Tampere University

JULIA PIETILÄ

Quantification of Physical Activity and Sleep Behaviors with Wearable Sensors

Analysis of a Large-Scale Real-World Heart Rate Variability Dataset

Tampere University Dissertations 209

Tampere University Dissertations 209

JULIA PIETILÄ

Quantification of Physical Activity and Sleep Behaviors with Wearable Sensors

Analysis of a Large-Scale Real-World Heart Rate Variability Dataset

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Medicine and Health Technology
of Tampere University,
for public discussion in the TB109
of the Tietotalo, Korkeakoulunkatu 1, Tampere,
on 14 February 2020, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Medicine and Health Technology
Finland

Responsible supervisor and Custos	Docent, Senior Research Fellow Ilkka Korhonen Tampere University Finland	
Supervisor	Postdoctoral Researcher Elina Helander Tampere University Finland	
Pre-examiners	Adjunct Professor, Research Director Harri Sievänen The UKK Institute for Health Promotion Research Finland	Docent, Research Professor Mark van Gils VTT Technical Research Center of Finland Finland
Opponent	PhD Marco Altini HRV4Training Ltd The Netherlands	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Cover design: Roihu Inc.

ISBN 978-952-03-1432-3 (print)
ISBN 978-952-03-1433-0 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-1433-0

PunaMusta Oy – Yliopistopaino
Tampere 2020

"Science is more than a body of knowledge. It's a way of thinking."

Carl Sagan's last interview on 27 May 1996

PREFACE

The research presented in this thesis has been conducted at Tampere University (formerly Tampere University of Technology) in collaboration with Firstbeat Technologies Oy (Jyväskylä, Finland) during years 2013–2019. The research have been financially supported by the doctoral school of Tampere University (2016–2019) and by the Finnish Funding Agency for Innovation (TEKES) for Health Data Mining Project (2014–2016).

First, I would like to thank my supervisor, Adjunct Professor Ilkka Korhonen, for giving me the opportunity back in 2013 to start working on in the research group of Personal Health Informatics as a Master thesis worker and so introducing me to this interesting research field. He has given me insightful advices throughout my research activities, and set a good example to me how to present the research in an interesting and understandable way and how to put science into practice.

Secondly, I would give my greatest thanks to my instructor, D.Sc. (Tech.) Elina Helander, with whom I have had the pleasure to be closely working for several years. She has basically introduced me how to conduct research in theory as well as in practice. She has provided me her valuable support throughout my doctoral studies, which I am truly grateful.

I would also like to thank Firstbeat Technologies Oy and especially, M.Sc. Tero Myllymäki from the good collaboration we have had during these years. I am truly grateful that you have enabled this whole research by letting me to work on the Firstbeat dataset.

I also wish to thank my pre-examiners Docent, Research Professor Mark van Gils and Adjunct Professor, Research Director Harri Sievänen for their insightful comments and objective criticism of my thesis.

I would also like to thank all of my colleagues and co-authors in the publications of this thesis. Professor Urho Kujala, Ph.D Tiina Föhr, and Ph.D. Harri Lindholm from whom I have learned many insightful aspects about sports and exercise medicine as well as occupational health. I would also like to thank Professor Holly Jimison and Professor Misha Pavel who have provided me a great support and thought me to focus on the details without missing the big picture.

My gratitude goes also to my research group of Personal Health Informatics. In the research group, I have had the pleasure to work with many skillful and inspiring persons during the years. Especially, I would like to thank D.Sc. (Tech.) Hannu Nieminen who has always been positive and supportive. Also I would also like to thank my former colleague D.Sc. (Tech.) Jakub Parak and M.Sc. Milla Jauhiainen who have given me peer-support during my thesis writing process.

Last but not least, I am grateful to my family and friends who have always been there for me when I have needed them. My family, especially my mother and sister, for whom I can truly trust and who will always support and give their best advices when I need them. I am also grateful to all my friends who have been there to support me in every aspect of my life, and with whom I can relax and just be myself!

Espoo, November 2019

Julia Pietilä

ABSTRACT

Wearable monitoring devices, such as smartwatches, are used for monitoring personal health, fitness, health behaviors and well-being in daily life. Nowadays, wearable devices are popular and many consumers use them, in particular, to record their physical activity and sleep. Data recorded with wearable devices is an example of real-world data that can provide practical observations and insights on health and wellness, but its analyses pose challenges for research. Consumers conduct continuous recordings with wearable devices in non-research settings. Hence, any analysis of wearable real-world monitoring data must take into account the limitations and inaccuracies of the data, as well as sampling biases and incomplete representativeness of the population that arise from the uncontrolled data collection setting. To date, there are no well-established methods for analyzing health behaviors and well-being from continuous wearable monitoring data. Consequently, real-world health monitoring data is not commonly used for research although it could provide valuable observations and insights on health behaviors and well-being.

This thesis work aims at analyzing a large-scale real-world dataset of wearable heart rate variability (HRV) recordings to quantify the behaviors of physical activity (PA) and sleep that are one of the most important health behaviors. Specifically, the thesis focuses on the quantification methods and temporal patterns of PA behavior, as well as the associations that PA, alcohol intake and other lifestyles have with sleep. In addition, this thesis work aims to evaluate the feasibility to use real-world wearable monitoring data with applicable analysis methodologies for scientific research, and to demonstrate the observations and data-driven hypotheses that the results provide.

The study material was an anonymized real-world HRV monitoring dataset of 52,273 Finnish employees, which was gathered and prepared by Firstbeat Technologies Oy (Jyväskylä, Finland), a Finnish company providing and developing HRV analytics for stress, recovery and exercise. The dataset included three-day continuous HRV recordings performed in free-living settings combined with self-reports of alcohol intake, work and sleep times. The recordings were originally performed for a routine wellness program (Firstbeat Lifestyle Assessment) provided for the employees by their employers as a part of preventive occupational healthcare and health promotion program.

For the analysis of this thesis, PA behavior was quantified from the recordings using an HRV-based estimate of the oxygen uptake. Sleep was quantified by the regulation of the autonomic nervous system (ANS) using traditional HRV parameters and novel HRV-based indices of recovery. Both statistical and machine-learning methods were employed in the analysis for the thesis results.

Temporal variations in PA behavior were observed: the amount of PA was highest at the weekends and at the beginning of the year. The amount of PA quantified by the *absolute* oxygen consumption was higher for men than for women, and higher for younger than older subjects, and also higher for individuals of normal weight than obese. However, PA levels were more similar between the subjects when their physical fitness level was considered in quantifying PA. Moreover, PA behavior was associated with sleep. After a day including PA, the parasympathetic regulation of the ANS and recovery during sleep were diminished, but regular PA seemed to increase parasympathetic regulation of the ANS and aid recovery during sleep.

The most important predictor for ANS regulation during sleep was, however, acute alcohol intake. Acute alcohol intake dose-dependently diminished the parasympathetic regulation of the ANS and recovery during sleep, an effect that was already observable after only 1–2 standardized units of alcohol. Moreover, the same alcohol intake, normalized by the body weight, seemed to affect the ANS regulation more in younger subjects than in the older ones, but was similar for both sedentary and physically active subjects, as well as for both men and women.

Many of the results obtained in this thesis accord with the findings of previous studies, such as the higher PA level on weekends, the higher amount of *absolute* intensity PA in men, younger and normal weight subjects, and the relationship of PA and alcohol intake with the ANS regulation during sleep. On the other hand, the results of this thesis provide new observations, for example, about the interaction between alcohol intake and subject's background characteristics that could not have been studied before due to the limited and homogenous study populations.

In conclusion, the results of this thesis demonstrates that real-world wearable monitoring data can be feasible for scientific research and its results not only supports the findings of existing studies but also provides new observations, insights and data-driven hypotheses. The real-world evidence facilitates our understanding of aspects of health behaviors and wellness that cannot be studied in the more traditional, controlled research settings. These real-world insights can be further used for designing more personalized and targeted health interventions and as tools for promoting health and well-being.

TIIVISTELMÄ

Puettavia mittalaitteita, kuten älykelloja, voidaan käyttää arjessa oman terveydentilan, fyysisen kunnon, terveyskäyttäytymisen sekä hyvinvoinnin seuraamiseen. Puettavien mittalaitteiden käyttö on nykyisin suosittua, ja kuluttajat mittaavat niillä yleensä liikuntaa ja unta. Puettavien mittalaitteiden keräämä mittausaineisto on esimerkki arkielämän aineistoista (real-world data), jotka voivat tarjota käytännönläheisiä havaintoja terveydestä ja hyvinvoinnista. Arkielämässä kerättyjen aineistojen hyödyntäminen tutkimustarkoituksiin on kuitenkin haastavaa, sillä kuluttajat käyttävät puettavia mittalaitteita vapaaehtoisesti arkielämän olosuhteissa. Siksi aineiston käsittelyssä on otettava huomioon aineiston keräyksen kontrolloimattomat tutkimusasetelmien ulkopuoliset olosuhteet, jotka aiheuttavat mittausaineistoon tyypillisesti epätarkkuutta ja puutteellisuutta sekä otospopulaation valikoituneisuutta. Puettavien mittalaitteiden tuottamille jatkuva-aikaisille aineistoille ei myöskään toistaiseksi ole vakiintuneita käsittelytapoja. Näiden tekijöiden vuoksi puettavien mittalaitteiden keräämiä aineistoja käytetään nykyisin vielä vain vähän tutkimuksissa, vaikka ne voivat tarjota uusia havaintoja terveyskäyttäytymisestä ja hyvinvoinnista.

Väitöstyössä hyödynnetään puettavan sydämen sykevälivaihtelua mittaavan laitteen tuottamaa arkielämän suurta aineistoa määrittämään liikuntaan ja uneen liittyvää käyttäytymistä. Liikunta ja uni ovat tärkeitä terveyskäyttäytymisen tekijöitä, ja väitöstyössä tutkitaan erityisesti liikunnan määrittämisen menetelmiä, liikuntakäyttäytymisen ajallista vaihtelua, sekä liikunnan, alkoholin nauttimisen ja muiden elämäntapojen vaikutusta uneen. Lisäksi väitöstyön tavoitteena on arvioida puettavien mittalaitteiden tuottamien suurten arkielämän aineistojen ja niiden hyödyntämisen soveltuvuutta tieteellisen tutkimukseen sekä osoittaa näiden aineistojen tarjoamia uusia havaintoja ja näkökulmia terveydestä ja hyvinvoinnista.

Väitöstutkimuksen aineistona käytettiin 52 273 suomalaisen työntekijän tunnisteettomia arkielämässä tehtyjä sydämen sykevälivaihtelun mittauksia, jotka oli alun perin tehty osana terveyttä edistävää ja ennaltaehkäisevää terveydenhuoltoa. Aineisto on kerätty Firstbeat Technologies Oy:n toimesta, joka kehittää ja tarjoaa sykevälivaihtelun analyysimenetelmiä liikunnan, stressin ja palautumisen arviointiin. Aineisto sisälsi kolmipäiväisiä jatkuva-aikaisia mittauksia sydämen sykevälivaihtelusta sekä itseraportointeja nautitusta alkoholin määrästä sekä työ- että nukkumisajoista.

Väitöstyössä liikunnan määrittämisessä hyödynnettiin sykevälivaihteluun perustuvaa hapenoton arviota. Unta arvioitiin autonomisen hermoston säätelyn kautta käyttäen perinteisiä sykevälivaihtelumuuttujia sekä uudenlaisia sykevälivaihteluun perustuvia palautumismuuttujia. Väitöstyön tulokset pohjautuvat sekä perinteisiin tilastollisiin että koneoppimisen menetelmiin.

Liikuntakäyttäytymisessä havaittiin ajallista vaihtelua: liikunnan määrä oli korkein viikonloppuisin sekä alkuvuonna. Kun liikuntaa arvioitiin *absoluuttisella* hapenotolla, liikunnan määrä oli korkeampi miehillä kuin naisilla, ja nuoremmilla kuin vanhemmilla sekä normaalipainoisilla kuin lihavilla henkilöillä. Toisaalta kun liikunnan määrää arvioitiin ottaen huomioon henkilöiden kuntotaso, erot liikunnan määrässä henkilöiden välillä pieneni huomattavasti. Lisäksi liikuntakäyttäytymisellä havaittiin olevan yhteys uneen. Päivällä harrastettu liikunta näytti heikentävän autonomisen hermoston parasympaattista säätelyä unen aikana, mutta säännöllinen liikunta näytti lisäävän parasympaattista säätelyä ja palautumista unen aikana.

Unen aikaisen autonomisen hermoston säätelyn kannalta tärkein tekijä oli kuitenkin päivän aikana nautittu alkoholi. Jo 1–2 alkoholiannosta heikensi autonomisen hermoston parasympaattista säätelyä unen aikana ja tämä säätely heikkeni sitä enemmän, mitä useampia alkoholiannoksia päivän aikana nautittiin. Painoon suhteutettu, sama alkoholimäärä näytti vaikuttavan autonomisen hermoston säätelyyn enemmän nuoremmilla kuin vanhemmilla henkilöillä, mutta samalla tavalla sekä paljon että vähän liikuntaa harrastavilla henkilöillä, ja sekä miehillä että naisilla.

Monet väitöstyön tulokset tukevat aiempia tutkimustuloksia, kuten esimerkiksi havainnot suuremmasta liikunta-aktiivisuudesta viikonloppuisin, miesten, nuorten ja normaalipainoisten suuremmasta liikuntamäärästä *absoluuttisella* hapenottomäärällä mitattuna, sekä liikunnan ja alkoholin yhteydestä autonomisen hermoston säätelyyn unen aikana. Toisaalta väitöstyössä havaittiin esimerkiksi myös alkoholin nauttimisen ja henkilön taustatekijöiden yhteisvaikutuksia autonomisen hermoston säätelyyn, joita ei ole voitu aiemmin tutkia pienten tutkimuspopulaatioiden vuoksi.

Kokonaisuudessaan väitöstyö osoittaa, että puettavien mittalaitteiden tuottamat arkielämän aineistot soveltuvat tieteelliseen tutkimukseen ja tulokset tukevat aiempia tutkimustuloksia, mutta tarjoavat myös uusia havaintoja sekä näkemyksiä. Tosielämän tieto voikin parantaa terveyskäyttäytymisen ja hyvinvoinnin tuntemusta, erityisesti niiltä osin, joihin perinteiset tutkimusasetelmat eivät sovellu. Käytännössä tosielämän havaintoja ja tietoa voidaan käyttää havainnollistamaan käyttäytymisen vaikutusta terveyteen ja hyvinvointiin, sekä tukemaan terveyskäyttäytymisen muutosta entistä henkilökohtaisemmin ja kohdennetummin.

