))

# own

dfm_FR$Age <- dfm_FR$ALTER
dff_FR$Age <- dff_FR$ALTER

# males
fitm.05 <- rq(REL_V02_MAX ~ Age + I(Age^2), data = dfm_FR, tau = 0.05)
fitm.50 <- rq(REL_V02_MAX ~ Age + I(Age^2), data = dfm_FR, tau = 0.50)
fitm.95 <- rq(REL_V02_MAX ~ Age + I(Age^2), data = dfm_FR, tau = 0.95)

own_m$m.05 <- predict(fitm.05, newdata = own_m)
own_m$m.50 <- predict(fitm.50, newdata = own_m)
own_m$m.95 <- predict(fitm.95, newdata = own_m)

own_m_long <- melt(own_m, id.vars = "Age")

# females
fitf.05 <- rq(REL_V02_MAX ~ Age + I(Age^2), data = dff_FR, tau = 0.05)
fitf.50 <- rq(REL_V02_MAX ~ Age + I(Age^2), data = dff_FR, tau = 0.50)
fitf.95 <- rq(REL_V02_MAX ~ Age + I(Age^2), data = dff_FR, tau = 0.95)

own_f$f.05 <- predict(fitf.05, newdata = own_f)
own_f$f.50 <- predict(fitf.50, newdata = own_f)
own_f$f.95 <- predict(fitf.95, newdata = own_f)

own_f_long <- melt(own_f, id.vars = "Age")

g <- function(dat, tit, xtit, ytit, class = T){
  # plot
  gp <- ggplot(data = dat, aes(x = Age, y = value, group = factor(variable))) +
    geom_line() +
    theme(panel.grid.major = element_line(colour = "grey60", size = 0.3),
          panel.background = element_rect(fill = "white")) +
    scale_y_continuous(limits = c(10,55), breaks = seq(from=10, to = 55, by = 5)) +
    labs(title = tit, x = xtit, y = ytit)

  # modifying x axis
  if(class == F){
    gp <- gp + scale_x_continuous(limits = c(25,70), breaks = seq(from= 25, to = 70, by = 5))
  }else{
    gp <- gp + geom_point()
  }
}

```

```
return(gp)
}

# plotting males
g1 <- g(own_m_long, "Prevention First", "Age [years]", REL_LAB, class = F)
g2 <- g(ship_m_long, "SHIP Study", "Age class [years]", "")
g3 <- g(kaminsky_m, "FRIEND Study", "Age class [years]", "")

# agganging plots
gm <- grid.arrange(g1, g2, g3, ncol = 3, top = "Men")

# plotting females
g1 <- g(own_f_long, "Prevention First", "Age [years]", REL_LAB, class = F)
g2 <- g(ship_f_long, "SHIP Study", "Age class [years]", "")
g3 <- g(kaminsky_f, "FRIEND Study", "Age class [years]", "")

# agganging plots
gf <- grid.arrange(g1, g2, g3, ncol = 3, top = "Women")

g_final <- grid.arrange(gf, gm, ncol = 1)

# saving plots
SAVE.PLOT(g_final, "Comp.png", 9, 9)

#####
#####
# web application
#####
#####

save.image()