CONTENTS

List of Figures

List of Tables

ABBREVIATIONS

ANOVA	analysis of variance
ANS	autonomic nervous system
BMI	body-mass-index
DAN	diabetic autonomic neuropathy
ECG	electrocardiogram
EE	energy expenditure
HF	high-frequency component of heart rate variability
HR	heart rate
HR_{max}	maximal heart rate
HRV	heart rate variability
H_0	null hypothesis
H_1	alternative hypothesis
LF	low-frequency component of heart rate variability
LF/HF ratio	the ratio between the low- and high-frequency components in the heart rate variability
$LightPA_{VO2max}$	number of minutes during which the oxygen uptake is within 20–30% of the maximal oxygen uptake
LR	linear regression model
MAE	mean absolute error
MET	metabolic equivalent
MI	myocardial infarction
$MVPA_{MET}$	the number of minutes during which the oxygen uptake is at least three times higher than during rest (≥3 METs)
$MVPA_{MET}$10min	number of minutes during which the oxygen uptake is at least three times higher than during rest (≥3 METs) for at least 10 minutes, with exception of one minute in which the oxygen uptake may me lower (<3 METs)
$MVPA_{VO2R}$	number of minutes during which the oxygen uptake exceeds the limit of 40% of the oxygen uptake reserve
MSE	mean squared error

NN-interval	time interval between two consecutive normal sinus beats in ECG
NREM	non-rapid eye movement
OOB	out-of-bag
PA	physical activity
PA_{VO2max}	number of minutes during which the oxygen uptake exceeds the limit of 30% of the maximal oxygen uptake
PA class	self-reported physical activity class
PNS	parasympathetic nervous system
PPG	photo plethysmography
PSD	powers spectral density
R^2	coefficient of determination
RCT	randomized controlled trial
REM	rapid eye-movement
RF	random forest
RFE	recursive feature elimination
RMSSD	the root mean square of successive differences in the RR-intervals
RR-interval	time interval between two consecutive R peaks in the ECG
RSA	respiratory sinus arrhythmia
RWE	real-world evidence
RWD	real-world data
R^2	coefficient of determination
SA	sinoatrial
SD	standard deviation
SDNN	standard deviation of the NN-intervals
SNS	sympathetic nervous system
ULF	ultra low-frequency component of heart rate variability
VLF	very low-frequency component of heart rate variability
VO_2	oxygen uptake
VO_2R	oxygen uptake reserve
VO_{2max}	maximal oxygen uptake
VO_{2rest}	oxygen uptake at rest
% VO_2R	percentage of oxygen uptake reserve

ORIGINAL PUBLICATIONS

Publication I PIETILÄ, J., MUTIKAINEN, S., HELANDER, E., MYLLYMÄKI, T., KUJALA, U.M. and KORHONEN, I., 2015. Methods to Use Big Wearable Heart Rate Data for Estimation of Physical Activity in Population Level, I. LACKOVIC and D. VASIC, eds. In: *6th European Conference of the International Federation for Medical and Biological Engineering 2015*, Springer, pp. 66-69.

Publication II KUJALA, U., PIETILÄ, J., MYLLYMÄKI, T., MUTIKAINEN, S., FÖHR, T., KORHONEN, I. and HELANDER, E., 2017. Physical activity: Absolute intensity vs. relative-to-fitness-level volumes. *Medicine and Science in Sports and Exercise*, **49**(3): pp. 474-481.

Publication III PIETILÄ, J., HELANDER, E., MYLLYMÄKI, T., KORHONEN, I., JIMISON, H. and PAVEL, M., 2015. Exploratory analysis of associations between individual lifestyles and heart rate variability-based recovery during sleep, *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2015*, IEEE, pp. 2339-2342.

Publication IV PIETILÄ, J., HELANDER, E., KORHONEN, I., MYLLYMÄKI, T., KUJALA, U.M. and LINDHOLM, H., 2018. Acute Effect of Alcohol Intake on Cardiovascular Autonomic Regulation During the First Hours of Sleep in a Large Real-World Sample of Finnish Employees: Observational Study. *JMIR Mental Health*, **5**(1), pp. e23.

Publication V FÖHR, T., PIETILÄ, J., HELANDER, E., MYLLYMÄKI, T., LINDHOLM, H., RUSKO, H. and KUJALA, U.M., 2016. Physical activity, body mass index and heart rate variability-based stress and recovery in 16 275 Finnish employees: a cross-sectional study. *BMC Public Health*, **16**:701, doi: 10.1186/s12889-016-3391-4.

AUTHOR'S CONTRIBUTIONS

Publication I The author participated in designing the data preprocessing and analysis schemes, and was responsible for conducting the data preprocessing and analysis. Moreover, the author was responsible for outlining and writing the paper.

Publication II The author was responsible for designing and implementing the statistical methods and writing the methods and results sections of the journal paper. U.M. Kujala was responsible for writing the introduction, conclusions and discussion sections of the conference paper.

Publication III The author participated in designing the data quantification and analysis procedures, and the author was responsible for writing the conference paper.

Publication IV The author participated in designing and conducting the data analysis, and was responsible for writing the final version of the paper. E. Helander was responsible for initiating the data analysis and writing the methods, results and discussion sections. H. Lindholm was responsible for initiating writing the introduction section of the paper.

Publication V The author was responsible for designing and implementing the statistical data analysis of the paper. The author also wrote the methods and results sections of the paper. T. Föhr participated in designing the overall statistical analysis, and was responsible for the study design and for writing the introduction, discussion and conclusions sections of the paper. This Publication has been also included in the doctoral dissertation of T. Föhr titled "The relationship between leisure-time physical activity and stress on workdays with special reference to heart rate variability analyses", published in 2016 by the Faculty of Sports and Health Sciences of the University of Jyvaskyla.

1 INTRODUCTION

During recent years, the market has been flooded with wearable sensors, i.e. small health-monitoring devices or sensors embedded into clothing (Korhonen, Pärkkä and van Gils, 2003; Swan, 2012; Piwek et al., 2016). Advances in technology have resulted in unobtrusive, non-invasive, affordable and user-friendly wearables, such as smartwatches and wristbands, which are feasible for long-term monitoring of physiological signals in free-living settings (Korhonen, Pärkkä, and van Gils, 2003; Majumder, Mondal and Deen, 2017). Based on the monitored signals, the wearable devices calculate parameters related to the users' health, wellness and behaviours (Figure 1) (Piwek et al., 2016; Majumder, Mondal and Deen, 2017). The parameters presented by the current wearable devices are typically only some simple summary statistics, such as the number of steps or duration of sleep (Piwek, 2016). The simple parameters provided by the wearable sensors always require an interpretation that takes into account the individual context before any insights about the subject's well-being and health behaviors can be generated (Figure 1).

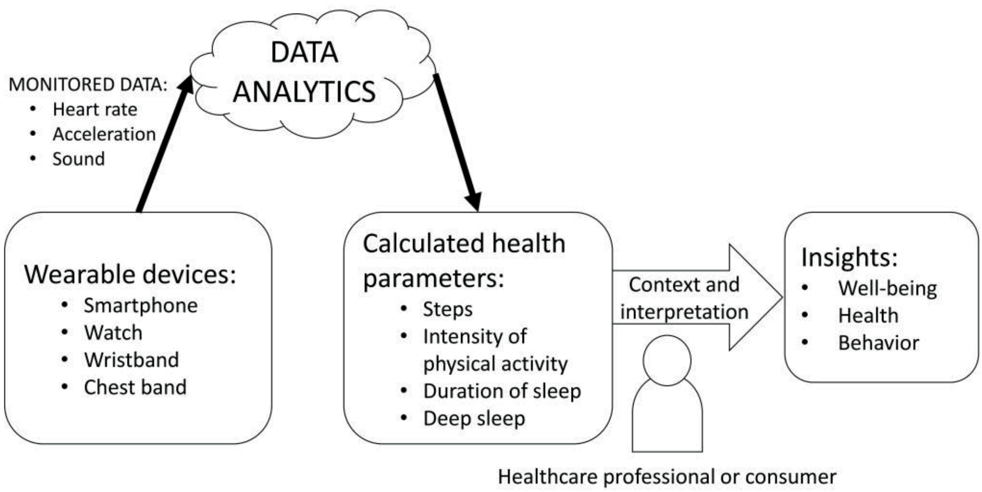


Figure 1. The operating principle of wearable devices for assessing health and well-being. Wearable devices monitor signals from which the algorithms calculate health-related parameters that require interpretation to have insights on the subject's health and well-being.

As the wearable device market has grown, wearable devices have gained popularity among consumers and become increasingly common among the general public, but especially in healthy individuals who are interested in their health and well-being (Swan, 2012; Piwek et al., 2016). The increased availability of wearable consumer devices has also significantly expanded research in the area of wearable monitoring (Silfee et al., 2018). For example, a PubMed search for the keyword "wearable devices", yielded 244 articles published in 2008, yet in 2018, there were 1,951 such articles published.

Physical activity (PA) and sleep are the most common health behaviours and wellness indicators monitored by the currently available wearable consumer devices. PA and sleep are key health behaviours, and both physical inactivity and poor sleep have been associated with adverse health outcomes (Haskell, Blair and Hill, 2009; Porkka-Heiskanen, Zitting and Wigren, 2013). For example, chronic poor sleep contributes to harmful changes in biological processes that may predispose the sufferer to cardiovascular disease and type 2 diabetes (Luyster et al., 2012), while PA is associated with biomarker profiles that help prevent those diseases (Physical Activity Guidelines Advisory Committee, 2008). There also seems to be a clear interaction between PA and sleep; physically active individuals sleep better than physically inactive ones (Physical Activity Guidelines Advisory Committee, 2018). Physical inactivity and obesity tend also to form a vicious circle: physical inactivity triggers obesity that may lead to less activity (Pietiläinen et al., 2008). The global trend for the increasing prevalence of obesity and insomnia-related symptoms can also be observed in the Finnish adult population (Ng et al., 2014; Helldán and Helakorpi, 2015; Pallesen et al., 2014; Ford et al., 2015; Kronholm et al., 2016). Thus, the promotion of good sleep and physically active lifestyles are crucial for the general public health (Physical Activity Guidelines Advisory Committee, 2018).

The increased popularity of wearable monitoring data has opened up new possibilities for research. In the health domain, research has traditionally relied on controlled research studies, especially on randomized controlled trials (RCTs). These RCTs are good for studying causality and finding mechanisms for associations, for example, between health parameters and behaviours (Sherman et al., 2016; Berger et al., 2017). However, due to the specific and limited populations of such studies, the results of RCTs cannot always be generalized to the whole population. In addition, long-term monitoring in large-scale RCTs may not be feasible (Sherman et al., 2016). Thus, the knowledge gained from RCTs can be complemented by so-called real-world evidence (RWE) (Booth and Tannock, 2014; Sherman et al., 2016). RWE consists of the results from the analysis of real-world data (RWD) (Sherman et al., 2016; Berger et al., 2017). RWD is data gathered outside traditional research studies and is typically from a broad population (Sherman et al., 2016; Berger et al., 2017). In the context of wearable monitoring, a great source for RWD can be the wearable monitoring data from ordinary consumers who give consent for their personally recorded data to be used for research and development purposes (Piwek et al., 2016). The RWE gained from wearable monitoring data may provide valuable observations and insights on the health behaviour and wellness of people in their everyday lives; evidence which could not be obtained from more traditional research studies (Sherman et al., 2016). Furthermore, RWD analysis can also help to formulate hypotheses about the causalities, which could be further studied and verified with RCTs (Sherman et al., 2016). RCT and RWE studies can be regarded as complementary in that they study the same phenomenon, but in different populations under different settings and with different data and methods (Kim, Lee and Kim, 2018; Maissenhaelter, Woolmore and Schlag, 2018). Table 1 summarizes and compares the most relevant characteristics of RCT and RWE studies.

Table 1. Characteristics of randomized controlled trials and real-world data studies.

Characteristic	Randomized controlled trials (RCTs)	Real-world data (RWD) studies
Population	• Limited and specific • ”Ideal study population”	• Large-scale and broad • ”Real-world study population”
Settings	• Controlled • Data collected over a limited time period	• Uncontrolled • Data may be collected over extended time periods
Design	• Typically interventional • Well-designed beforehand	• Typically observational • Adjusted for the available data
Data	• High quality data collected for the study	• Low quality • Secondary use of data
Analyses	• Hypothesis-driven analyses with traditional statistical methods using small samples	• Exploratory analyses with statistical and machine-learning methods using large samples
Results	• Good internal validity	• Good external validity
New knowledge	• Causality and mechanisms in health-related issues	• Observations and data-driven hypotheses on health-related issues
Disadvantages	• The study population does not represent a clinical sample • The results are not generalizable • Expensive to conduct	• The data may be poor quality including confounding factors • Poor internal validity • Data fishing is possible
Advantages	• High internal validity of the results and randomization ensures unbiased results for the study population	• Generalizable results providing practical insights • Cost-effective in enabling long-term studies

Despite the great opportunities the analysis of real-world wearable monitoring data presents, it also poses some challenges. Firstly, continuous wearable monitoring is still a relatively new field of science, and the methods used to quantify health behaviours from the recorded data are, to say the least, under development, i.e., there are no universally accepted 'gold standards'. In addition, there are more general concerns about RWD analysis, which also apply to wearable monitoring data. For example, the huge quantity of data raises the issue of "data dredging" (or "data fishing"), which arises from the performance of multiple analyses on the data in order to achieve the desired results (Smith and Ebrahim, 2002; Berger et al., 2017). Furthermore, the uncontrolled settings for RWD data collection may introduce bias in the selection of the subjects (Vandenbroucke et al., 2007), uncertainties in the quality of the data (Sherman et al., 2016), and the presence of confounding factors potentially causing distortion in the studied associations (Onghena and van der Noortgate, 2005). However, these concerns can be addressed, to some extent, by rigorous study design and statistical modelling (Berger et al., 2009).

This thesis work is based on the results of five, scientific peer-reviewed studies published between 2014 and 2018. The material used in the publications is an anonymized, real-world dataset consisting of heart rate variability (HRV) monitoring data. The HRV data in free-living settings was collected as part of a routine wellness program (Firstbeat Lifestyle Assessment) provided by Finnish employers as part of a preventative occupational healthcare program. By June 2015, 52,273 Finnish employees had participated in the three-day HRV monitoring procedure. In addition to the wearable HRV monitoring data, the dataset included the subjects' background characteristics, and their self-reported work and sleep times as well as alcohol intake per day. The five publications have all used statistical data analysis and machine learning on this HRV monitoring dataset in order to study health behaviours and physical activity, and the associations of lifestyle choices and physical activity with sleep.

A review of the relevant literature on the autonomic nervous system (ANS), the physiological background to assessing heart rate (HR) and HRV, the wearable monitoring and analysis for HRV, as well as aspects of RWD and RWE are discussed in Chapter 2. Chapter 3 outlines the aims of the study, and Chapter 4 describes the study material and the methods employed. Chapter 5 presents the main results of the research, which are discussed in Chapter 6. Chapter 7 summarizes the results of the research and draws some conclusions about the physical activity and sleep behaviours observed in the Finnish employees.

2 REVIEW OF THE LITERATURE

2.1 Autonomic nervous system and heart rate variability

2.1.1 Autonomic nervous system

The purpose of the autonomic nervous system (ANS) is, in collaboration with the endocrine system, to maintain homeostasis in the body (Furness, 2009). To achieve homeostasis, i.e. a steady physiological state in the body (Koizumi, 2009), the ANS is continuously controlling the bodily functions (Furness, 2009). The ANS consists of afferent and efferent nerves, which are innervated into almost all organs (Ernst, 2017). The afferent nerves gain sensory information from the organs (Ernst, 2017) down to the lowest-level parts of the central nervous system, including the spinal cord, the brain stem and the hypothalamus (Gabella, 2012). The lowest-level parts of the central nervous system involuntarily control the reactions transmitted via the efferent nerves to the target organs (Ernst, 2017). Traditionally, the ANS is divided into two branches: the sympathetic (SNS) and parasympathetic (PNS) nervous systems (Ernst, 2017). SNS and PNS typically affect the target organs in opposite but complementary ways (Figure 2) (Furness, 2009; Vinik, 2012). The opposite effects of SNS and PNS on the target organs enable efficient adjustments for the requirements of the body and its environment in order to maintain homeostasis (Furness, 2009).

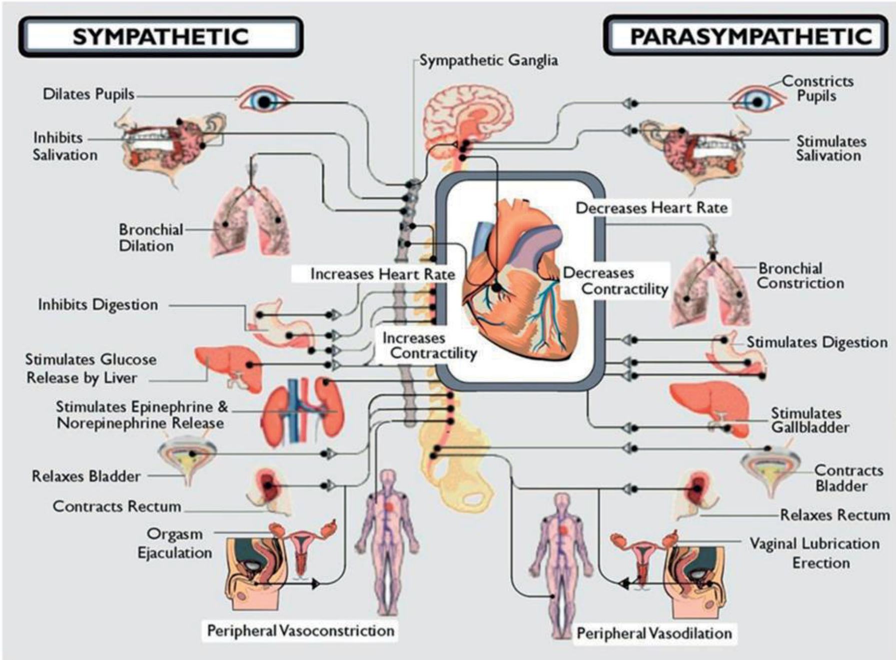


Figure 2. The sympathetic and parasympathetic branches of the ANS and their effects on the target organs (Vinik, 2012). Licensed under Creative Commons Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0).

Traditionally, SNS and PNS regulation has been regarded as reciprocal; the activation of one branch is linked to the inhibition of the other branch (Malliani et al., 1998). This reciprocal behavior of SNS and PNS is observed in many physiological conditions (Montano et al., 2009), but the co-activation of both SNS and PNS also occurs in various physiological conditions, e.g. during exercise (Koizumi et al., 1982; Malliani and Montano, 2002; Paton et al., 2005). Thus, the interaction between SNS and PNS activation is complex and non-linear: PNS activation may be associated with either inhibited, increased or unchanged SNS activation, and vice versa (Shaffer and Ginsberg, 2017). In other words, the activation of SNS and PNS should be interpreted as a model with a two-dimensional surface, instead of as a one-dimensional vector (Berntson, Cacioppo and Quigley, 1991). It has been suggested that this complex and non-symmetrical activation of the SNS and PNS provides the most efficient adaptation to maintain homeostasis (Koizumi et al., 1982; Paton et al., 2005).

A dynamic balance of the SNS and the PNS is a marker of a healthy organism (Shaffer, McCraty and Zerr, 2014). A healthy organism has the capacity to respond and adjust to the autonomic balance according to the requirements for maintaining homeostasis (McCraty and Shaffer, 2015). The lack of dynamic balance is a risk factor for various diseases (Thayer and Sternberg, 2006; Montano et al., 2009). In autonomic imbalance, one ANS branch dominates the other; typically the SNS over the PNS (Thayer and Sternberg, 2006). Autonomic imbalance has been associated with hypertension and an elevated risk of cardiovascular disease, for example (Brook and Julius, 2000).

2.1.2 Electrocardiogram, heart rate and heart rate variability

The heart is one of the target organs of the ANS (Gabella, 2012). The functions of the heart are largely regulated by the ANS (Thayer and Sternberg, 2006; Montano et al., 2009). However, the regulation of the autonomic heart functions is a complex phenomenon including several interconnected biological systems and mechanisms (Shaffer, McCraty and Zerr, 2014). The heartbeat is a contraction of the heart muscle that originates from the spontaneously depolarized pacemaker cells in the sinoatrial (SA) node of the heart (e.g. Vander, Sherman and Luciano, 1990). From the pacemaker cells, the action potential propagates through the heart causing the heart muscle to contract and pump blood into vessels.

Electrocardiography is the gold standard for measuring the electrical activity of the heart (e.g. Vander, Sherman and Luciano, 1990). In electrocardiography, electrodes are attached to the skin and the measured electrical activity of the heart is shown as an electrocardiogram (ECG), a graph of voltage over time. The events in the ECG correspond to the electrical activity of the heart (Figure 3). The P-wave in the ECG corresponds to the atrial depolarization, the QRS-complex corresponds to the ventricular depolarization and simultaneous atrial repolarization, and the T-wave corresponds to the ventricular repolarization. The duration of a heartbeat is typically derived from the time between the consecutive R-waves in the ECG (Georgiou et al., 2018). The time intervals between consecutive R-waves are called RR-intervals (Vander, Sherman and Luciano, 1990). The instantaneous heart rate (HR) is the number of contractions of a heart muscle in a time unit, and it can be determined from the duration of a heartbeat, which in ECG measurements is equivalent to the RR-intervals (Vander, Sherman and Luciano, 1990).

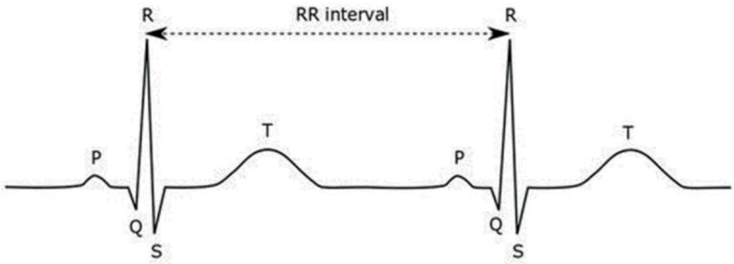


Figure 3. An example of ECG signal showing a P-wave, a QRS-complex and a T-wave as well as an RR-interval or an inter-beat interval (Pang and Igasaki, 2018). Licensed under CC Attribution 4.0 International (CC BY 4.0).

The intrinsic HR generated by the pacemaker cells in the SA node is around 100 beats per minute (bpm) (Vander, Sherman and Luciano, 1990; Opthof, 2000). However, HR may vary in adults from as low as 30 bpm during rest to up to 200 bpm during exercise, through ANS regulation (Shaffer, McCraty and Zerr, 2014). In general, SNS activation increases HR while PNS activation decreases it (Gabella, 2012). Thus, the PNS is in primary control of the heart during rest, even though the SNS and PNS are in constant interaction (Vander, Sherman and Luciano, 1990; Malik et al., 1996; McCraty and Shaffer, 2015). Compared to the SNS stimulus, the PNS stimulus has immediate, short-term effects on HR (Shaffer, McCraty and Zerr, 2014; Draghici and Taylor, 2016). The PNS stimulus affects HR with a delay of less than one second and the effect only lasts for one or two heart beats (Nunan, Sandercock and Brodie, 2010; Shaffer, McCraty and Zerr, 2014). The SNS stimulus has a delay of more than five seconds and its effects last for 5–10 seconds (Nunan, Sandercock and Brodie, 2010; Shaffer, McCraty and Zerr, 2014).

As heart functions are largely under the control of the ANS, resting HR and the decline in HR after exercise can be used as rough estimates of the ANS regulation (Thayer and Sternberg, 2006; Montano et al., 2009; Shaffer, McCraty and Zerr, 2014). A more sophisticated estimate of the ANS regulation can be obtained with the beat-to-beat variation in instantaneous HR or RR-intervals (Malik et al., 1996). The variation in the instantaneous HR or RR-intervals is known as heart rate variability (HRV) (Malik et al., 1996).

Both HR and HRV are non-invasive and indirect markers of ANS regulation in healthy and diseased subjects (Malik et al., 1996; Thayer, Yamamoto and Brosschot, 2010; Billman, 2011). Typically, HRV is assessed from an ECG as the variation of RR-intervals by detecting the QRS complexes in the ECG (Malik et al., 1996). The QRS complexes are detected, for example, by the fiducial point (Pan and Tompkins,

1985) and, using the time points of QRS complexes, the RR-intervals may be extracted. Both the RR-interval tachogram, which represents the RR-intervals with respect to time, and the RR-intervals presented as a function of the heartbeat index number can be used as the basis for HRV analysis (Figure 4).

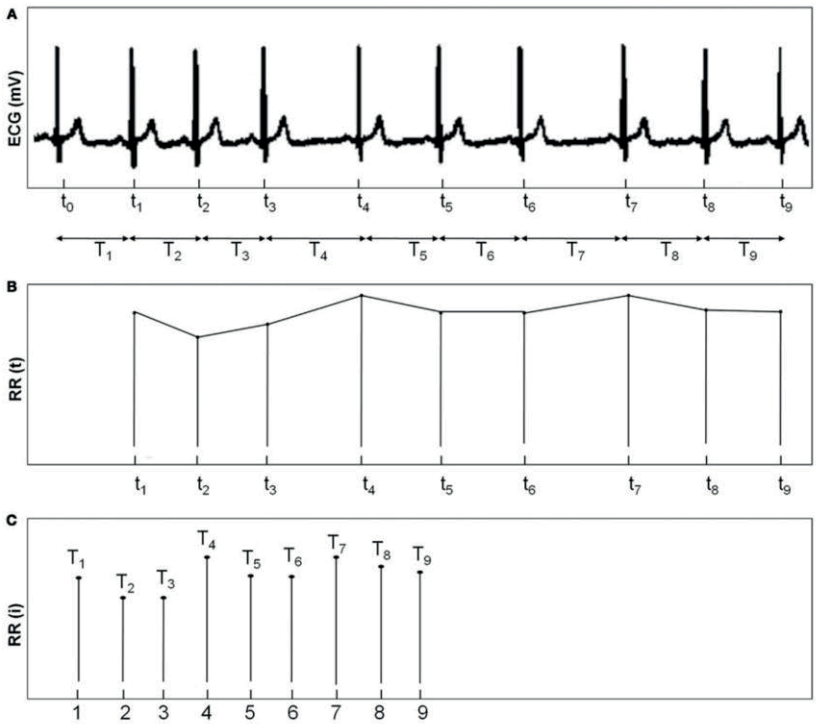


Figure 4. Example of an ECG signal (A) from which the time between the consecutive R peaks are extracted. The interpolated RR-interval tachogram (B) and the RR-intervals as a function of the heartbeat index number (C) can used for the HRV analysis. (Peltola, 2012). Licensed under CC BY-NC 3.0.

Although the principal effects of PNS and SNS on the HR are widely and commonly accepted, the exact effects and contributions of PNS and SNS on the HRV are still under investigation and open to debate (Billman, 2011; Billman et al., 2015). The two main mechanisms responsible for the HRV are baroreflex and respiration (Karemaker, 2017). Baroreflex contributes to HRV via PNS activation (Karemaker, 2017). Fluctuations of HR during a respiration cycle, known as respiratory sinus arrhythmia (RSA), has also been linked to PNS regulation (Denver, Reed and Porges, 2007; Karemaker, 2017). However, the PNS activity is not the only factor affecting RSA, as the SNS mechanisms also modulate it (Taylor et al., 2001; Karemaker, 2017).

In addition to PNS and SNS, the chemo- and mechano-sensory neurons located in the heart are part of the intra-cardiac nervous system which ultimately controls the pacemaker cells in the SA node (Billman, 2011; Shaffer, McCraty and Zerr, 2014; Campos et al., 2018). Thus, the autonomic control of the heart includes regulatory functions from the brain to the heart, and both HR and HRV represent the net effect of the autonomic control of the heart (Shaffer, McCraty and Zerr, 2014; McCraty and Shaffer, 2015; Campos et al., 2018).

HR and HRV have been shown to have a non-linear and inverse relationship known as the cycle length dependence (Tsuji et al., 1996; Monfredi et al., 2014; McCraty and Shaffer, 2015). This is justified for both physiological and mathematical reasons (Zaza and Lombardi, 2001; Melenovsky et al., 2005; Sacha and Pluta, 2008). It has been proposed that the SA pacemaker cells have an intrinsic property causing the RR-intervals to be non-linearly dependent on the cardiac activity: the same level of vagal activity triggers increased prolongation of RR-intervals at longer baseline RR-intervals (Zaza and Lombardi, 2001; Melenovsky et al., 2005). The mathematical dependence between HRV and RR-intervals arises from the non-linear inverse relationship between the average HR and RR-intervals (Sacha and Pluta, 2008). At low average HRs, the fluctuations of the RR-intervals can be higher than at higher average HRs (Sacha, 2014). In addition, the same fluctuations in the HR have a relatively greater impact on the RR-intervals at low average HRs than at higher average HRs (Billman et al., 2015). Thus, HRV is higher at low-average HRs and lower when the average HR is higher (Sacha and Pluta, 2008).

To remove the mathematical dependence between the HRV and HR, the HRV may be normalized with respect to the HR level (Sacha and Pluta, 2008). A simple HRV normalization method is to divide the RR-intervals by the average RR-interval (Sacha and Pluta, 2008). If the HRV is not normalized, the HR should be reported together with the HRV (Shaffer, McCraty and Zerr, 2014). It is especially important to notice the cycle length dependence when comparing the HRV between subjects with different average HRs, or during interventions which may alter the HR (Billman et al., 2015).