# see separate code
```

## 8 Bibliography

- [1] American Thoracic Society (ATS), American College of Chest Physicians (ACCP) (2003) ATS/ACCP statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med* 167:211–277. ISSN 1073-449X, 1535-4970. doi:10.1164/rccm.167.2.211. URL <http://www.atsjournals.org/doi/abs/10.1164/rccm.167.2.211>
- [2] Aragon TJ (2012) epitools: Epidemiology Tools. URL <https://CRAN.R-project.org/package=epitools>. R package version 0.5-7
- [3] Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSM) und der Deutschen Gesellschaft für Epidemiologie (DGEpi) (2014) Gute Praxis Sekundärdatenanalyse (GPS) Leitlinien und Empfehlungen. URL [http://dgepi.de/fileadmin/pdf/leitlinien/GPS\\_revision2-final\\_august2014.pdf](http://dgepi.de/fileadmin/pdf/leitlinien/GPS_revision2-final_august2014.pdf)
- [4] Assmann G, Cullen P, Schulte H (2002) Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Münster (procam) study. *Circulation* 105:310–5. doi:10.1161/hc0302.102575
- [5] Aune D, Norat T, Leitzmann M, Tonstad S, Vatten LJ (2015) Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. *Eur J Epidemiol* 30:529–542. ISSN 0393-2990, 1573-7284. doi:10.1007/s10654-015-0056-z. URL <http://link.springer.com/10.1007/s10654-015-0056-z>
- [6] Bader DS, Maguire TE, Balady GJ (1999) Comparison of ramp versus step protocols for exercise testing in patients  $\geq 60$  years of age. *J Am Coll Cardiol* 83:11–14. ISSN 00029149. doi:10.1016/S0002-9149(98)00774-7. URL <http://linkinghub.elsevier.com/retrieve/pii/S0002914998007747>
- [7] Balady GJ, Arena R, Sietsema K, Myers J, Coke L, Fletcher GF, Forman D, Franklin B, Guazzi M, Gulati M, Keteyian SJ, Lavie CJ, Macko R, Mancini D, Milani RV, on behalf of the American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee of the Council on Clinical Cardiology, Council on Epidemiology and Prevention, Council on Peripheral Vascular Disease, Interdisciplinary Council on Quality of Care and Outcomes Research (2010) Clinician’s guide to cardiopulmonary exercise testing in adults: A scientific statement from the American Heart Association. *Circulation* 122:191–225. ISSN 0009-7322, 1524-4539. doi:10.1161/CIR.0b013e3181e52e69. URL <http://circ.ahajournals.org/cgi/doi/10.1161/CIR.0b013e3181e52e69>

- 
- [8] Borg A, Sariyar M (2016) RecordLinkage: Record Linkage in R. URL <https://CRAN.R-project.org/package=RecordLinkage>. R package version 0.4-10
- [9] Catapano AL, Graham I, De Backer G, Wiklund O, Chapman MJ, Drexel H, Hoes AW, Jennings CS, Landmesser U, Pedersen TR, Reiner Å, Riccardi G, Taskinen MR, Tokgozoglul L, Verschuren WMM, Vlachopoulos C, Wood DA, Zamorano JL (2016) 2016 ESC/EAS guidelines for the management of dyslipidaemias. *Eur Heart J* 37:2999–3058. ISSN 0195-668X, 1522-9645. doi:10.1093/eurheartj/ehw272. URL <https://academic.oup.com/eurheartj/article-lookup/doi/10.1093/eurheartj/ehw272>
- [10] Chang W, Cheng J, Allaire J, Xie Y, McPherson J (2017) shiny: Web Application Framework for R. URL <https://CRAN.R-project.org/package=shiny>. R package version 1.0.3
- [11] Chiriboga DE, Ockene IS (2016) Fitness in young adults as an independent predictor of risk for cardiovascular disease. *JAMA Intern Med* 176:95. ISSN 2168-6106. doi:10.1001/jamainternmed.2015.6819. URL <http://archinte.jamanetwork.com/article.aspx?doi=10.1001/jamainternmed.2015.6819>
- [12] Conroy R, Pyörälä K, Fitzgerald A, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetière P, Jousilahti P, Keil U, Njølstad I, Oganov R, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham I, SCORE project group (2003) Estimation of ten-year risk of fatal cardiovascular disease in europe: the score project. *Eur Heart J* 24:987–1003. doi:10.1016/S0195-668X(03)00114-3
- [13] D’Agostino RS, Vasan R, Pencina M, Wolf P, Cobain M, JM M, Kannel W (2008) General cardiovascular risk profile for use in primary care: the framingham heart study. *Circulation* 117:743–53. doi:10.1161/CIRCULATIONAHA.107.699579
- [14] Davino C, Furno M, Vistocco D (2014) Quantile regression: theory and applications. Wiley series in probability and statistics, Wiley, Chichester, West Sussex. ISBN 978-1-119-97528-1
- [15] Deutsche Gesellschaft für Sportmedizin und Prävention (2007) S1-Leitlinie Vorsorgeuntersuchung im Sport. URL [http://daten2.verwaltungsportal.de/dateien/seitengenerator/leitlinie\\_vorsorgeuntersuchung\\_4.10.2007-1-19.pdf](http://daten2.verwaltungsportal.de/dateien/seitengenerator/leitlinie_vorsorgeuntersuchung_4.10.2007-1-19.pdf)
- [16] Dickhuth HH, Gollhofer A (2007) Allgemeine medizinische Trainingslehre. In: HH Dickhuth, G Badtke (eds.) Sportmedizin für Ärzte: Lehrbuch auf der Grundlage des Weiterbildungssystems der Deutschen Gesellschaft für Sportmedizin und Prävention (DGSP) ; mit 109 Tabellen. Dt Ärzte-Verl, Köln. ISBN 978-3-7691-0472-1. OCLC: 180173071

- 
- [17] Edvardsen E, Scient C, Hansen BH, Holme IM, Dyrstad SM, Anderssen SA (2013) Reference values for cardiorespiratory response and fitness on the treadmill in a 20- to 85-year-old population. *Chest* 144:241–248. ISSN 1931-3543. doi:10.1378/chest.12-1458
- [18] Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, Nieman DC, Swain DP (2011) Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Med Sci Sports Exerc* 43:1334–1359. ISSN 0195-9131. doi:10.1249/MSS.0b013e318213febf. URL <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00005768-201107000-00026>
- [19] Gosswald A, Lange M, Kamtsiuris P, Kurth BM (2012) DEGS: Studie zur Gesundheit Erwachsener in Deutschland. Bundesweite Quer- und Längsschnittstudie im Rahmen des Gesundheitsmonitorings des Robert Koch-Instituts. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 55:775–780. ISSN 1436-9990. doi:10.1007/s00103-012-1498-z
- [20] Gulati M, Pandey DK, Arnsdorf MF, Lauderdale DS, Thisted RA, Wicklund RH, Al-Hani AJ, Black HR (2003) Exercise capacity and the risk of death in women: the st james women take heart project. *Circulation* 108:1554–1559. ISSN 1524-4539. doi:10.1161/01.CIR.0000091080.57509.E9
- [21] Hamer M, Chida Y (2009) Physical activity and risk of neurodegenerative disease: a systematic review of prospective evidence. *Psychol Med* 39:3. ISSN 0033-2917, 1469-8978. doi:10.1017/S0033291708003681. URL [http://www.journals.cambridge.org/abstract\\_S0033291708003681](http://www.journals.cambridge.org/abstract_S0033291708003681)
- [22] Hao L, Naiman DQ (2007) Quantile regression. Number 149 in *Quantitative applications in the social sciences*, Sage Publications, Thousand Oaks, Calif. ISBN 978-1-4129-2628-7
- [23] Hawkins S, Wiswell R (2003) Rate and mechanism of maximal oxygen consumption decline with aging: implications for exercise training. *Sports Med* 33:877–888. ISSN 0112-1642
- [24] Huang G, Wang R, Chen P, Huang SC, Donnelly JE, Mehlferber JP (2016) Dose-response relationship of cardiorespiratory fitness adaptation to controlled endurance training in sedentary older adults. *Eur J Prev Cardiol* 23:518–529. ISSN 2047-4873, 2047-4881. doi:10.1177/2047487315582322. URL <http://journals.sagepub.com/doi/10.1177/2047487315582322>
- [25] Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF, Tibshirani R (2014) Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet* 383:166–175. ISSN 01406736. doi:10.

- 1016/S0140-6736(13)62227-8. URL <http://linkinghub.elsevier.com/retrieve/pii/S0140673613622278>
- [26] Itoh H, Ajisaka R, Koike A, Makita S, Omiya K, Kato Y, Adachi H, Nagayama M, Maeda T, Tajima A, Harada N, Taniguchi K (2013) Heart rate and blood pressure response to ramp exercise and exercise capacity in relation to age, gender, and mode of exercise in a healthy population. *J Cardiol* 61:71–78. ISSN 09145087. doi:10.1016/j.jjcc.2012.09.010. URL <http://linkinghub.elsevier.com/retrieve/pii/S0914508712002687>
- [27] Jackson A, Pollock M (1985) Practical assessment of body composition. *Phys Sportsmed* 13:76–90. doi:10.1080/00913847
- [28] JBS3 Board (2014) Joint british societies’ consensus recommendations for the prevention of cardiovascular disease (jbs3). *Heart* 100:ii1–ii67. doi:10.1136/heartjnl-2014-305693
- [29] Jurczyk P, Lu JJ, Xiong L, Cragan JD, Correa A (2008) FRIL: A tool for comparative record linkage. *Annu Symp Proc* pp. 440–444. ISSN 1942-597X
- [30] Kaminsky LA, Arena R, Myers J (2015) Reference standards for cardiorespiratory fitness measured with cardiopulmonary exercise testing. *Mayo Clin Proc* 90:1515–1523. ISSN 00256196. doi:10.1016/j.mayocp.2015.07.026. URL <http://linkinghub.elsevier.com/retrieve/pii/S0025619615006424>
- [31] Kaminsky LA, Imboden MT, Arena R, Myers J (2017) Reference standards for cardiorespiratory fitness measured with cardiopulmonary exercise testing using cycle ergometry: Data from the fitness registry and the importance of exercise national database (FRIEND) registry. *Mayo Clin Proc* 92:228–233. ISSN 00256196. doi:10.1016/j.mayocp.2016.10.003. URL <http://linkinghub.elsevier.com/retrieve/pii/S0025619616306243>
- [32] Kamtsiuris P, Lange M, Hoffmann R, Schaffrath Rosario A, Dahm S, Kuhnert R, Kurth B (2013) Die erste Welle der Studie zur Gesundheit Erwachsener in Deutschland (DEGS1): Stichprobendesign, Response, Gewichtung und Repräsentativität. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 56:620–630. ISSN 1436-9990, 1437-1588. doi:10.1007/s00103-012-1650-9. URL <http://link.springer.com/10.1007/s00103-012-1650-9>