2.1.3 Individual characteristics affecting heart rate and heart rate variability

In addition to HR, age has been reported to be an important determinant of HRV (Tsuji et al., 1996; Umetani et al., 1998). In one study, the HR and age together were reported to account for up to 50% of the variance in HRV: HR accounting for up to 25% and age accounting for up to 40% of the variance in the HRV (Tsuji et al., 1996). It has been clearly shown that HRV decreases with age (Tsuji et al., 1996; Fagard, Pardaens and Staessen, 1999; Nunan, Sandercock and Brodie, 2010; Abhishekh et al., 2013). The total 24-hour HRV is reported to be up to 20% higher in young adults than in elderly ones (Umetani et al., 1998; Bonnemeier et al., 2003). The decrease in HRV with age indicates the predominance of the SNS and the withdrawal of the PNS in ANS regulation (Bonnemeier et al., 2003; Abhishekh et al., 2013). With aging, the decrease in HRV is particularly noticeable during sleep as a diminished circadian pattern of the HRV (Bonnemeier et al., 2003).

HRV also varies between genders (Koenig and Thayer, 2016). The resting HR is significantly higher in women than men (Fagard, Pardaens and Staessen, 1999; Bonnemeier et al., 2003; Koenig and Thayer, 2016), and women's HRV, especially in long-term recordings, is lower than men's (Koenig and Thayer, 2016). The total 24-hour HRV has been reported to be up to 15% higher in men than in women (Umetani et al., 1998; Bonnemeier et al., 2003). Even though the total HRV is lower in women, they seem to have a higher proportion of short-term fluctuations in HR than men (Koenig and Thayer, 2016). It has thus been concluded that women have a relative dominance of PNS while men have a relative dominance of SNS in ANS regulation (Koenig and Thayer, 2016). These gender differences in the HRV can be explained both by humoral and mechanical factors (Bonnemeier et al., 2003), as well as hormones and differences in neural control of the heart (Koenig and Thayer, 2016).

Moreover, the interaction between age and gender affects the HR and HRV (Fagard, 2001). The HR declines with aging in women but not in men, while the decline in HRV with aging is steeper for men than for women (Umetani et al., 1998). Thus, the gender differences in the HR and HRV diminish with age (Umetani et al., 1998; Bonnemeier et al., 2003).

Although the HRV is clearly and consistently associated with age and gender, it is highly individual (Umetani et al., 1998; Bonnemeier et al., 2003; Nunan, Sandercock and Brodie, 2010; Koenig and Thayer, 2016). Thus, the HRV may vary greatly between healthy subjects, even if the measurement protocol is controlled and the subject group is homogenous (Nunan, Sandercock and Brodie, 2010).

2.1.4 Associations between clinical conditions and heart rate variability

The clinical importance of HRV has been acknowledged since the 1960s but knowledge and evidence of the associations between ANS regulation, HRV and various diseases is continuously increasing (Malik et al., 1996; Thayer and Sternberg, 2006; Billman et al., 2015). Broadly speaking, a decline in HRV is associated with various clinical conditions while an increase in HRV is associated with health (Thayer and Sternberg, 2006; Thayer, Yamamoto and Brosschot, 2010). This observation is in line with the observation that an increase in the resting HR is an independent marker of all-cause mortality (Zhang, Shen and Qi, 2016). As with HR, HRV is a feasible, non-invasive, inexpensive and easy-to-obtain clinical parameter (Buccelletti et al., 2009; Zhang, Shen and Qi, 2016).

In cardiology, the first practical use of HRV was in the risk assessment of cardiac events after acute myocardial infarction (MI) (Malik et al., 1996). In the 1980s, decreased HRV after acute MI was associated with increased mortality (Kleiger et al., 1987). Decreased HRV after MI was later also shown to be a risk factor for both sudden and non-sudden cardiac deaths and other adverse outcomes, such as heart failures (Buccelletti et al., 2009; Huikuri and Stein, 2012). In the general population, decreased HRV and increased resting HR have both been independently associated with cardiovascular mortality and morbidity (Hillebrand et al., 2013; Zhang, Shen and Qi, 2016). In addition, it has been reported that attenuated HR recovery and lower HRV during recovery from physical exercise are significant predictors for overall mortality in patients with a high risk of cardiovascular events and in patients after acute MI (Nissinen et al., 2003; Pradhapan et al., 2014).

Another pioneering practical use of HRV was in diabetic autonomic neuropathy (DAN), a complication of diabetes mellitus (Malik et al., 1996). In DAN, the HRV is already reduced in the early stages of the disease, and is thus considered to be the earliest indicator of the disease (Vinik et al., 2003; Dimitropoulos, Tahrani and Stevens, 2014). Tests for cardiovascular ANS regulation are recommended as a key part of the guidelines for diagnosing DAN (Spallone et al., 2011). Even without any complications of the disease, diabetic patients have been reported as having an increased minimum 24-hour HR and decreased HRV (Ewing et al., 1983; Kudat et al., 2006; Benichou et al., 2018).

Reduced HRV has also been reported in various psychiatric disorders, such as in acute psychosis (Valkonen-Korhonen et al., 2003), schizophrenia (Bär et al., 2007), anxiety (Chalmers et al., 2014), and depressive disorders (Kemp et al., 2010; Kemp et al., 2014). In depression, the HRV decreases in line with the severity of the depression, and in major depression an increase in the resting HR has been reported (Kemp et al., 2010; Kemp et al., 2014).

2.2 Monitoring and analysis of heart rate variability

2.2.1 Wearable monitoring of heart rate variability

Recent technological advances have made HR and HRV measurements readily available for consumers in the form of wearable sensors i.e. sensors embedded into small items of apparel, such as watches or jewelry (Korhonen, Pärkkä, and van Gils, 2003; Quintana, Alvares and Heathers, 2016). In fact, there is currently a wide variety of wearable health-monitoring devices on the market that can be used for monitoring various health and well-being aspects, such as sleep, PA or emotional stages (Peake, Kerr and Sullivan, 2018). The wearable sensor market has growth exponentially in the past years and is still constantly growing (Hernando et al., 2018).

The wearable devices for HR and HRV monitoring that are currently on the market typically employ ECG or photoplethysmography (PPG) technologies (Georgiou et al., 2018; Singh et al., 2018). ECG-based technologies that employ a chest strap monitoring, e.g. from Polar (Polar Electro, Kempele, Finland), have been on the market for decades and are nowadays widely available for consumers with affordable price (Achten and Jeukendrup, 2003; Parak and Korhonen, 2014). The more recent PPG-based technology is used in some common examples of wearable sensors e.g., AppleWatch (Apple Inc., California, USA) and ŌURA ring (Oura, Oulu, Finland) (Hernando et al., 2018; Peake, Kerr and Sullivan, 2018).

PPG is an optical method in which a light source illuminates the skin and a light detector measures the absorption of the light in the tissue under the skin (Allen, 2007). The changes in the absorbed light are synchronous with the heartbeats, and thus, HR and HRV can be extracted from PPG measurements (Georgiou et al., 2018). A few studies have investigated the reliability of wearables using PPG for HR and HRV recording (Georgiou et al., 2018; Hernando et al., 2018; Parak, 2018). The PPG technology seems to provide reliable HR recording in free-living settings but the accuracy of the measurements may decrease during physical exercise (Weiler et al., 2017; Parak, 2018). The current PPG technology is highly susceptible to motion artefacts, so reliable HRV estimates can only be obtained during rest and light physical exercise conditions (Allen, 2007; Weiler et al., 2017; Georgiou et al., 2018;

Hernando et al., 2018; Parak, 2018). Therefore, ECG-based devices remain the most accurate method for monitoring HR and HRV in free-living settings (Singh et al., 2018).

A fundamental element affecting the reliability of the HR and HRV analysis is the quality and accuracy of the RR-interval data (Malik et al., 1996). In ECG recordings, the RR-intervals are typically extracted based on detection of the QRS complexes (Malik et al., 1996). The jitter in the QRS complex detection can be minimized by using a sufficiently high sampling frequency for the measurement device (Malik et al., 1996). Recommendations for the minimum sampling frequency vary from 250 Hz to 500–1000 Hz but even lower sampling frequencies (≥100 Hz) with appropriate signal reconstruction methods may be adequate (Malik et al., 1996; Quintana, Alvares and Heathers, 2016; Kwon et al., 2018). Nowadays, the commercially available devices have sampling frequencies that are well over the minimum limits, ranging from 1–8 kHz (Quintana, Alvares and Heathers, 2016).

2.2.2 Artefact correction in heart rate variability analysis

Figure 5 shows a flow chart of a typical procedure for conducting HRV analysis on recorded ECG data. Firstly, the RR-interval data is extracted from the ECG signal by detecting the QRS complexes (see section 2.1.2). Ideally, the HRV analysis is performed on the RR-interval data that only contains normal sinus beats. Normal sinus beats are heartbeats triggered by depolarization of the SA node, and the intervals between consecutive normal sinus beats are called NN-intervals (Malik et al., 1996). The non-sinus beats, also known as abnormal beats or artefacts, are excluded from the NN-intervals (Malik et al., 1996; Peltola, 2012; Shaffer, McCraty and Zerr, 2014). In practice, however, the HRV analysis is typically performed on artefact-corrected RR-interval data where the abnormal beats have been corrected using signal processing techniques (Peltola, 2010). Figure 5 gives an overview of the steps between the raw ECG signal and the HRV analysis. This section, 2.2.2, describes the artefact correction of the RR-intervals while the following section, 2.2.3, describes the calculation of the HRV parameters in different domains (Figure 5, bottom row).

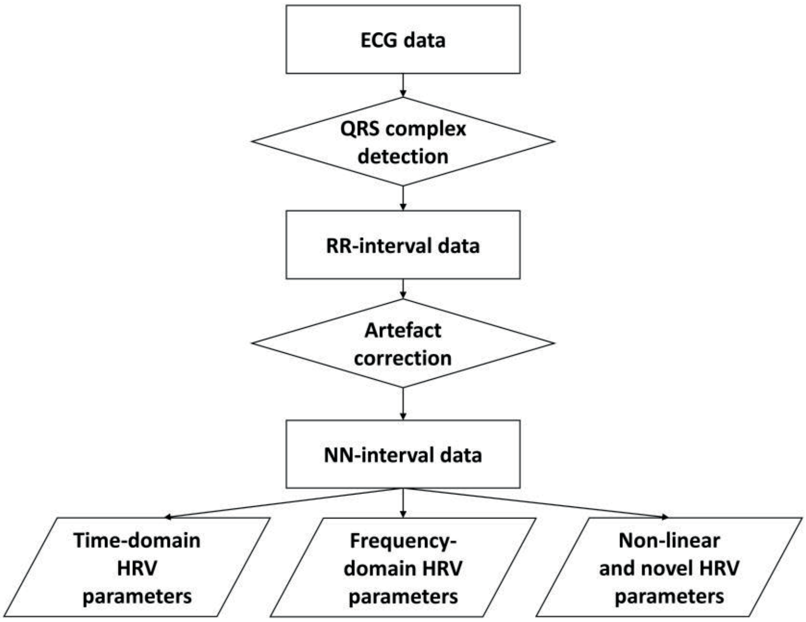


Figure 5. The flow chart of HRV analysis from recorded ECG data to the HRV parameters.

Artefacts are unavoidable for wearable ECG recordings in free-living settings, so they need to be taken into account in order to avoid any distortion of the HRV analysis (Peltola, 2012; Choi and Shin, 2018; Pang and Igasaki, 2018). In ECG, artefacts may originate from technical issues e.g. poorly attached electrodes or sudden movements of the subject, or from physiology e.g. ectopic beats (Peltola, 2012; Choi and Shin, 2018). With the RR-intervals, the artefacts arising from the instrumentation are typically observed from a relatively long sequence of abnormal RR-intervals, while the ectopic beats are typically observed from a compensatory pause after an ectopic beat (Peltola, 2010). Both technical artefacts and ectopic beats are also observed in healthy individuals (Bikkina, Larson and Levy, 1992; Peltola, 2010). For long-term recordings, it is not practical to carry out visual inspections and manual editing in the artefact correction of the RR-intervals (Karlsson et al., 2012; Peltola, 2012). However, automatic artefact correction has been shown to provide sufficiently accurate HRV analysis results, and this is the method that should be used, especially in long-term recordings (Karlsson et al., 2012; Peltola, 2012).

Many of the algorithm-based artefact correction methods employ strategies of deletion or interpolation (Peltola, 2012). In the deletion-based methods, the erroneous RR-intervals are removed, while in the interpolation-based methods, the erroneous RR-intervals are replaced by the RR-intervals interpolated from the normal RR-intervals (Salo, Huikuri and Seppänen, 2001). In the case of interpolation-based methods and ectopic beats, for HRV analysis it is important to edit both the ectopic beat and the following compensatory pause (Peltola, 2012). In general, interpolation methods are superior to deletion methods (Peltola, 2012). However, different artefact-correction methods have different effects on the different HRV measurement estimates, so there is no one gold-standard artefact-correction scheme (Salo, Huikuri and Seppänen, 2001). It is good practice to report the relative number and duration of the artefact-corrected or -deleted RR-intervals in the recordings (Malik et al., 1996). For reliable HRV analysis, those parts of the recordings which have many erroneous RR-intervals should be excluded (Peltola, 2012). A minimum threshold of 80% for the proportion of normal sinus beats in the recording is a typical level in scientific studies (Peltola, 2012).

2.2.3 Heart rate variability analysis

As mentioned, the HRV parameters are calculated from the NN-intervals or from the artefact-corrected RR-intervals (Malik et al., 1996; Shaffer, McCraty and Zerr, 2014). The most widely used HRV parameters are calculated in the time and frequency domains (Malik et al., 1996; Shaffer and Ginsberg, 2017). In addition, some non-linear measures of HRV have been developed and used (Malik et al., 1996; Shaffer and Ginsberg, 2017).

The time-domain HRV measures are based on statistical analysis of the NN-intervals, while the frequency-domain measures are based on the power spectral density (PSD) estimate of the NN-intervals (Billman, 2011). In a non-linear analysis of HRV, the idea is to describe the non-linear dynamics of HR changes by measuring, for example, the complexity of the NN-interval time series with entropy measures (Huikuri et al., 2009).

The most traditional HRV parameters are the time- and frequency-domain parameters (Malik et al., 1996). The time-domain HRV parameters are the statistical parameters of the NN-intervals (Malik et al., 1996; Shaffer, McCraty and Zerr, 2014). Typical time-domain HRV measures widely used and reported for HRV studies include the standard deviation of the NN-intervals (SDNN) and the root-mean square of successive differences in the consecutive NN-intervals (RMSSD) (Malik et al., 1996). Due to the simplicity of the calculations, the time-domain HRV measures can easily be reproduced (Billman, 2011). In addition, the simple time-domain HRV measures, such as SDNN, are more robust to artefacts in RR-intervals and artefact correction methods than the more complex time-domain methods, such as RMSSD, or to the frequency-domain HRV parameters (Peltola, 2010).

The frequency-domain HRV parameters require signal processing methods as the frequency power spectrum is typically estimated from the interpolated equidistant RR-intervals (Malik et al., 1996; Peltola, 2012). Due to the natural variation in the duration of heartbeats, the RR-interval data is spontaneously irregularly sampled (Malik, 1996; Peltola, 2012). To obtained uniformly resampled RR-intervals, the irregular RR-intervals can be first interpolated by using, e.g. spline interpolation methods, and thereafter be resampled with a uniform time intervals (Peltola, 2012). Once the interpolated equidistant RR-interval data is available, PSD estimates can be obtained using either non-parametric methods, e.g. Fast Fourier Transform, or parametric ones, e.g. autoregressive model, methods (Malik et al., 1996; Quintana, Alvares and Heathers, 2016). The advantage of the non-parametric methods are their simplicity and high processing speed, while the non-parametric methods produce smoother PSD estimates and more accurate PSD estimates with low number of samples than the non-parametric methods but the parametric methods require effort to verify the suitability of the selected model (Malik et al., 1996). However, both non-parametric and parametric provide similar results in general and almost equivalent estimates for the high-frequency bands (Hayano et al., 1991; Malik et al., 1996).

The frequency-domain HRV parameters describe the power of the PSD in specific frequency ranges (Malik et al., 1996). In adults, the frequency ranges of interest are: ultra-low frequencies (ULF) of ≤0.003 Hz, very-low frequency (VLF) of 0.003–0.04 Hz, low frequencies (LF) of 0.04–0.15 Hz, high frequencies (HF) of 0.15–0.4 Hz, and the total frequency range of ≤0.4 Hz (Malik et al., 1996; Quintana, Alvares and Heathers, 2016). In other words, the fluctuations in the RR-intervals are over 5 minutes for the ULF band components, 25–300 seconds for the VLF, 7–25 seconds for the LF and less than 7 seconds for the HF components (Quintana, Alvares and Heathers, 2016). It is recommended to report the frequency-domain HRV parameters in absolute units of power (ms^2) (Malik et al., 1996). Additionally, the power in the LF and HF components can be expressed by normalizing them against their total power (Malik et al., 1996; Shaffer and Ginsberg, 2017). The ratio between the LF and HF powers i.e. the LF/HF ratio (Heathers, 2014) is equivalent to the normalized LF and HF powers. LF and HF powers expressed in normalized units, or as an LF/HF ratio may be better for comparing the frequency range powers between the subjects. This is because the HRV may vary greatly between the subjects and due to contextual factors (Shaffer and Ginsberg, 2017).

Recently, a novel analysis for personalized HRV indices has also been studied (Uusitalo et al., 2011). The personalized HRV indices take into account the individual's baseline resting HR and HRV to describe the ANS regulation (Uusitalo et al., 2011). They employ traditional time- and frequency-domain methods (Föhr et al., 2015) to compare the HR and HRV responses to stimuli against the baseline resting HR and HRV levels (Uusitalo et al., 2011). Based on the individual changes in the HR and HRV, the ANS regulation reactions can be categorized into stress, recovery and physical activity, for example (Föhr et al., 2015). Thus, these novel HRV indices provide information about ANS regulation reactions that are independent of the subject's background characteristics, such as age and gender (Uusitalo et al., 2011).

The HRV analysis of the RR-intervals can be done either in short, typically 5-minute, or long, typically 24-hour, time windows (Malik et al., 1996). However, the length of the analysis window affects the HRV parameter values (Malik et al., 1996; Shaffer and Ginsberg, 2017). Thus, short-term and long-term HRV values are not interchangeable and cannot be easily compared (Malik et al., 1996; Shaffer and Ginsberg, 2017). The length of the time window determines the length of cyclic variations and the frequency components in the HRV that can be analyzed. The longer the time window, the lower the frequencies that can be analyzed (Heathers, 2014). On the other hand, the time window also affects the level of HRV. For example, HRV analyzed over a long (24-hour) time window includes responses to a great number of stimuli while HRV analyzed over a short (5-minute) time window represents the response to only a limited number of stimuli (Shaffer and Ginsberg, 2017). In other words, the HRV analysis in long (24-hour) time windows is more likely to violate the stationarity assumption, i.e. mechanisms responsible for specific HR modulations are unchanged during the recording (Malik et al., 1996; Laborde, Mosley and Thayer, 2017). This makes any interpretation of the frequency-domain HRV parameters calculated over long (24-hour) time windows particularly complicated (Malik et al., 1996; Laborde, Mosley and Thayer, 2017).

The current recommendation for long (24-hour) recordings is to analyze the HRV in short (5-minute) non-overlapping time windows (Laborde, Mosley and Thayer, 2017). In a 5-minute time window, the time-domain as well as the VLF, LF and HF frequency components can be studied (Malik et al., 1996). Based on the calculated 5-minute HRV parameters, the lower frequency variations, such as circadian rhythms, can also be indirectly estimated (Laborde, Mosley and Thayer, 2017).

2.2.4 Heart rate variability parameters and their relation to autonomic nervous system regulation

HRV is a complex measure for HR fluctuations introduced by the ANS, as it is for other factors (Shaffer and Ginsberg, 2017). HRV reflects the ANS modulation only indirectly (Billman, 2011). Thus, HRV should be interpreted more as a qualitative, than a quantitative measure of ANS regulation (Billman, 2011). Moreover, the limitations of the HRV measures in distinguishing between the SNS and PNS branches in the ANS regulation should be noted (Billman, 2011) although different HRV measures have been associated with different origins of HR fluctuations (Malik et al., 1996).

The SDNN measure describes the total variability of the NN-intervals in a time window, typically 5 minutes for a short-term interval or 24 hours for a long-term recording (Shaffer and Ginsberg, 2017). SDNN is thus a measure of the total HRV, including the variability generated by both PNS and SNS activation, as well as any other mechanisms that influence the heartbeat (Shaffer, McCraty and Zerr, 2014). In short-term recordings at rest, SDNN is primarily modulated by the RSA, while in long-term recordings, the circadian rhythm of HRV has a major effect on the SDNN (Kleiger, Stein and Bigger, 2005; Shaffer, McCraty and Zerr, 2014). In general, the SDNN values tend to increase with longer analysis time windows as the longer time windows include a wider variety of stimuli as well as the low frequency variations of HRV (Malik et al., 1996; Shaffer and Ginsberg, 2017).

The RMSSD is calculated as the square root of the mean of the squared differences between the successive NN-intervals (Shaffer and Ginsberg, 2017). As evident from the definition, the RMSSD measures the beat-to-beat variation in the NN-intervals and corresponds to the HF parameter in the frequency domain (Malik et al., 1996). RMSSD is a well-established time-domain measure for the PNS-mediated changes in the NN-intervals (Shaffer, McCraty and Zerr, 2014).

Traditionally, the LF peak in the PSD has been taken to describe the SNS modulation, while the HF peak has been taken to describe the PNS modulation (Malliani et al., 1998; Billman, 2013). However, this simple association between LF and SNS has been challenged (Goldstein et al., 2011). Firstly, the SNS has been observed to modulate HR with frequencies $\leq$0.1 Hz, while the PNS modulates the HR at frequencies $\geq$0.05 Hz (McCraty and Shaffer, 2015). Some interventions that are thought to increase SNS activation, such as physical exercise, have been shown not to increase the LF component (Houle and Billman, 1999; Sandercock and Brodie, 2006). In addition, respiration introduces modulation of the HR in the LF range, especially in the case of slow and deep breaths (Brown et al., 1993; Shaffer and Ginsberg, 2017). In short-term resting recordings, the LF is primarily thought to reflect baroreflex activity (Reyes del Paso et al., 2013; Shaffer, McCraty and Zerr, 2014; Goldstein et al., 2011). However, SNS activation may also significantly modulate the LF, in both short-term and long-term recordings (Shaffer, McCraty and Zerr, 2014). In summary, the LF component is a complex measure of HRV. It is made up mostly of the PNS (≥50%), the SNS (≤25%), and substantial contributions from other, as yet unidentified, factors (Billman, 2013).

The HF component has been strongly associated with PNS activation in the majority of studies, although this association is still a matter of debate (Malik et al., 1996; Taylor et al., 2001; Thayer, Yamamoto and Brosschot, 2010; Billman, 2011). The HF range of HRV corresponds to fluctuations in the HR due to the RSA (Shaffer and Ginsberg, 2017). The RSA has traditionally been thought to reflect PNS activity, but evidence shows that SNS activity also modulates the RSA (Taylor et al., 2001; Cohen and Taylor, 2002). Thus, it has been concluded that the HF power is largely due to the PNS modulation (Billman, 2013). The PNS and SNS are estimated to contribute around 90% and 10% to the HF, respectively (Billman, 2013).

Consistent with the traditional interpretation of LF and HF powers, as being markers of SNS and PNS activation, the LF/HF ratio has been interpreted to represent the "sympatho-vagal balance" (Malliani et al., 1998). Along with discussion over how to interpret the LF and HF powers, there has also been much debate about how to interpret the LF/HF ratio (Billman, 2013; Heathers, 2014). From a mathematical perspective, it is self-evident that the LF/HF ratio may have similar values or changes with respect to different LF and HF values and changes (Heathers, 2014). Because both PNS and SNS have been reported as contributing to changes in both the LF and HF components in a complex manner, the physiological basis for the LF/HF ratio cannot be determined in detail (Billman, 2013). Thus, the LF/HF ratio cannot be used to interpret the ANS regulation on its own (Heathers, 2014), so it is recommended to report the LF/HF ratio together with the raw values of the HRV frequency components (Malik et al., 1996).

While the LF and HF components of HRV have gained a lot of attention in the past, the origins for VLF and ULF have been less well studied (Shaffer, McCraty and Zerr, 2014). Nevertheless, the VLF and ULF bands account for 95% of the total power in long-term (24-hour) recordings (Malik et al., 1996). The VLF band reflects the long-term variations in HR and it has been proposed these originate from thermoregulation and hormonal factors, but recent evidence shows that VLF is intrinsically generated by the heart (Shaffer, McCraty and Zerr, 2014). The research has suggested that the ULF band originates primarily from circadian regulation, and from other slow regulatory mechanisms, such as the core body temperature (Shaffer, McCraty and Zerr, 2014).

The time-domain HRV measures have their counterparts in the frequency-domain and vice versa, because of their mathematical relationships and the same physiological phenomena influencing them (Kleiger, Stein and Bigger, 2005). The highly correlated time- and frequency-domain HRV measures are, however, not one-to-one equivalents (Kleiger, Stein and Bigger, 2005; Billman, 2011). The frequency-domain measures provide more detailed information than the time-domain measures (Billman, 2011). The SDNN correlates highly with the total power in the frequency domain, and the total power is the frequency-domain equivalent to the time-domain SDNN (Bigger et al., 1992). Both the RMSSD and the HF mainly reflect the PNS modulation and they are highly correlated (Bigger et al., 1992; Kleiger, Stein and Bigger, 2005). However, the RMSSD is not affected by respiration, unlike the HF (Penttilä et al., 2001; Hill et al., 2009). A summary of the most common time- and frequency-domain HRV parameters is given in Table 2.

Table 2. A summary table of the most common heart rate variability (HRV) parameters: SDNN, RMSSD, LF, HF, and LF/HF ratio.

HRV parameter	Definition	Physiological interpretation	Domain	Unit	Typical range[1]	Correlated parameters
SDNN	the standard deviation of NN-intervals	Reflects the overall ANS regulation in both PNS and SNS regulation	time	ms	30–100 ms	Total power
RMSSD	the root-mean-square of successive differences of RR-intervals	Reflects PNS regulation	time	ms	20–80 ms	HF
LF power	the low-frequency (0.04–0.15 Hz) power	Reflects mainly PNS regulation, but also SNS regulation and other factors.	frequency	ms^2	190–1000 ms^2	
HF power	the high-frequency (0.15–0.4 Hz) power	Affected mainly by RSA: reflects, mainly but not purely, PNS regulation.	frequency	ms^2	30–3600 ms^2	RMSSD
LF/HF ratio	the ratio between the low and high frequency powers	Reflects PNS and SNS regulation in a complex, non-linear manner.	frequency	-	1–12	

[1] = in short-term recordings (Nunan, Sandercock and Brodie, 2010)

2.3 Physical activity, sleep, and other behaviors and contextual factors linked to heart rate variability

In addition to personal background characteristics (Bonnemeier et al., 2003; Nunan, Sandercock and Brodie, 2010) and clinical conditions (Malik et al., 1996), HRV is also associated with various behaviors and contextual factors. These include physical activity (Sandercock and Brodie 2006; Nunan, Sandercock and Brodie, 2010), sleep (Tobaldini et al., 2013), smoking (Barutcu et al., 2005), alcohol use (Sagawa et al., 2011; Quintana et al., 2013) and psychological stress (Föhr et al., 2015). Many of these factors, such as smoking, alcohol dependence and poor sleep have also been associated with an increased risk of cardiovascular diseases and events (Barutcu et al., 2005; Quintana et al., 2013; Tobaldini et al., 2013). Figure 6 summarizes the effects of the most important personal characteristics, behaviors, and clinical conditions affecting HRV. However, it should be noted that HRV has been reported to be highly individual (Nunan, Sandercock and Brodie, 2010).

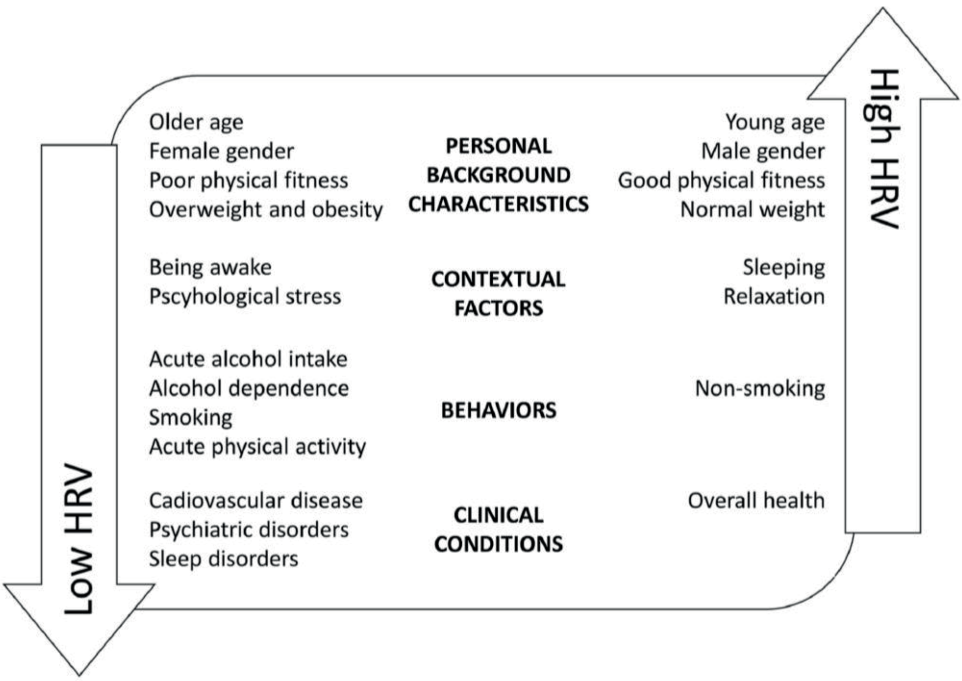


Figure 6. A summary of the effects on the HRV of the subjects' background characteristics, contextual factors, behaviors and clinical conditions.