- [33] Kim ES, Ishwaran H, Blackstone E, Lauer MS (2007) External prognostic validations and comparisons of age- and gender-adjusted exercise capacity predictions. *J Am Coll Cardiol* 50:1867–1875. ISSN 07351097. doi:10.1016/j.jacc.2007.08.003. URL <http://linkinghub.elsevier.com/retrieve/pii/S0735109707025697>

- 
- [34] Koch B, Schäper C, Ittermann T, Spielhagen T, Dörr M, Völzke H, Opitz CF, Ewert R, Gläser S (2009) Reference values for cardiopulmonary exercise testing in healthy volunteers: the SHIP study. *Eur Respir J* 33:389–397. ISSN 1399-3003. doi:10.1183/09031936.00074208
- [35] Kodama S (2009) Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: A meta-analysis. *JAMA* 301:2024. ISSN 0098-7484. doi:10.1001/jama.2009.681. URL <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2009.681>
- [36] Koenker R (2016) quantreg: Quantile Regression. URL <https://CRAN.R-project.org/package=quantreg>. R package version 5.29
- [37] Kokkinos P (2012) Physical activity, health benefits, and mortality risk. *ISRN Cardiol* 2012:1–14. ISSN 2090-5599. doi:10.5402/2012/718789. URL <http://www.hindawi.com/journals/isrn/2012/718789/>
- [38] Lampert T, von der Lippe E, Müters S (2013) Verbreitung des Rauchens in der Erwachsenenbevölkerung in Deutschland: Ergebnisse der Studie zur Gesundheit Erwachsener in Deutschland (DEGS1). *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 56:802–808. ISSN 1436-9990, 1437-1588. doi:10.1007/s00103-013-1698-1. URL <http://link.springer.com/10.1007/s00103-013-1698-1>
- [39] Lin X, Zhang X, Guo J, Roberts CK, McKenzie S, Wu WC, Liu S, Song Y (2015) Effects of exercise training on cardiorespiratory fitness and biomarkers of cardiometabolic health: A systematic review and meta-analysis of randomized controlled trials. *J Am Heart Assoc* 4. ISSN 2047-9980. doi:10.1161/JAHA.115.002014
- [40] Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, Greenlund K, Daniels S, Nichol G, Tomaselli GF, Arnett DK, Fonarow GC, Ho PM, Lauer MS, Masoudi FA, Robertson RM, Roger V, Schwamm LH, Sorlie P, Yancy CW, Rosamond WD, on behalf of the American Heart Association Strategic Planning Task Force and Statistics Committee (2010) Defining and setting national goals for cardiovascular health promotion and disease reduction: The American Heart Association's strategic impact goal through 2020 and beyond. *Circulation* 121:586–613. ISSN 0009-7322, 1524-4539. doi:10.1161/CIRCULATIONAHA.109.192703. URL <http://circ.ahajournals.org/cgi/doi/10.1161/CIRCULATIONAHA.109.192703>
- [41] Löllgen H, Böckenhoff A, Knapp G (2009) Physical activity and all-cause mortality: An updated meta-analysis with different intensity categories. *Int J Sports Med* 30:213–224. ISSN 0172-4622, 1439-3964. doi:10.1055/s-0028-1128150. URL <http://www.thieme-connect.de/DOI/DOI?10.1055/s-0028-1128150>

- [42] Marsh L, Cormier DR (2001) Spline regression models. Number 07-137 in Quantitative applications in the social sciences, Sage Publications, Thousand Oaks, Calif. ISBN 978-0-7619-2420-3
- [43] Mensink G, Schienkiewitz A, Haftenberger M, Lampert T, Ziese T, Scheidt-Nave C (2013) Übergewicht und adipositas in deutschland: Ergebnisse der studie zur gesundheit erwachsener in deutschland (DEGS1). Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 56:786–794. ISSN 1436-9990, 1437-1588. doi:10.1007/s00103-012-1656-3. URL <http://link.springer.com/10.1007/s00103-012-1656-3>
- [44] Meyer T, Kindermann W (1999) Standards in der sportmedizin: Die maximale sauerstoffaufnahme ( $\text{VO}_2\text{max}$ ). DZSM 50:285–286
- [45] Midgley AW, McNaughton LR, Polman R, Marchant D (2007) Criteria for determination of maximal oxygen uptake: a brief critique and recommendations for future research. Sports Med 37:1019–1028. ISSN 0112-1642
- [46] Morris CK, Myers J, Froelicher VF, Kawaguchi T, Ueshima K, Hideg A (1993) Nomogram based on metabolic equivalents and age for assessing aerobic exercise capacity in men. J Am Coll Cardiol 22:175–182. ISSN 0735-1097
- [47] Myers J, Buchanan N, Walsh D, Kraemer M, McAuley P, Hamilton-Wessler M, Froelicher VF (1991) Comparison of the ramp versus standard exercise protocols. J Am Coll Cardiol 17:1334–1342. ISSN 0735-1097
- [48] Neuhauser H, Thamm M, Ellert U (2013) Blutdruck in deutschland 2008-2011: Ergebnisse der studie zur gesundheit erwachsener in deutschland (DEGS1). Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 56:795–801. ISSN 1436-9990, 1437-1588. doi:10.1007/s00103-013-1669-6. URL <http://link.springer.com/10.1007/s00103-013-1669-6>
- [49] Online Appendix (2018) Quantile reference values for peak oxygen uptake: Cross-sectional study of 9,354 adult participants of cardiopulmonary exercise tests using cycle ergometry in three German cities. URL <https://github.com/rappdaniel/vo2peak>