2.3.1 Associations of physical activity behavior and heart rate variability

Physical activity (PA) is defined as the bodily movement produced by skeletal muscles and increased energy expenditure (EE) (Caspersen, Powell and Christenson, 1985). Physical fitness describes one's ability to perform physical activities (Caspersen, Powell and Christenson, 1985). Cardiorespiratory fitness describes the ability to supply oxygen to working muscles via the cardiovascular and respiratory systems during heavy aerobic physical exercise (Howley, 2001). Cardiorespiratory fitness is typically assessed with maximal oxygen uptake (VO_{2max}), which is the highest oxygen uptake rate that the subject achieves during heavy aerobic physical exercise (Howley, 2001). The HR and HRV parameters associated with all of these concepts are related to PA behavior, which describes the way an individual conducts PA and modifies their physical fitness (modified from Oxford English Dictionary Online, 2019).

It has been proposed that the variety of short-term resting HRV values observed across studies in healthy adults is partly due to the differences in the subjects' habitual physical activity or fitness levels (Nunan, Sandercock and Brodie, 2010). There is evidence of cross-sectional differences between fit and unfit subjects (Sandercock and Brodie, 2006). For example, VO_{2max} has been identified as having a high positive correlation with HRV (Goldsmith et al., 1997). Moreover, a negative correlation has been established between HRV and the body-mass-index (BMI) (Molfino et al., 2009; Koenig et al., 2014). Cardiorespiratory-fitness-enhancing training also decreases resting HR (Wilmore et al., 2001; Yamamoto et al., 2001; Borresen and Lambert, 2008) and submaximal HR (Wilmore et al., 2001; Borresen and Lambert, 2008). Meta-analysis and review articles have concluded that physical exercise training increases HRV (Sandercock, Bromley and Brodie, 2005; Hautala, Kiviniemi and Tulppo, 2009). A number of cross-sectional studies with a limited number of subjects have reported higher short-time resting HRV and lower HR for subjects with higher cardiorespiratory fitness, or those who are engaged in regular physical activity, although these have not always been statistically significant (Melanson, 2000; Rennie et al., 2003; Buchheit et al., 2005; Buchheit and Gindre, 2006; Zaffalon et al., 2018). In fact, a few longitudinal randomized controlled trials in older men with a follow-up time of multiple years have not shown that cardiorespiratory fitness-enhancing training affects the HR or the HRV (Uusitalo et al., 2004; Tuomainen et al., 2005). This may be due to too small an increase in cardiorespiratory fitness (Tuomainen et al., 2005), or simply ageing, which overwhelms the effect of increased cardiorespiratory fitness (Uusitalo et al., 2004). Overall, the effects of training on

HRV are still rather inconclusive (Borresen and Lambert, 2008). This may be explained by the different training programs and the non-optimal HRV indices and analyses used in the studies (Sandercock and Brodie, 2006; Borressen and Lambert, 2008).

During physical exercise, HR increases linearly with the intensity of the PA but the increase depends on the subject's fitness level (Sandercock and Brodie, 2006). The HRV shows a curvilinear decrease as PA intensity increases (Michael, Graham and Davis, 2017). This decrease is highest between rest and moderately-to-vigorously intense PA (Michael, Graham and Davis, 2017). The effect of the PA intensity seems to outweigh the effect of the PA duration or the mode of the HRV (Michael, Graham and Davis, 2017). Compared to a 30-minute moderate PA, a prolonged 90-minute moderate PA further increased the HR and decreased the HRV but the most drastic differences with the resting HR and HRV levels were already observable after only a 30-minute PA (Moreno et al., 2013). Moreover, different moderate-intensity training modalities and similar HR responses have been shown to produce relatively small, although statistically significant, differences in the HRV responses (Leicht, Sinclair and Spinks, 2008).

2.3.2 Assessment of physical activity

The outcome of PA is increased EE and thus, the assessment of PA is based on the estimated EE (Hills, Mokhtar and Byrne, 2014). The gold standard for the estimation of total EE in free-living settings is doubly labelled water (Schoeller and van Santen, 1982; Hills, Mokhtar and Byrne, 2014), a technique to track the elimination of the administered isotopes of hydrogen and oxygen from the body using urine samples (Wareham and Rennie, 1998; Hills, Mokhtar and Byrne, 2014). However, this method cannot provide any detailed information about the EE related to PA, so measurement approaches based on indirect calorimetry are the reference method for estimating EE in laboratory and field studies (Strath et al., 2013; Hills, Mokhtar and Byrne, 2014). In indirect calorimetry, the total energy production of the body is measured (Ainslie, Reilly and Westerterp, 2003). Nowadays, indirect calorimetry methods typically estimate oxygen consumption by measuring the quantity of inhaled gases with a known oxygen concentration (Strath et al., 2013; Hills, Mokhtar and Byrne, 2014). For each liter of oxygen consumed, the body utilizes a certain amount of energy (ca. 5 kilocalories), and this can be used for estimating the EE at rest, during PA, and for other activities (Hills, Mokhtar and Byrne, 2014). The problem with this method is that the equipment needed is bulky and complicated, and therefore not easy to use outside the laboratory (Ainslie, Reilly and Westerterp, 2003). Although there are portable indirect calorimetry devices which can measure EE accurately, they are expensive and can only be used for a limited time (Ainslie, Reilly and Westerterp, 2003).

Due to the limitations of measuring EE with indirect calorimetry, various other methods have been developed, from subjective questionnaires to the monitoring of acceleration and HR (Strath et al., 2013; Hills, Mokhtar and Byrne, 2014). Subjective questionnaires are particularly useful in large population-based studies (Ainslie, Reilly and Westerterp, 2003). In fact, a variety of subjective questionnaires are the most widely used method for assessing PA, but their reliability and validity may be limited (Shephard, 2003; Westerterp, 2009). It has been concluded that although the results of such questionnaires tend to agree with more objective measures for assessing vigorous-intensity PA levels, they are less accurate with low-intensity ones (Strath et al., 2013). Thus, the best way to use subjective questionnaires is for categorizing the subjects between sedentary and active, although they can also be used for estimating physical fitness (Jackson et al., 1990; Westerterp, 2009; Hills, Mokhtar and Byrne, 2014). For example, VO_{2max} can be assessed reliably with self-reported PA behavior together with personal background characteristics of age, gender and BMI (Jackson et al., 1990).

Recent technological advances have popularized PA monitoring with wearable sensors using either accelerometers or HR monitoring (Hills, Mokhtar and Byrne, 2014). Both accelerometers and HR monitoring can provide objective and detailed information about PA, such as its intensity and duration (Hills, Mokhtar and Byrne, 2014). Accelerometers are wearable devices that detect the degree of acceleration caused by the PA (Hills, Mokhtar and Byrne, 2004). Acceleration is a measure of changes in velocity, and in PA it is proportional to the oxygen uptake (VO_2) (Hills, Mokhtar and Byrne, 2014). However, the relationship between acceleration and VO_2 varies between the modes of PA, and the accelerometers require individual calibration in order to provide an accurate assessment of the PA (Brage et al., 2003; Strath et al., 2013; Hills, Mokhtar and Byrne, 2014). For example, various methods estimating VO_2 from the raw acceleration data have been shown to have a high correlation (correlation coefficient of ≥ 0.75) with measured VO_2 but the PA activities have not really been representative of daily activities in free-living settings (de Almeida Mendes et al., 2018).

Using HR to assess PA is based on the well-known fact that HR increases during PA and there is a relationship between HR and VO_2 (Sandercock and Brodie, 2006;

Hills, Mokhtar and Byrne, 2014). The problem with this approach is that the relationship between HR and VO_2 is not linear during rest or low-intensity PA (Ainslie, Reilly and Westerterp, 2003). Moreover, it varies across individuals due to their background characteristics, such as gender, weight, age and physical fitness (Keytel et al., 2005; Hills, Mokhtar and Byrne, 2014). Individual calibration tests in laboratory settings can be used for estimating the relationship between an individual's HR and VO_2, in order to accurately estimate VO_2 (Hills, Mokhtar and Byrne, 2014). On the other hand, the estimation of VO_2 using HR and background characteristics has shown reasonable accuracy even without individual calibration (Rennie et al., 2001; Keytel et al., 2005). A neural network model for estimating VO_2 from the respiration rate, HR, the dynamics of HR and HRV, and the individual's background characteristics has also been proposed (Smolander et al., 2011). The VO_2 estimated with this method was shown to have a high correlation (correlation coefficient of ≥ 0.75) with the measured VO_2 over a number of different free-living daily activities ranging from low- to vigorous-intensity (Smolander et al., 2011; Robertson et al., 2015).

Due to the often complementary information that accelerometers and HR monitoring provide, it has been suggested that a combination of assessment techniques that can provide the most accurate results (Ainslie, Reilly and Westerterp, 2003; Altini et al., 2013; Strath et al., 2013; Hills, Mokhtar and Byrne, 2014). Recently, machine-learning-based algorithms predicting VO_2 from acceleration, HR and respiration patterns have been shown to provide high correlations (correlation coefficients of ≥ 0.87) with the measured VO_2 during typical daily activities (Beltrame et al., 2017; Lu et al., 2018).

Regardless of the method used to estimate VO_2, VO_2 levels during PA can be used to evaluate the intensity of the PA (Howley, 2001). PA intensity can be estimated in *absolute* or by *relative* terms. With *absolute* terms, the intensity of the PA is assessed by the actual VO_2 or EE, while in *relative* terms, the intensity of the PA is assessed by VO_2 uptake or EE *relative* to the subject's fitness level (Howley, 2001). The volume of PA can be estimated as a product of its intensity, duration and frequency (Howley, 2001).

A typical expression for PA intensity in *absolute* terms is metabolic equivalents (METs) (Howley, 2001). This means comparing the VO_2 level during an event with the VO_2 level when sitting quietly, or at rest (Hills, Mokhtar and Byrne, 2014). Typically, the VO_2 level at rest is assumed to be 3.5 ml/kg/min but the BMI, age and gender of the subject affect the resting VO_2 level (Byrne et al., 2005; Kozey et al., 2010). Thus, the MET values may be derived from background-corrected baseline VO_2 levels using, for example, the original Harris-Benedict formula (Harris and Benedict, 1918). With METs, the intensity of PA can be categorized into three types: moderate (3–6 METs), vigorous (≥6 METs), and moderate-to-vigorous (≥3 METs) (Garber et al., 2011).

In terms of *relative* PA intensity, the percentage of oxygen uptake reserve (% VO_2R) can be used (Howley, 2001) and it is defined as:

$$\% \, VO_2R = \frac{VO_2 - VO_{2rest}}{VO_{2max} - VO_{2rest}} \times 100, \tag{1}$$

where VO_2 is the oxygen uptake, VO_{2rest} is the oxygen uptake at rest, and VO_{2max} is the maximum oxygen uptake (Howley, 2001). With % VO_2R, the ranges for moderate, vigorous, and moderate-to-vigorous PA are 40–60% VO_2R, ≥60% VO_2R, and ≥40% VO_2R, respectively (Howley, 2001; Garber et al., 2011). The *relative* terms are more reliable for the assessment of PA intensity than the *absolute* terms, especially, for older and unfit subjects (Howley, 2001; Garber et al., 2011; Piercy et al., 2018).

Today, the recommended volume of PA in adults for substantial health benefits is at least 150–300 minutes per week of moderate-intensity aerobic PA or 75–150 minutes of vigorous-intensity aerobic PA, or other equivalent combinations of moderate and vigorous aerobic PA (Piercy et al., 2018). In the previous guidelines, the recommended volume of PA to improve health and cardiorespiratory fitness in adults was at least 150 minutes per week of moderate-intensity PA or 75 minutes of vigorous-intensity PA, or other combinations of moderate and vigorous PA resulting in a similar total EE (Garber et al., 2011). Moreover, the recommended volume of PA should be accumulated from bouts of activity lasting at least 10 minutes in the previous guidelines but now this requirement has been relaxed (Garber et al., 2011; Piercy et al., 2018). A large proportion of the global adult population does not achieve the recommended PA volumes (Guthold et al., 2008; Colley et al., 2011; Piercy et al., 2018). However, PA volumes lower than the recommended levels can also be beneficial, as there seems to be a dose-dependent relationship between PA and health outcomes (Garber et al., 2011; Piercy et al., 2018). The people who are moving the least will benefit even from a modest increase in PA (Piercy et al., 2018). Moreover, being physically active seems not to be enough, as sedentary behavior seems to be an independent risk factor for health, and thus, nearly everyone benefits from moving more and sitting less throughout the day (Piercy et al., 2018; Patterson et al., 2018; Stamatakis et al., 2019).

Since errors in estimating the PA's intensity, duration and frequency directly affect the estimated volumes of PA, more objective measures for assessing PA are recommended (Howley, 2001; Hills, Mokhtar and Byrne, 2014). Assessing a subject's habitual PA requires a comprehensive set of measures and long or sufficiently frequent observations (Levin et al., 1999; Strath et al., 2013; Hills, Mokhtar and Byrne, 2014). For example, it has been reported that PA levels can vary substantially from day to day and from season to season (Gretebeck and Montoye, 1992; Uitenbroek, 1993; Levin et al., 1999; Matthews et al., 2002; Pivarnik, Reeves and Rafferty, 2003). There are also differences in PA volumes between workdays and days off (Gretebeck and Montoye, 1992; Matthews et al., 2002; Mutikainen et al., 2014). Thus, any assessment of habitual PA should include both work and leisure time (Gretebeck and Montoye, 1992; Matthews et al., 2002; Jaeschke et al., 2018). At least one-week monitoring has been recommended for reliable objective assessment of habitual PA (Matthews et al., 2002; Jaeschke et al., 2018). Leisure-time PA seems to be the most important determinant of subject's PA level, and especially leisure-time PA is susceptible to seasonal changes (Shephard and Aoyagi, 2009). Both cross-sectional and longitudinal studies show that PA volumes seem to be lower in the winter than in the summer months (Uitenbroek, 1993; Levin et al., 1999; Pivarnik, Reeves and Rafferty, 2003; Shephard and Ayogi, 2009).

2.3.3 Sleep and heart rate variability

The circadian rhythm is one of the most distinctive variations observed in HR and HRV during everyday life (Malik et al., 1996; Bonnemeier et al., 2003; Kleiger, Stein and Bigger, 2005). In a normal circadian rhythm, HRV is lower and HR is higher when awake than asleep (Bonnemeier et al., 2003). The HRV starts to increase and the HR to decrease during the evening hours, and the turning point is reached in the early morning hours (Bonnemeier et al., 2003). By late morning, the HRV and HR levels achieve their plateau level, and they remain there until the evening hours (Bonnemeier et al., 2003). The difference in the levels of HR and HRV between being awake and asleep arises primarily from changes in the ANS regulation (Tobaldini et al., 2013). During non-rapid eye movement (NREM) sleep, the PNS, which supports rest and restoration, is predominant (Jones, 2009). Thus, NREM sleep is an optimal period for assessing ANS regulation in its most relaxed state (Brandenberger el al., 2005). Slow-wave sleep that is the deepest stages of NREM sleep is regarded as the most restorative sleep, during which the predominance of the PNS in ANS regulation is the highest (Brandenberger el al., 2005; Jones, 2009). Most slow-wave sleep occurs during the first third of a subject's sleep period, so that is the optimal time for assessing the ANS regulation (Brandenberger el al., 2005; Colten and Altevogt, 2006; Jones, 2009).