- [50] Paap D, Takken T (2014) Reference values for cardiopulmonary exercise testing in healthy adults: a systematic review. Expert Rev Cardiovasc Ther 12:1439–1453. ISSN 1477-9072, 1744-8344. doi:10.1586/14779072.2014.985657. URL <http://www.tandfonline.com/doi/full/10.1586/14779072.2014.985657>
- [51] Pappasavvas T, Bonow RO, Alhashemi M, Micklewright D (2016) Depression symptom severity and cardiorespiratory fitness in healthy and depressed adults: A systematic review and meta-analysis. Sports Med 46:219–230. ISSN 0112-1642, 1179-2035. doi:10.1007/s40279-015-0409-5. URL <http://link.springer.com/10.1007/s40279-015-0409-5>



- [52] Peng RD (2009) Reproducible research and biostatistics. *Biostatistics* 10:405–408. ISSN 1468-4357, 1465-4644. doi:10.1093/biostatistics/kxp014. URL <https://academic.oup.com/biostatistics/article-lookup/doi/10.1093/biostatistics/kxp014>
- [53] Pescatello LS, American College of Sports Medicine (eds.) (2014) *ACSM’s guidelines for exercise testing and prescription*. 9. edition. Wolters Kluwer/Lippincott Williams & Wilkins Health, Philadelphia. ISBN 978-1-60913-605-5
- [54] R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- [55] Rapp D, Scharhag J, Wagenpfeil S, Scholl J (2018) Reference values for peak oxygen uptake: cross-sectional analysis of cycle ergometry-based cardiopulmonary exercise tests of 10 090 adult german volunteers from the prevention first registry. *BMJ Open* 8:e018697. ISSN 2044-6055, 2044-6055. doi:10.1136/bmjopen-2017-018697. URL <http://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2017-018697>
- [56] Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, Collins GS (2016) External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 353:i3140. ISSN 1756-1833. doi:10.1136/bmj.i3140
- [57] Roglic G, World Health Organization (eds.) (2016) *Global report on diabetes*. World Health Organization. ISBN 978-92-4-156525-7. URL [http://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257\\_eng.pdf;jsessionid=82D5FECBA0F8595C383597B1CBB27069?sequence=1](http://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257_eng.pdf;jsessionid=82D5FECBA0F8595C383597B1CBB27069?sequence=1). OCLC: ocn948336981
- [58] Ross R, Blair SN, Arena R, Church TS, Després JP, Franklin BA, Haskell WL, Kaminsky LA, Levine BD, Lavie CJ, Myers J, Niebauer J, Sallis R, Sawada SS, Sui X, Wisløff U (2016) Importance of assessing cardiorespiratory fitness in clinical practice: A case for fitness as a clinical vital sign: A scientific statement from the american heart association. *Circulation* 134:e653–e699. ISSN 0009-7322, 1524-4539. doi:10.1161/CIR.0000000000000461. URL <http://circ.ahajournals.org/lookup/doi/10.1161/CIR.0000000000000461>
- [59] Scharhag-Rosenberger F, Schommer K (2013) Die Spiroergometrie in der Sportmedizin. *DZSM* 64:362–366. ISSN 03445925. doi:10.5960/dzsm.2013.105. URL <http://www.zeitschrift-sportmedizin.de/artikel-online/archiv-2013/heft-12/die-spiroergometrie-in-der-sportmedizin/>
- [60] Schmid D, Leitzmann MF (2015) Cardiorespiratory fitness as predictor of cancer mortality: a systematic review and meta-analysis. *Ann Oncol* 26:272–278. ISSN

- 0923-7534, 1569-8041. doi:10.1093/annonc/mdu250. URL <https://academic.oup.com/annonc/article-lookup/doi/10.1093/annonc/mdu250>
- [61] Shah RV, Murthy VL, Colangelo LA, Reis J, Venkatesh BA, Sharma R, Abbasi SA, Goff DC, Carr JJ, Rana JS, Terry JG, Bouchard C, Sarzynski MA, Eisman A, Neilan T, Das S, Jerosch-Herold M, Lewis CE, Carnethon M, Lewis GD, Lima JAC (2016) Association of fitness in young adulthood with survival and cardiovascular risk: The coronary artery risk development in young adults (CARDIA) study. *JAMA Intern Med* 176:87. ISSN 2168-6106. doi:10.1001/jamainternmed.2015.6309. URL <http://archinte.jamanetwork.com/article.aspx?doi=10.1001/jamainternmed.2015.6309>
- [62] Statistisches Bundesamt (2015) Bevölkerung und Erwerbstätigkeit: Ausgangsdaten der Bevölkerungsfortschreibung aus dem Zensus 2011. Statistisches Bundesamt, Wiesbaden, Germany. URL <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Bevoelkerungsstand/Bevoelkerungsstand.htm>
- [63] Statistisches Bundesamt (2016) Bevölkerung und Erwerbstätigkeit: Bevölkerungsfortschreibung auf Grundlage des Zensus 2011. URL [https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Bevoelkerungsstand/Bevoelkerungsfortschreibung2010130147004.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Bevoelkerungsstand/Bevoelkerungsfortschreibung2010130147004.pdf?__blob=publicationFile)
- [64] Steyerberg EW (2009) Clinical prediction models: a practical approach to development, validation, and updating. *Statistics for biology and health*, Springer, New York. ISBN 978-0-387-77243-1 978-0-387-77244-8. OCLC: ocn295033353
- [65] Subirana I, Sanz H, Vila J (2014) Building bivariate tables: The compareGroups package for R. *J Stat Softw* 57:1–16. URL <http://www.jstatsoft.org/v57/i12/>
- [66] Tanaka H, Monahan KD, Seals DR (2001) Age-predicted maximal heart rate revisited. *J Am Coll Cardiol* 37:153–156. ISSN 07351097. doi:10.1016/S0735-1097(00)01054-8. URL <http://linkinghub.elsevier.com/retrieve/pii/S0735109700010548>
- [67] Trappe HJ, Löllgen H (2000) Leitlinien zur ergometrie. *Z Kardiol* 89:821–831. ISSN 0300-5860
- [68] Tutz G, Gertheiss J (2014) Rating scales as predictors - the old question of scale level and some answers. *Psychometrika* 79:357–376. ISSN 0033-3123, 1860-0980. doi:10.1007/s11336-013-9343-3. URL <http://link.springer.com/10.1007/s11336-013-9343-3>
- [69] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative (2007) The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet* 370:1453–1457. ISSN 1474-547X. doi:10.1016/S0140-6736(07)61602-X

- 
- [70] Wasserman K, Hansen JE, Sue DY, Stringer WW, Sietsema KE, Sun X, Whipp BJ (eds.) (2012) Principles of exercise testing and interpretation: including pathophysiology and clinical applications. 5th ed edition. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia. ISBN 978-1-60913-899-8
- [71] Wasserman K, Hansen JE, Sue DY, Stringer WW, Whipp BJ (eds.) (2005) Principles of exercise testing and interpretation: including pathophysiology and clinical applications. 4th edition. Lippincott Williams & Wilkins, Philadelphia. ISBN 978-0-7817-4876-6
- [72] Weiß C (2013) Basiswissen Medizinische Statistik: mit 20 Tabellen ; [mit Epidemiologie]. 6., revised edition. Springer, Berlin. ISBN 978-3-642-34261-5
- [73] World Health Organization (2009) WHO child growth standards: growth velocity based on weight, length and head circumference: methods and development. World Health Organization, Department of Nutrition for Health and Development, Geneva, Switzerland. ISBN 978-92-4-154763-5. OCLC: ocn316014678
- [74] World Health Organization (2017) Cardiovascular diseases (CVDs). URL <http://www.who.int/mediacentre/factsheets/fs317/en/>
- [75] World Health Organization (2017) The WHO Child Growth Standards. URL <http://www.who.int/childgrowth/en/>
- [76] World Health Organization (2018) Obesity and overweight. URL <http://www.who.int/mediacentre/factsheets/fs311/en/>
- [77] Zaccardi F, O'Donovan G, Webb DR, Yates T, Kurl S, Khunti K, Davies MJ, Laukkanen JA (2015) Cardiorespiratory fitness and risk of type 2 diabetes mellitus: A 23-year cohort study and a meta-analysis of prospective studies. *Atherosclerosis* 243:131–137. ISSN 00219150. doi:10.1016/j.atherosclerosis.2015.09.016. URL <http://linkinghub.elsevier.com/retrieve/pii/S0021915015301258>

# A Appendix

## A.1 Random sample

**Table A.1:** Bivariate descriptive table of the full dataset and random sample.

	Full data N=9937	Random sample N=252	P value
<b><math>\dot{V}O_{2peak}</math></b>			
Relative [mLO <sub>2</sub> /min/kg]	32.8 [27.7;38.4]	31.9 [27.2;37.5]	0.211
Absolute [LO <sub>2</sub> /min]	2.63 [2.06;3.18]	2.56 [1.97;3.13]	0.218
<b>Anthropometric</b>			
Age [years]	45.0 [41.0;50.0]	45.0 [40.0;50.0]	0.546
Weight [kg]	80.0 [69.0;90.0]	79.0 [67.5;89.0]	0.615
Height [cm]	177 [170;183]	177 [168;183]	0.459
BMI [kg/m <sup>2</sup> ]	25.1 [22.9;27.7]	24.7 [22.7;28.0]	0.713
Sex:			0.756
Men	5916 (64.8%)	147 (63.6%)	
Women	3207 (35.2%)	84 (36.4%)	
Smoker			0.374
no	7810 (86.1%)	192 (83.8%)	
yes	1258 (13.9%)	37 (16.2%)	
Overweight			0.234
no	4434 (48.6%)	122 (52.8%)	
yes	4684 (51.4%)	109 (47.2%)	
<b>Study center</b>			
City			0.628
Frankfurt/Ruedesheim	7327 (80.3%)	189 (81.8%)	
Munich	1796 (19.7%)	42 (18.2%)	

**Note:** Full data denotes the overall dataset (n=10,189) excluding participants who were selected for the random sample (n=252). The random sample was drawn from the original dataset including 10,189 participants before exclusion of participants (fig. 3.2).

None of the obtained P values was < 0.05. Consequently, there were no statistically significant differences between the full data and the random sample. The random sample was hence assumed to represent the full dataset adequately. Quantitative variables are displayed as median [1st quartile; 3rd quartile], qualitative characteristics as n (percent).

Overweight: BMI  $\geq$  25 kg/m<sup>2</sup>, obesity: BMI  $\geq$  30 kg/m<sup>2</sup>.

## A.2 Selective dropout analysis

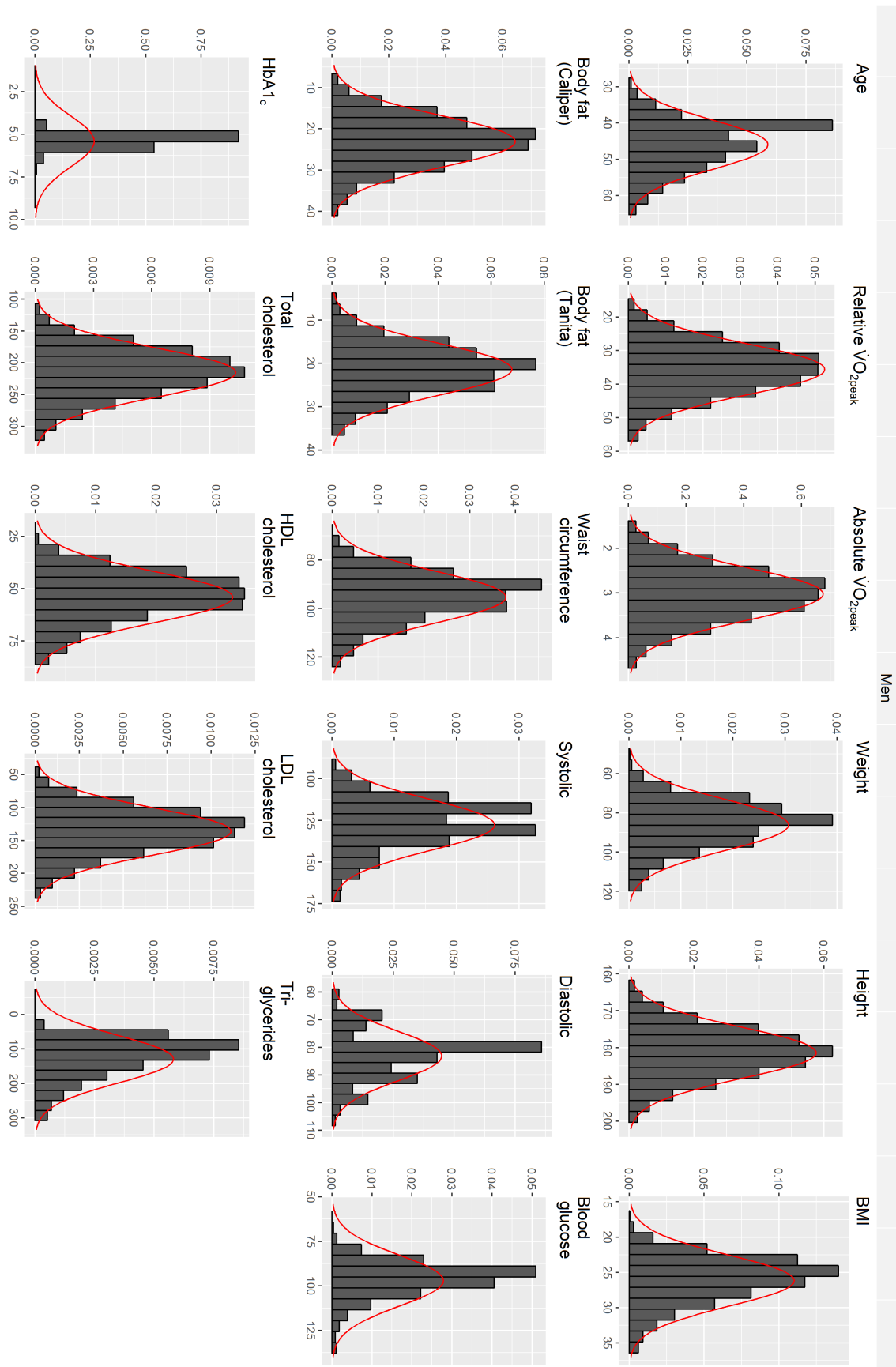
**Table A.2:** Bivariate descriptive table of the final dataset and excluded cases.

	Final dataset N=9354	Excluded cases N=835	P value
<b><math>\dot{V}O_{2\text{peak}}</math></b>			
Relative [mLO <sub>2</sub> /min/kg]	32.8 [27.7;38.4]	31.6 [26.4;36.8]	<0.001
Absolute [LO <sub>2</sub> /min]	2.63 [2.06;3.18]	2.29 [1.84;2.94]	<0.001