Even though the vagal regulation of the ANS continues throughout the sleep, fluctuations occur in the autonomic regulation during normal sleep, mainly in accordance with the sleep stage transitions (Somers et al., 1993; Trinder et al., 2001; Jones, 2009). Typically, sleep starts with the transition from wakefulness to NREM sleep (Colten and Altevogt, 2006). In NREM, the HR decreases and the HRV increases (Somers et al., 1993; Trinder et al., 2001; Bušek et al., 2005). During sleep, the periods of NREM sleep and rapid eye-movement (REM) sleep alternate, but there is more NREM sleep at the beginning of the sleep and more REM sleep at the end (Colten and Altevogt, 2006). In REM sleep, PNS regulation continues, but the SNS activity increases (Jones, 2009). Thus, the transition from NREM to REM sleep is associated with increased HR and decreased HRV (Somers et al., 1993; Trinder et al., 2001; Bušek et al., 2005). During REM sleep, the HR is at a similar level to when the subject is awake (Somers et al., 1993; Bušek et al., 2005). Although ANS regulation has systematic patterns over specific sleep stages, it has also been reported to vary between the sleep stages during the sleep (Trinder et al., 2001; Bušek et al., 2005). For example, the sympathetic modulation is shown to be higher during REM sleep at the end of the sleep, rather than at the start (Scholz et al., 1997). It has been proposed that ANS regulation during sleep varies in relation to the preceding sleep stage and to the length of time, the subject has been asleep (Bušek et al., 2005). On the other hand, the ANS regulation during sleep has not been reported to vary according to time within a sleep stage, which may indicate that the time-of-night effects in ANS regulation arise from the changes in the sleep stage distribution in the course of sleep (Trinder et al., 2001).

Because ANS regulation associates with sleep onset, the various stages of sleep, and pathological sleep, the HRV can be used to assess sleep (Tobaldini et al., 2013) and sleep behaviors of individuals i.e. how individuals typically sleep (modified from Oxford English Dictionary Online, 2019). For example, a systematic review of research studies reported that insomniacs have increased sympathetic activity both during sleep and in the daytime (Nano et al., 2017). Sleep is a crucial element of the body's restoration and recovery process and provides psychophysiological unwinding after exposure to effort (Geurts and Sonnentag, 2006; Lindholm, 2013). Those subjects who self-reported insufficient recovery have been shown to have decreased HRV during the first hours of sleep compared to those subjects who self-reported sufficient recovery (Lindholm et al., 2012). Moreover, subjective stress has been associated with increased HR and decreased HRV during sleep (Föhr et al., 2015).

Other behaviors have been reported to affect ANS regulation during sleep, such as alcohol intake and PA (Hynynen et al., 2010; Romanowicz et al., 2011; Porkka-Heiskanen, Zitting and Wigren, 2013). Their effects can be seen through the changes in HR and HRV (Porkka-Heiskanen, Zitting and Wigren, 2013). Acute alcohol intake before sleep appears to dose-dependently increase HR and decrease HRV, indicating the sympathetic predominance of ANS during sleep after alcohol intake (Sagawa et al., 2011). Moreover, alcohol-dependent subjects have been reported to show increased sympathetic activity during sleep than controls, and similar results have been obtained from awake subjects (Irwin et al., 2006; Spaak et al., 2010; Quintana et al., 2013). The results of these studies are consistent, but the studies have included only a limited number of subjects and have been performed in highly controlled research settings (Irwin et al., 2006; Spaak et al., 2010; Romanowicz et al., 2011; Sagawa et al., 2011; Quintana et al., 2013; Ralevski, Petrakis and Altemus, 2018).

Regarding PA, decreased HR and increased HRV in 24-hour measurements has been associated with higher VO_{2max} (Goldsmith et al., 1997). Thus, it can be hypothesized that physical fitness or exercise training can be an important determinant for HR and HRV during sleep. PA during the day has been shown to significantly alter HR and HRV during the next night's sleep (Hynynen et al., 2010; Myllymäki et al., 2011; Myllymäki et al., 2012).

2.3.4 Other health behaviors and contextual factors impacting heart rate variability

HRV analysis has benefited from sustained interest in psychological and behavioral studies (Quintana and Heathers, 2014). For example, the effects of smoking (Barutcu et al., 2005) and psychological stress and recovery (Föhr et al., 2015) on HRV have been studied. Other health behavior-related factors, such as the consumption of caffeine or alcohol, also have an effect on HRV measurements, although very few studies report or take this into account (Nunan, Sandercock and Brodie, 2010).

Smoking is a common health behavior associated with reduced HRV (Barutcu et al., 2005). HRV decreases and HR increases after acute cigarette smoking in non-smokers (Karakaya et al., 2007). Habitual smokers have decreased HRV compared to non-smokers but if they give up smoking, their HRV increases and their HR decreases (Yotsukura et al., 1998; Minami, Ishimitsu and Matsuoka, 1999; Barutcu et al., 2005).

In stress response, various changes occur in physiological systems, such as the predominance of the SNS in the ANS regulation (Cullinan, 2009). Both physiological stress, such as exercise, as well as pathological stress, such as sepsis infection, have been reported to be associated with decreased HRV, although the decrease in HRV was greater for exercise than for sepsis (Bravi et al., 2013). The effect of psychological stress on HRV has been studied by simulating acute stressors, such as with mental arithmetic tests, and obtaining the perceived stress levels through questionnaires (Berntson and Cacioppo, 2007). Various studies have shown alterations in the ANS regulation with self-reported psychological stress (Lucini et al., 2005; Berntson and Cacioppo, 2007; Jarczok et al., 2013; Föhr et al., 2015). Both acute and chronic psychological stress have been associated with increased HR and decreased HRV (Lucini et al., 2005; Berntson and Cacioppo, 2007). Acute psychophysiological stress has also been associated with decreased PNS modulation and increased SNS modulation during NREM sleep, which may be a pathway for disturbed sleep (Hall et al., 2004). As a subcategory of psychological stress, work-related stress and effort at work have also been shown to be associated with reduced HRV (Thayer, Yamamoto and Brosschot, 2010; Uusitalo et al., 2011).

2.4 Use of real-world observational health monitoring data

Traditionally, all good research studies focus on a specific research question with a carefully designed research protocol (Sherman et al., 2016). In health research, randomized controlled trials (RCTs) are the most reliable method for studying causality between a treatment and an outcome (Sibbald and Roland, 1998). To achieve reliable results about the causal interference, RCTs typically study a limited number of homogenous subjects in controlled settings (West et al., 2008; Sherman et al., 2016). The study subjects are randomly assigned to the treatment groups, and different measures are taken in RCTs to control variability and to ensure the quality of the collected data (Sibbald and Roland, 1998; Sherman et al., 2016). All these practices are intended to ensure that no confounding factors affect the results, so RCTs have good internal validity (West et al., 2008; Rothwell, 2005; Sherman et al., 2016).

However, RCTs have been criticized for their poor external validity, as the results of an RCT may not be generalizable to larger and more heterogeneous patient populations (Rothwell 2005; Berger et al., 2009; Sherman et al., 2016). The results of an RCT can only be generalized to the population from which the RCT subjects are taken. As RCTs are typically small and highly selective, the very small proportion of patients who might participate in the study may differ significantly from the broader patient population as a whole (Rothwell, 2005; Sherman et al., 2016; Kim, Lee and Kim, 2018; Maissenhaelter, Woolmore and Schlag, 2018). In addition to the selection bias, the settings and protocol of RCTs typically differ from routine clinical practice, and this also affects the external validity of the RCT (Rothwell, 2005). Thus, the knowledge gained from RCTs is valuable, but it is incomplete, and so traditional research studies should be complemented with real-world evidence (RWE) gained by analyzing real-world data (RWD) (Booth and Tannock, 2014; Sherman et al., 2016; Berger et al., 2017).

2.4.1 Characteristics and challenges of real-world health data

As with wearable devices, RWD is still a relatively new but rapidly growing research area; a PubMed search with the key term "real-world data" yielded 293 publications in 2008 but 2,855 publications for 2018. However, there is currently no globally accepted exact definition for RWD (Makady et al., 2017). In the health domain, RWD is generally considered to mean the health data gathered outside of typical research studies (Sherman et al., 2016; Berger et al., 2017). Figure 7 comprises the typical sources and characteristics of RWD as well as the concerns with, and advantages of, RWE.

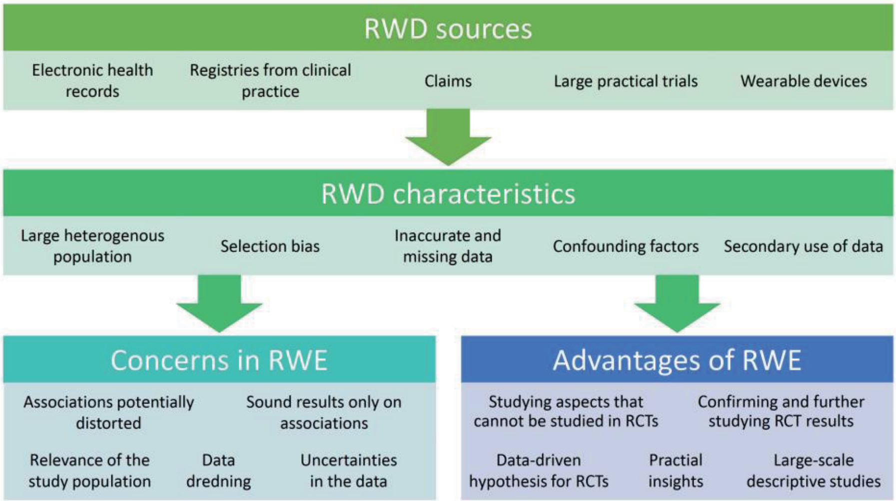


Figure 7. Typical sources and characteristics of real-world data (RWD) as well as typical concerns and advantages related to real-world evidence (RWE).

RWD may originate from various sources. Some typical RWD sources include observational, uncontrolled or non-experimental studies (Makady et al., 2017), registries from routine clinical, health and wellness practices (Cox et al., 2009; Dreyer and Garner, 2009; Mutikainen et al., 2014), interventional studies conducted in real-world settings with heterogeneous population (Sherman et al., 2016), and large practical trials (Garrison et al., 2007). Thus, RWD may have been collected originally for either research or non-research purposes, and the RWD studies may be conducted prospectively and retrospectively (Garrison et al., 2007; Cox et al., 2009; Sherman et al., 2016; Berger et al., 2017). Moreover, the secondary use of data is common in RWD analysis, which means that the RWD is often analyzed for another purpose than that for which it was originally collected (Berger et al., 2009). Based on the literature, the most common sources of RWD are electronic health records and registries, and the resulting RWE is particularly useful for decision making in healthcare (Cox et al., 2009; Berger et al., 2017; Makady et al., 2017). However, in addition to these data sources, RWD can now be gathered from wearable devices and the social media used by the consumers in their day-to-day lives, and this can be used to provide valuable, large-scale RWE (Sherman et al., 2016; Swift et al., 2018).

Despite the great variety of RWD, some typical characteristics for health-related RWD exist (Figure 7). These include large and heterogeneous populations (Motheral et al., 2003; Booth and Tannock, 2014; Sherman et al., 2016), and the presence of a variety of confounding factors (Garrison et al., 2007; Cox et al., 2009; Dreyner and

Garner, 2009; Johnson et al., 2009; Sherman et al., 2016). It has been shown that RWD often includes some wrong or inaccurate data, and there are uncertainties about the data collection procedures, and other factors which affect its quality (Cox et al., 2009; Sherman et al., 2016). Missing data is another typical drawback with RWD (Garrison et al., 2007). Some data variables may be missing for some subjects, or some important variable may be completely missing from the data (Garrison et al., 2007; Johnson et al., 2009). Furthermore, RWD not originally collected for research purposes is typically characterized by systematic and random errors in the data, lack of precision in the data collection and varying degrees of accuracy in the data variables (Cox et al., 2009).

Despite all these drawbacks, RWD is still a valuable source of information, although it is challenging to utilize it. Typically, the first challenge in using RWD sources is the availability and privacy of the data (Sherman et al., 2016). There are also concerns related to the RWD analysis, and these arise mainly from the typical characteristics of the RWD and the possibility of "data dredging" (Smith and Ebrahim, 2002; Berger et al., 2017). This "data dredging" (or "data fishing") means conducting multiple analyses of the same dataset to gain a desired result (Berger et al., 2017). This is of particular concern with RWD studies due to their large datasets (Smith and Ebrahim, 2002; Berger et al., 2017). RWD also often has a selection bias because there is typically no randomization of the individuals who have been selected for, or participated in, the study (Smith and Ebrahim, 2002; Vandenbroucke et al., 2007; Berger et al., 2017). Thus, the study population and its relevance are potential concerns in RWD studies (Berger et al., 2009). Moreover, uncertainties in the data quality (Sherman et al., 2016), and the specificity of the measures may raise concerns (Berger et al., 2009). In RWD studies, the studied associations between data variables may be distorted by confounding factors (Cox et al., 2009). The confounding factors are some measured or unmeasured variables associated with the data variables which can distort the study of any perceived associations between the data variables (Onghena and van der Noortgate, 2005; Cox et al., 2009).

Currently, RWD is widely used for descriptive studies of large-scale populations (Berger et al., 2017). A number of recommendations for good practice when conducting and evaluating RWD studies have been proposed, which can enhance the confidence in RWE, (Motheral et al., 2003; Berger et al., 2009; Cox et al., 2009; Johnson et al., 2009; Berger et al., 2017). Most of the concerns related to RWE can be addressed with appropriate methodologies, but still the results of RWD studies are only reliable for establishing associations, rather than causations (Berger et al., 2017). In other words, RWD studies can only provide hypotheses about causal relationships (Berger et al., 2017), and these then need to be confirmed or disproved using more traditional and controlled research studies (Sherman et al., 2016). Thus, the real strength of RWD is that it can be used for exploratory analysis and data mining to generate hypotheses which can be tested with traditional research studies (Berger et al., 2009). Nevertheless, observational RWD studies can also be used to confirm the results of an RCT among a larger population, and for further exploratory study of a topic (Booth and Tannock, 2014). RWD studies complement RCTs by enabling the study of issues that would be too costly, impractical or even unethical to be studied with RCTs alone (Rothwell, 2005; Sherman et al., 2016; Kim, Lee and Kim, 2018).