<b>Anthropometric</b>			
Age [years]	45.0 [41.0;50.0]	46.0 [41.0;51.0]	<0.001
Weight [kg]	80.0 [69.0;90.0]	75.0 [65.0;86.0]	<0.001
Height [cm]	177 [170;183]	173 [166;181]	<0.001
BMI [kg/m <sup>2</sup> ]	25.1 [22.9;27.7]	24.5 [22.2;27.2]	<0.001
Sex			<0.001
Men	6063 (64.8%)	449 (53.8%)	
Women	3291 (35.2%)	386 (46.2%)	
Smoker			0.722
no	8002 (86.1%)	710 (86.6%)	
yes	1295 (13.9%)	110 (13.4%)	
Overweight			<0.001
no	4556 (48.7%)	463 (55.7%)	
yes	4793 (51.3%)	368 (44.3%)	
<b>Study center</b>			
City			0.805
Frankfurt/Ruedesheim	7516 (80.4%)	126 (79.2%)	
Munich	1838 (19.6%)	33 (20.8%)	

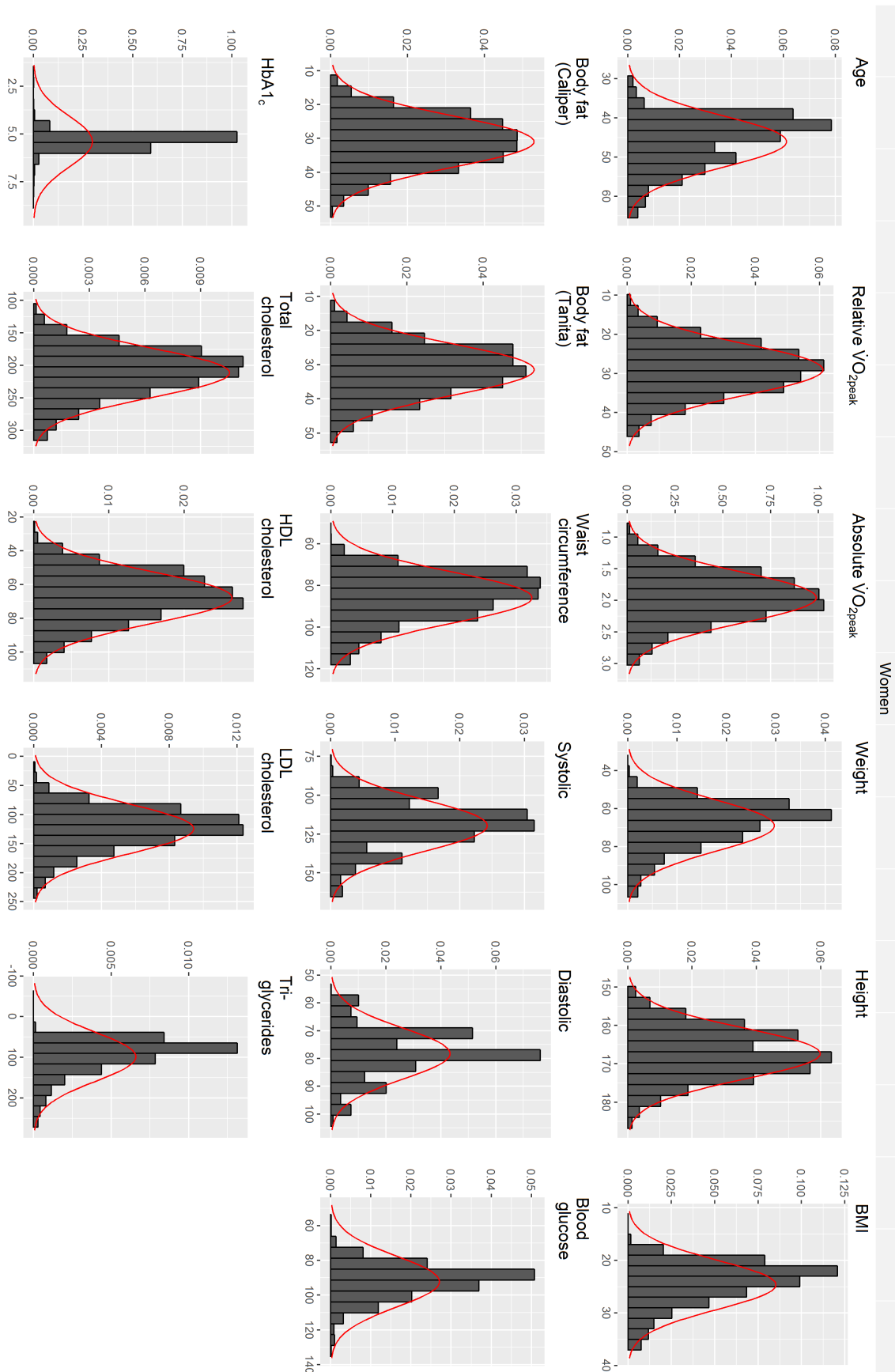
**Note:** Final dataset denotes the main study dataset (n=9,354) excluding participants who were excluded due to missing values (n=835) (fig. 3.2).

Quantitative variables are displayed as median [1st quartile; 3rd quartile], qualitative characteristics as n (percent).

Overweight: BMI  $\geq$  25 kg/m<sup>2</sup>, obesity: BMI  $\geq$  30 kg/m<sup>2</sup>.



**Figure A.1:** Histograms of continuous variables for men.  
**Note:** Men from Frankfurt/Rüdesheim and Munich were included ( $n=6,063$ ).



**Figure A.2:** Histograms of continuous variables for women.  
**Note:** Women from Frankfurt/Rüdesheim and Munich were included (n=3,291).

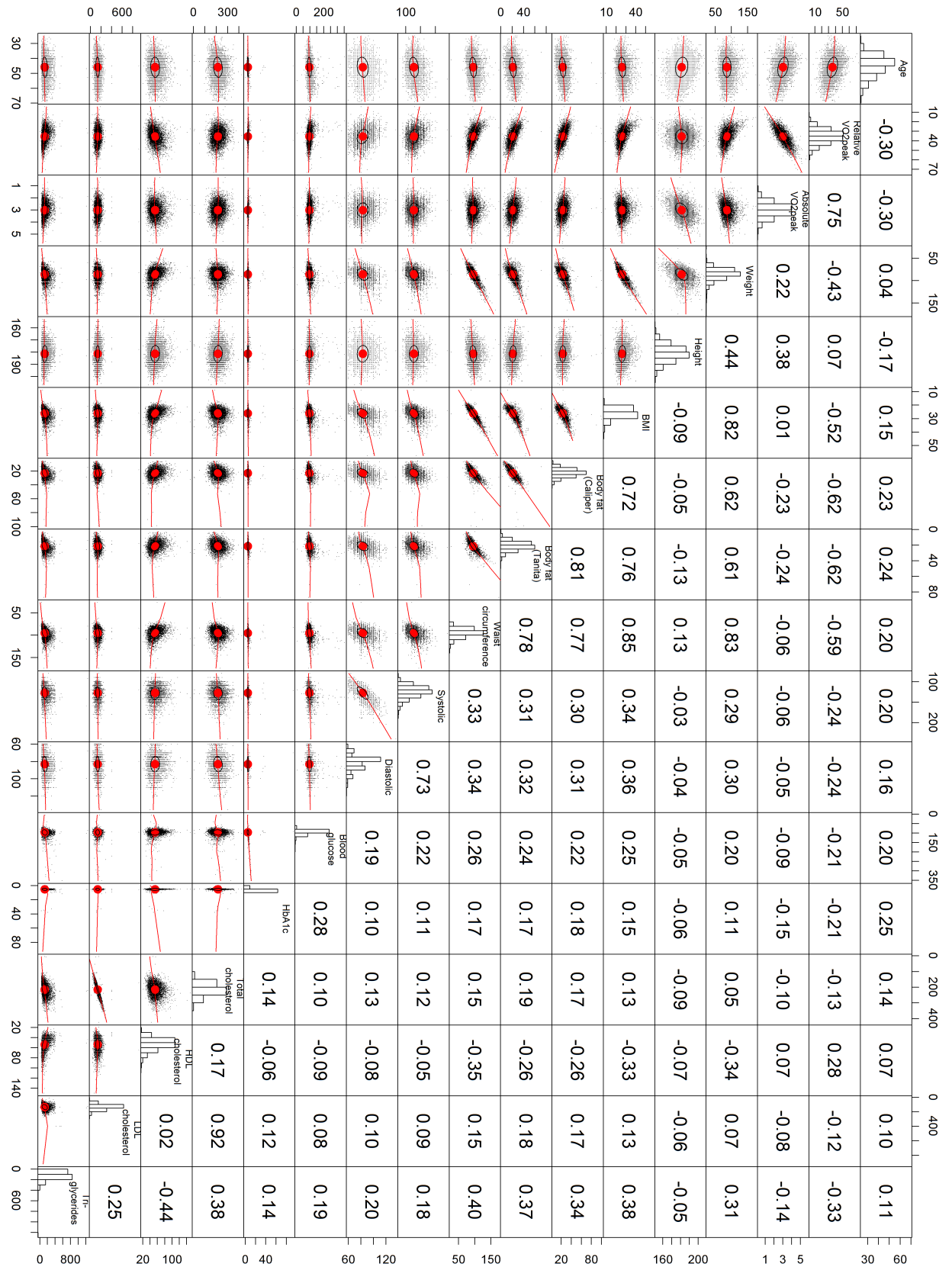


Figure A.3: Correlation matrix of continuous variables for men.

Note: Scatter plots including ellipsoids and Loess curves are plotted on the bottom-left, histograms on the diagonal and Spearman correlation coefficients on the top-right.

All 6,063 men from Frankfurt/Rüdesheim and Munich were included.



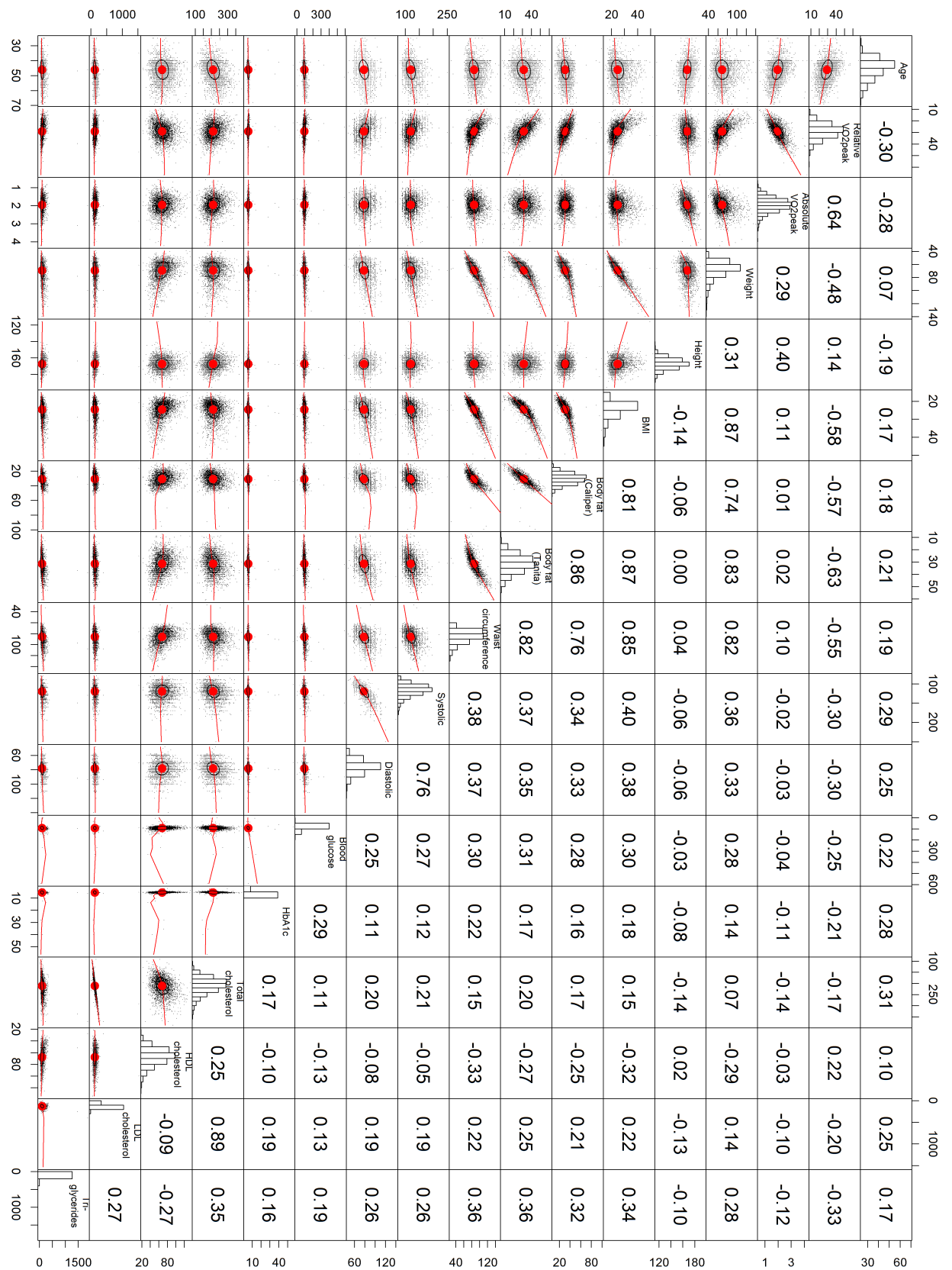


Figure A.4: Correlation matrix of continuous variables for women.

Note: Scatter plots including ellipsoids and Loess curves are plotted on the bottom-left, histograms on the diagonal and Spearman correlation coefficients on the top-right.

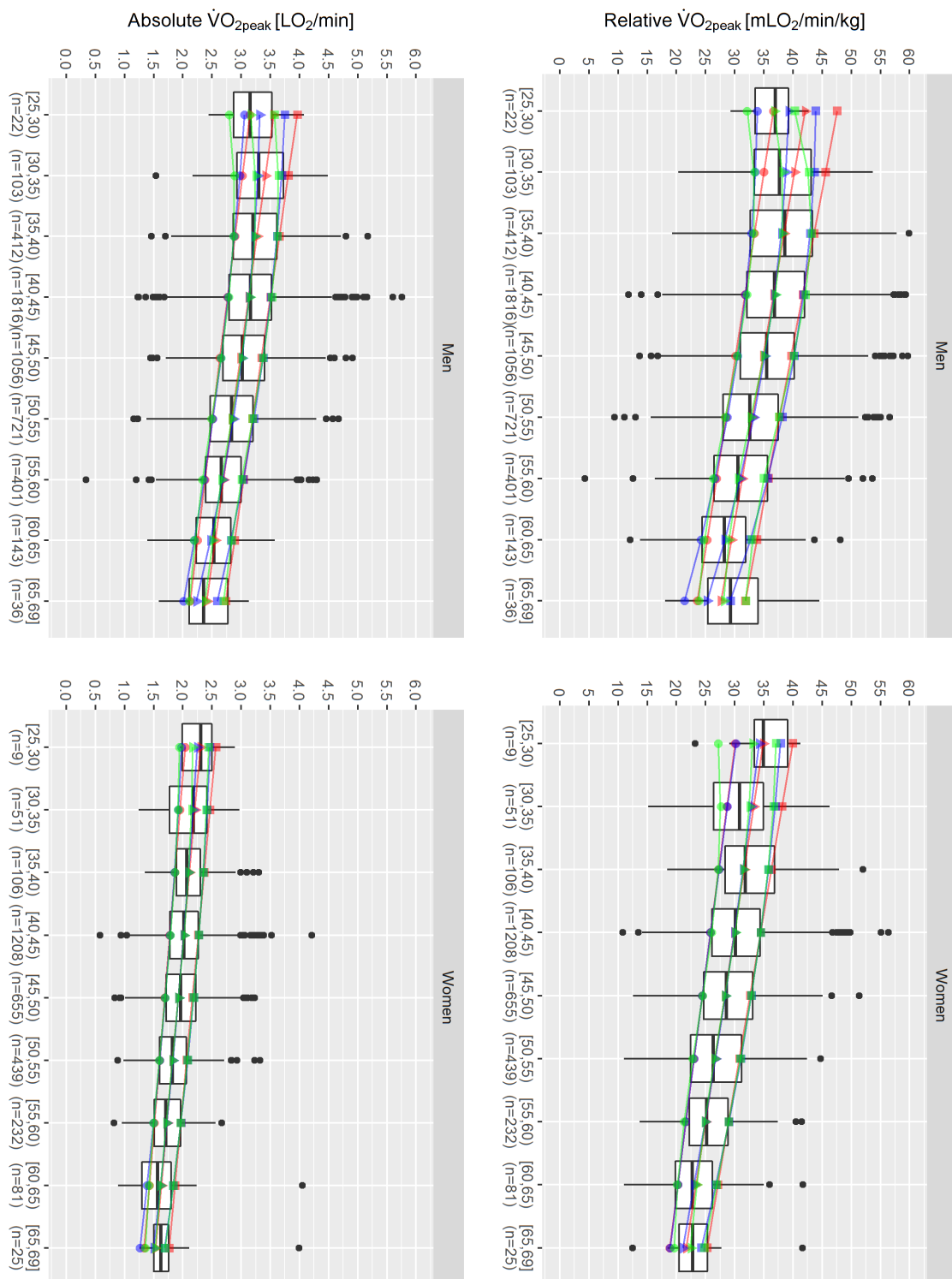
All 3,291 women from Frankfurt/Rüdesheim and Munich were included.

**Table A.3:** Skewness and quantile-based skewness of all eligible quantitative variables

	Men		Women	
	Skewness	QSK	Skewness	QSK
<b>Outcome measures</b>				
Relative $\dot{V}O_{2\text{peak}}$ [mLO <sub>2</sub> /min/kg]	0.19	0.06	0.30	0.09
Absolute $\dot{V}O_{2\text{peak}}$ [LO <sub>2</sub> /min]	0.09	0.08	0.43	0.04
<b>Anthropometrics</b>				
Age [years]	0.50	0.83	0.60	1.00
Weight [kg]	0.96	0.31	1.30	0.82
Height [cm]	-0.01	0.12	-0.10	0.03
BMI [kg/m <sup>2</sup> ]	1.27	0.53	1.42	0.93
Body fat (Caliper) [%]	1.15	0.04	0.64	0.00
Body fat (Tanita) [%]	0.55	0.13	0.16	0.11
Waist circumference [cm]	0.86	0.36	0.82	0.58
<b>Blood pressure</b>				
Systolic [mmHg]	0.82	0.38	1.04	0.00
Diastolic [mmHg]	0.57	0.30	0.52	-0.17
<b>Glucose metabolism</b>				
Blood glucose [mg/dL]	5.72	0.40	13.63	0.30
HbA1 <sub>c</sub> [%]	41.38	-0.00	25.73	0.33
<b>Lipid metabolism</b>				
Total cholesterol [mg/dL]	0.31	0.13	0.66	0.30
HDL Cholesterol [mg/dL]	0.98	0.50	0.56	0.38
LDL Cholesterol [mg/dL]	2.29	0.18	11.75	0.28
Triglycerides [mg/dL]	3.89	1.22	10.65	1.37

Quantile-based skewness (QSK) was calculated according to [22, p. 14]. All 3,291 women and all 6,063 men from Frankfurt/Rüdesheim and Munich were included.

### A.3 Regression modelling



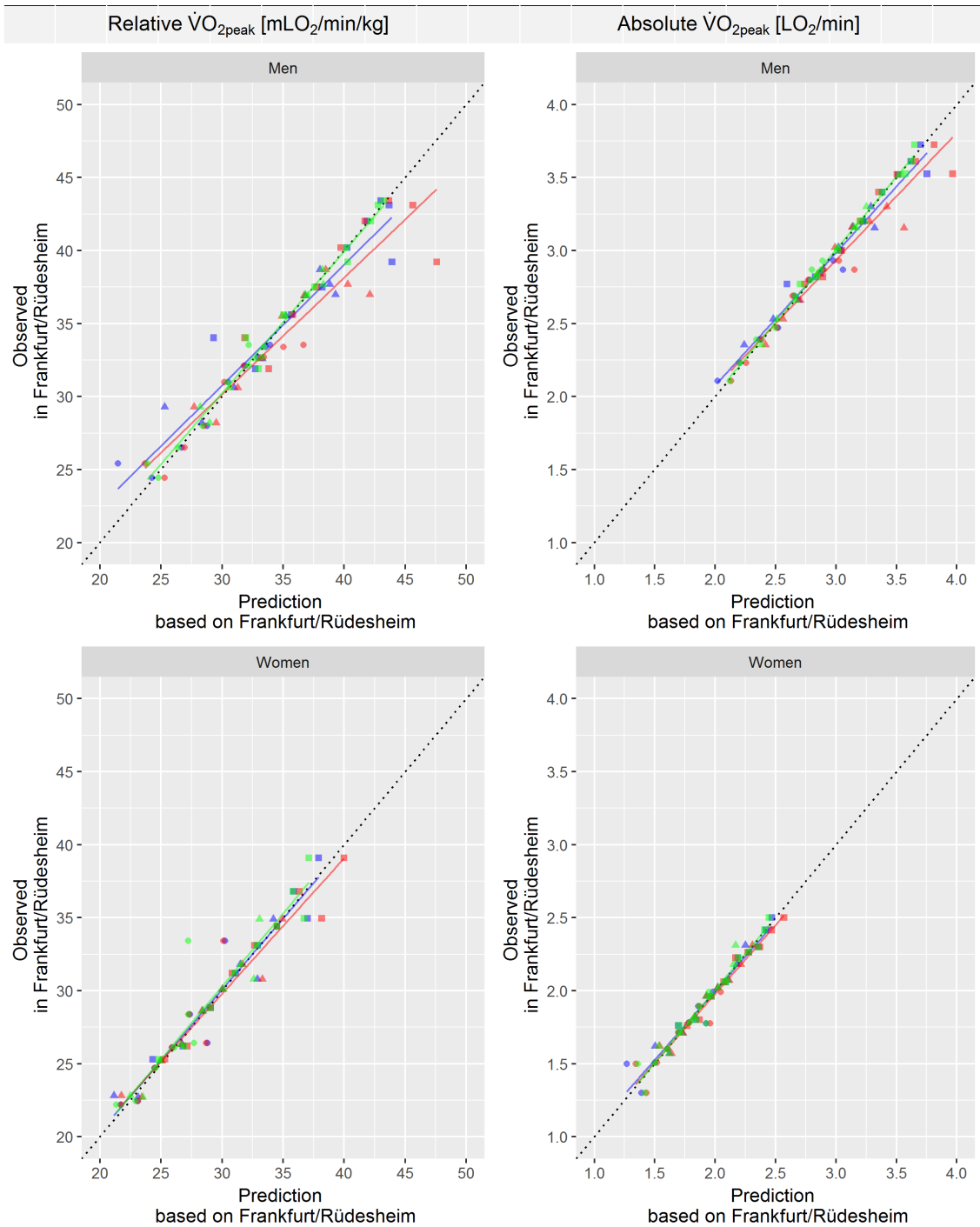
**Figure A.5:** Apparent validation box plots: Comparing regression models.

**Note:** Quantiles are displayed as  $\bullet = 0.25$ ,  $\blacktriangle = 0.5$ ,  $\blacksquare = 0.75$ . For apparent validation, the box plots as well as the regression predictions were based on participants from Frankfurt/Rüdesheim. Age was modelled in classes.

**Green:** quantile regression using b-spline smoothing

**Red:** linear quantile regression model

**Blue:** polynomial quantile regression model



**Figure A.6:** Apparent validation calibration plots.

**Note:** Quantiles are displayed as  $\bullet = 0.25$ ,  $\blacktriangle = 0.5$ ,  $\blacksquare = 0.75$ .

For apparent validation, the observed values as well as the regression predictions were based on participants from Frankfurt/Rüdesheim.

**Green:** quantile regression using b-spline smoothing

**Red:** linear quantile regression model

**Blue:** polynomial quantile regression model

**Table A.4:** Regression coefficients and R squared for linear regression in calibration data for apparent validation. Comparison of three regression models.

	Absolute $\dot{V}O_{2\text{peak}}$			Relative $\dot{V}O_{2\text{peak}}$		
	Linear	Spline	Polynomial	Linear	Spline	Polynomial
<b>Women</b>						
Intercept	0.09 [-0.1 to 0.27]	0.01 [-0.19 to 0.19]	0.11 [-0.12 to 0.33]	2.03 [-0.35 to 4.13]	0.27 [-2.12 to 2.96]	1.06 [-1.29 to 3.35]
Slope	0.94 [0.86 to 1.04]	1 [0.9 to 1.1]	0.94 [0.83 to 1.05]	0.93 [0.85 to 1.01]	1 [0.91 to 1.09]	0.97 [0.88 to 1.05]
R squared	0.96 [0.92 to 0.99]	0.96 [0.92 to 0.98]	0.95 [0.92 to 0.99]	0.95 [0.89 to 0.98]	0.92 [0.79 to 0.98]	0.95 [0.9 to 0.98]
<b>Men</b>						
Intercept	0.32 [0.02 to 0.64]	0.02 [-0.08 to 0.13]	0.26 [0.07 to 0.45]	6.21 [0.99 to 11.11]	1.16 [-0.62 to 3]	5.94 [1.35 to 9.91]
Slope	0.87 [0.76 to 0.98]	0.99 [0.96 to 1.03]	0.91 [0.84 to 0.97]	0.8 [0.65 to 0.96]	0.97 [0.92 to 1.02]	0.83 [0.71 to 0.96]
R squared	0.94 [0.88 to 0.99]	0.99 [0.99 to 1]	0.97 [0.95 to 0.99]	0.89 [0.8 to 0.96]	0.98 [0.96 to 0.99]	0.92 [0.85 to 0.97]

95% confidence intervals in square brackets.

**Table A.5:** Akaike Information Criterion AIC for different regression models.

Quantile	Absolute $\dot{V}O_{2\text{peak}}$			Relative $\dot{V}O_{2\text{peak}}$		
	Linear	Spline	Polynomial	Linear	Spline	Polynomial
<b>Female</b>						
.25	18562.02	18562.13	18563.99	2832.95	2830.88	2831.54
.50	18357.10	18358.27	18357.55	2623.22	2623.27	2623.08
.75	18794.05	18791.95	18790.79	3205.03	3201.94	3200.07
<b>Male</b>						
.25	32693.45	32659.89	32672.15	8611.33	8597.82	8604.32
.50	32425.59	32388.06	32406.82	8175.13	8145.55	8152.83
.75	33269.80	33219.23	33242.66	8975.45	8954.71	8957.62

## A.4 Informed consent



### Datenschutzrechtliche Einwilligungserklärung

Mir ist bekannt, dass **Prevention First** gesetzlich verpflichtet ist, im Rahmen der Untersuchung Daten über meine Person zu erheben und festzuhalten. Diese Daten werden in meiner Akte im Praxis-EDV-System gespeichert und im Abschlussbericht über meine Untersuchung dokumentiert, den ich nach dem Check-up erhalten werde.

Für allgemeine wissenschaftliche Auswertungen, z.B. über den Anteil von Patienten mit Bluthochdruck oder Diabetes, erfasst Prevention First die Check-up-Daten in einer eigenen Datenbank. Ich bin damit einverstanden, dass meine Daten zentral in dieser Datenbank gesammelt und ggf. anonymisiert statistisch ausgewertet werden.

Die Auswertungen erfolgen unter der vollen Wahrung der ärztlichen Schweigepflicht und des Datenschutzes. Mir ist bewusst, dass diese Einverständniserklärung freiwillig erfolgt. Bei einer Verweigerung entstehen mir keine Nachteile. Meine Einwilligung kann ich jederzeit widerrufen.

- Hiermit erkläre ich mein Einverständnis für die anonymisierte Datenauswertung.
- Ich stimme der anonymisierten Datenauswertung nicht zu.

\_\_\_\_\_  
Ort, Datum

\_\_\_\_\_  
Name, Unterschrift

**Prevention First** versendet per Email dreimal jährlich einen Newsletter mit aktuellen Informationen und Tipps zur Gesundheitsförderung und Prävention. Ich weiß, dass ich meine Einwilligung jederzeit widerrufen kann.

- Ich bin mit dem Versand des Newsletters an mich einverstanden.
- Ich stimme dem Newsletter-Versand an mich nicht zu.

\_\_\_\_\_  
Ort, Datum

\_\_\_\_\_  
Name, Unterschrift

Figure A.7: Informed consent

## B Danksagung

An dieser Stelle möchte ich meinen besonderen Dank an Personen richten, ohne die die Erstellung dieser Dissertation nicht möglich gewesen wäre.

An erster Stelle möchte ich mich bei Herrn Univ.-Prof. Dr. Stefan Wagenpfeil bedanken. Ihre freundliche Unterstützung und Hilfe war die Basis dafür, dass ich diese Arbeit begann und auch beenden konnte. Für Ihre Ratschläge bin ich besonders dankbar, da ich immer das Gefühl hatte, dass sie zu meinem Besten waren. Herzlichen Dank für Ihre Unterstützung.

Des Weiteren möchte ich mich bei Herrn Dr. med. Johannes Scholl und Herrn apl. Prof. Dr. med. Jürgen Scharhag für Ihre Hilfe sowohl an Werktagen als auch an Wochenenden bedanken. Herr Dr. med. Johannes Scholl hatte bereits im Jahre 2001 in weiser Voraussicht begonnen Routinedaten zu erheben. Dieses Engagement hat zu der beeindruckenden Datengrundlage dieser Arbeit geführt. Zusammen mit Herrn apl. Prof. Dr. med. Jürgen Scharhag wurde die Idee der Erstellung von Referenzwerten entwickelt. Besonders eindrücklich und konstruktiv werden mir die persönlichen Treffen zusammen mit Herrn Univ.-Prof. Dr. Stefan Wagenpfeil, Herrn Dr. med. Johannes Scholl und Herrn apl. Prof. Dr. med. Jürgen Scharhag in Rüdesheim und Heidelberg in Erinnerung bleiben.

Dissertationen Korrektur zu lesen ist eine mühsame Arbeit. Deshalb möchte ich mich ganz herzlich bei Chrisi, Jakob, Karin und Walde bedanken. Besonders dankbar bin ich auch meiner Freundin Frauke, die trotz der vielen Stunden Arbeit an dieser Dissertation immer ein offenes Ohr für Probleme hatte und mich immer unterstützte. Ungleich dankbarer als für die Hilfe bei dieser Arbeit bin ich jedoch für die vielen schönen Momente mit Euch.

Zuletzt danke ich meiner Mutter und meiner Familie, durch die ich in meinem ganzen Leben bedingungslose Unterstützung erfahren habe.