2.4.2 Methodological aspects of real-world data analysis

The methodological aspects of RWD analysis touch on all the key issues in any research study, namely, the research question and the datasets used as well as the statistical study designs and their analysis (Berger et al., 2009). All of these aspects are interconnected and methodologically they should be considered as a complete entity (Vandenbroucke et al., 2007; Berger et al., 2009). For example, the primary concerns of observational RWD analysis results are related to confounding and selection biases (Berger et al., 2009). These issues can be addressed in the design and implementation of the study (Vandenbroucke et al., 2007; Berger et al., 2009) as well as in the statistical analysis (Johnson et al., 2009).

The basis for an RWD study, as for any other research study, is a relevant, rational, specific and novel research question (Berger et al., 2009). However, the research question in exploratory or descriptive RWD studies is not usually formulated as a study hypothesis, unlike in typical RCTs (Berger et al., 2009). The study should be carefully designed to answer the research question in a feasible way (Berger et al., 2009). In RWD studies, typical study designs include cross-sectional, cohort and case-control designs (Vandenbroucke et al., 2007; Berger et al., 2009). Cross-sectional study designs, which include data from subjects at a specific time point, can be used for studying associations between the variables in a dataset (Berger et al., 2009). In addition, such studies are good for generating hypotheses about associations which can then be tested by further studies (Berger et al., 2009). In cohort studies, groups of subjects are compared for a particular outcome while the other variables are made as similar as possible (Berger et al., 2009). It is the same principle for individuals in case-control studies (Berger et al., 2009). There is another popular approach, known as case-crossover or within-subject study design (Berger et al., 2009). This can be used for making comparisons of the variable of interest within the subject at different time points with respect to a specific condition (Berger et al., 2009). These within-subject designs are suitable for studying the acute effect of a condition on the variable of interest, but they typically severely limit the number of subjects for the study, as each subject needs to be measured for all conditions of interest (Berger et al., 2009).

The RWD source must be selected or collected, preprocessed and analyzed to answer the research question as adequately as possible (Berger et al., 2009). In the selection of an RWD source for retrospective and secondary analysis, special attention should be paid to evaluating the relevance of the study population, the amount, richness and quality of the data, and the timeframe of the observations (Berger et al., 2009). The description of the selected dataset and its original purpose, as well as all the preprocessing and analysis steps should be reported in detail to facilitate the understanding of the applicability and possible limitations of the study results (Vandenbroucke et al., 2007).

The selection bias affecting the internal validity of the study results is a concern, especially in any retrospective analysis of observational RWD (Vandenbroucke et al., 2007; Berger et al., 2009). It is good practice to replicate the analysis with a different database and/or population (Berger et al., 2017), although this does not guarantee the generalizability or robustness of the results (Madigan et al., 2013). Restricting the study population can increase the internal validity of the study (Vandenbroucke et al., 2007; Cox et al., 2009). The researcher can apply inclusion and/or exclusion criteria to the study subjects, and this may help to ensure a homogeneous study population (Vandenbroucke et al., 2007; Cox et al., 2009). However, the tight restriction of the study population may also reduce the generalizability of the RWD study results (Cox et al., 2009).

Confounding factors are another major concern in observational RWD studies (Berger et al., 2009). The selection and inclusion of all possible confounding factors requires a thorough literature review (Johnson et al., 2009). These confounding factors may then be accounted for in the design and implementation of the study by a thorough data collection (Vandenbroucke et al., 2007; Cox et al., 2009). Confounding factors may also be reduced through the application of well-defined eligibility criteria (Cox et al., 2009) and rigorous statistical approaches to the data analysis (Johnson et al., 2009). Rich datasets can be used to extract a variety of potential confounding factors (Cox et al., 2009). Because the omission of any confounding factor in the data analysis can cause a bias in the results, a proxy for the confounding factor should be used if the true values of the confounding factors are missing (Johnson et al., 2009). If neither true values nor proxies are available, the missing confounding factors should be acknowledged and written up as a limitation to the study, and any assumed effects on the results should be discussed (Johnson et al., 2009).

Missing data in an RWD dataset is another cause for concern (Berger et al., 2017). Although a well-designed and implemented study may reduce the likelihood of missing data, it is still a common characteristic of observational RWD (Vandenbroucke et al., 2007). Especially in RWD collected originally for informal purposes e.g. wearable monitoring, missing data is typically unavoidable as the data collection still requires manual input and is perfomed voluntarily without any ensurance for the adherence to the data collection (Swan, 2012; Sherman et al., 2016). It is good practice to report the reasons for any missing data, and the amount of missing data for each variable of interest (Vandenbroucke et al., 2007). However, important is also to assess the potential underlying mechanisms responsible for the missing data and to remember that in some cases missing data itself can also be informative (Helander et al., 2014; Salgado et al., 2016).

In the data analysis, missing data points can be discarded or imputed (Vandenbroucke et al., 2007; Johnson et al., 2009). Discarding missing data and only using the "complete data" reduces the amount of data available for the analysis and may cause a bias in the results, especially if the missing data is not missing at random (Vandenbroucke et al., 2007). Alternatively, imputation methods can be used to account for missing data (Johnson et al., 2009). For example, missing data points can be replaced by a mean value, or a value predicted from regression analysis, or by any more sophisticated method that will preserve the variability of the data (Johnson et al., 2009).

2.4.3 Exploratory data analysis and data mining for real-world data

As stated above, RWD is often used for exploratory analysis and data mining to generate a data-driven hypothesis for further studies (Berger et al., 2009). Exploratory data analysis is a statistical approach employing a variety of techniques in order to find patterns in the data, with no predefined association of interest (Hinterberger, 2009). Data mining can be seen as "the process of discovering knowledge or patterns from massive amounts of data" (Han, 2009). Both data mining and exploratory data analysis employ a variety of techniques for data visualization and statistics (Han, 2009; Hinterberger, 2009), and data mining further employs artificial intelligence techniques (Han, 2009). Data mining and exploratory data analysis can also be seen as complementary processes: exploratory analysis provides a basic understanding of the data, while data mining is more oriented towards applications in the analysis of large datasets (Han, 2009; Hinterberger, 2009). In addition to traditional statistical approaches, such as regression and statistical tests, data mining employs advanced statistical learning methods, such as random forests (RFs) and neural networks (Han, 2009). Given the variety of statistical learning methods, data mining typically aims for simple-to-construct, robust, accurate, and easy-to-interpret models providing qualitative information about the relationships of variables (Hastie, Tibshirani and Friedman, 2009). These requirements are best served with an ensemble of decision tree-based methods, such as RFs, and have become one of the most popular statistical learning methods in data mining (Hastie, Tibshirani and Friedman, 2009).

In observational studies, a fundamental exploratory data analysis method is stratified analysis in which statistical analyses are conducted on subcategories of the data (Johnson et al., 2009). To stratify the data, typical variables include any confounding factors and covariates, which are variables having an effect on the outcome but are not of primary interest (Motheral et al., 2003; Johnson et al., 2009). Relevant descriptive statistics including simple statistical measures such as the mean, the median and the standard deviation are calculated for the subcategories of the data (Vandenbroucke et al., 2007). Stratified analysis is a good first step in RWD analysis as it gives an overview of the data and may provide important information about the covariates and their effects on the studied outcome (Johnson et al., 2009). Stratified analysis can also be used as a simple method for controlling confounding factors with minimal statistical assumptions (Greenberg and Kleinbaum, 1985). However, stratified analysis becomes cumbersome and inaccurate when there are multiple covariates and confounding factors (Greenberg and Kleinbaum, 1985; Johnson et al., 2009). Moreover, categorization of the variable values is not ideal for continuous variables as information will be lost and accuracy will be decreased by the categorization (Greenberg and Kleinbaum, 1985; Johnson et al., 2009).

Multivariate regression analysis can be used for studying the association of interest by simultaneously adjusting all the available covariates and confounding factors (Johnson et al., 2009). Multivariate regression is a powerful analysis technique which is typically used to control or adjust any time-independent confounding factors (Johnson et al., 2009). There are various regression models that can be used, depending on the characteristics of data (Johnson et al., 2009). The most typical regression model used for continuous outcomes is ordinary least squares regression (Johnson et al., 2009). This assumes the observations in the data to be independent, the association between the inputs and outcome to be linear and the outcome to be normally distributed (Johnson et al., 2009). Logistic regression models are typically used for binary outcomes, (Johnson et al., 2009). Generalized linear models can be used for data in which the observations are not independent (Johnson et al., 2009), and the Tobit regression model can be used in the case of limited dependent variables (Austin, Escobar and Kopec, 2000). In the case of non-normally distributed outcomes, data transforms, such as the Box-Cox transformation, can be employed (Osborne, 2010).

In order to make any regression analysis fully transparent, the full regression model with all its covariates and confounders and the coefficient of determination (R^2) should always be reported (Johnson et al., 2009). The R^2 shows how well the model explains the data (Johnson et al., 2009). However, it should be noted that in the case of large datasets the R^2 may be very small, even though the regression estimates are not biased (Johnson et al., 2009). In statistical software packages, regression models are built using maximum likelihood estimation, and the fulfillment of the regression model assumptions, including normality, linearity, homogeneity of variance and the absence of multi-collinearity, all of which can easily be ensured with the provided regression diagnostics (Johnson et al., 2009).

If there is a non-linear association between an input and outcome variable, a categorization of the input variable may be an easy solution for regression analysis (Johnson et al., 2009). On the other hand, data mining methods, such as RFs, can be applied to the data to study complex and non-linear associations (Hastie, Tibshirani and Friedman, 2009). RF is a method that uses an ensemble of regression or classification trees built on bootstrapped samples of data (Hastie, Tibshirani and Friedman, 2009). These RFs are fast to construct, can handle mixtures of continuous and categorical data with missing values, tolerate predictor outliers, provide an internal feature selection through variable importance metrics, and have good accuracy in prediction (Hastie, Tibshirani and Friedman, 2009).

The purpose of the statistical analysis is to investigate statistical inference by using the concept of the null hypothesis (Lin, Lucas and Shmueli, 2013). The null hypothesis typically presents a situation where there is no association and the parameter of interest is typically set to zero, and the association between the study variables is concluded only if the distance between the null hypothesis and the data-based estimate is large enough (Vidakovic, 2011; Lin, Lucas and Shmueli, 2013). The distance measure between the data and the null hypothesis is called a p-value and in statistical analysis, the level for statistical significance is typically set to be 0.05 (Lin, Lucas and Shmueli, 2013). The level of statistical significance sets the upper threshold for the p-values that are interpreted to show a statistically significant difference between the data and the null hypothesis (Lin, Lucas and Shmueli, 2013). However, the p-value is typically measured through standard deviations of the estimate i.e. standard errors, and the standard errors of consistent estimators decrease with increasing sample size (Lin, Lucas and Shmueli, 2013). Thus, seemingly negligible distances between the estimate and a null hypothesis become statistically significant (p-value <0.05) with large sample sizes (Lin, Lucas and Shmueli, 2013). In other words, the statistical significance or p-value is dependent on the sample size (Sullivan and Feinn, 2012). Consequently, the statistical interferences in RWD and large datasets should not be interpreted only by the p-values and the estimate sign, but also by the effect size (Lin, Lucas and Shmueli, 2013). The effect size describes the magnitude of the predictors' estimates on the outcome (Lin, Lucas and Shmueli, 2013), and it is independent of the sample size (Sullivan and Feinn, 2012). To enhance the transparency of the results of statistical interferences in large samples, the studies should report the estimates' effect sizes and their confidences intervals, and preferably, also calculate them for various sample sizes to show that the statistical interference is consistent with smaller sample sizes (Lin, Lucas and Shmueli, 2013). Furthermore, any interpretation of the effect sizes should be illustrated with practical examples and marginal analysis by showing the outcome variable values after alternating one predictor variable and holding all the others constant, especially for non-linear models (Lin, Lucas and Shmueli, 2013).

2.4.4 Present and future perspectives for real-world wearable health monitoring data analysis

Currently, the importance of RWE as a complement to traditional clinical trials is widely recognized in healthcare decision-making (Garrison et al., 2007; Khosla et al., 2018). Ten years ago, under the Task Force initiative, a number of researchers (Berger et al., 2009; Cox et al., 2009; Johnson et al., 2009) published reports providing guidance about good practice for the retrospective analysis of RWD. These reports were particularly focused on studies of drug treatment effects using secondary data sources. Although RWE is not yet overly used for making healthcare guidelines, there does seem to be an increasing trend towards the uptake of RWE in healthcare decision-making (Oyinlola, Campbell and Kousoulis, 2016). The use of RWE is also recommended in the drug and medical device regulation practices (Khosla et al., 2018; Swift et al., 2018).

In addition to RWD sources from clinical practice, such as digitalized health records and claims, nowadays wearable sensors can also be used as a source of RWD (Sherman et al., 2016; Swift et al., 2018). These wearable devices include, for instance, chest straps, wristbands and smartwatches (Majumder, Mondal and Deen, 2017). Another resource is smartphone applications, which can be used for monitoring health parameters in free-living settings without interrupting the daily activities of the individuals (Majumder, Mondal and Deen, 2017). Recent technological advances in compact, low-power electronics have brought low-cost, unobtrusive and non-invasive wearable health-monitoring devices onto the market that can be used for monitoring a wide variety of health and well-being aspects, such as sleep, PA or emotional stages (Majumder, Mondal and Deen, 2017; Peake, Kerr and Sullivan, 2018). These wearable consumer devices have become increasingly common, especially among health-conscious individuals who are interested in their own well-being, as they can use the wearables for self-monitoring (Swan, 2012; Piwek et al., 2016). However, although self-monitoring may be part of a lifestyle for some people, the take-up and continued engagement among the wider public is still quite low and (Swan, 2012; Piwek et al., 2016; Marin et al., 2019). Therefore, RWD gathered from these devices typically shows a selection bias among the users (Piwek et al., 2016). It has been suggested that people who practice self-monitoring are usually younger, wealthier, better educated, more motivated towards a healthy lifestyle, and more interested in quantifying their progress than ordinary people (Helander, Wansink and Chieh, 2016; Piwek et al., 2016; Sperrin et al., 2016).

Although also research interest in wearable devices has expanded during the last few years, there has been very little external scientific validation of their accuracy and efficiency, or of the algorithms they use (Piwek et al., 2016; Peake, Kerr and Sullivan, 2018; Silfee et al., 2018). Clearly, the majority of the wearable sensors available in the market for consumers has not been independently validated or used in research (Peake, Kerr and Sullivan, 2018). For example, in PA monitoring some devices show reliable and valid results, but the accuracy of the devices on the market varies greatly and the error in the estimated PA may be up to 25% (Case et al., 2015; An et al., 2017). In addition, a consumer's tendency to use, and adhere to wearable devices also requires more studying (Sieverink, Kelders and van Gemert-Pijnen, 2017; Marin et al., 2019).

Despite the number of studies employing wearable sensors has increased, the real-life studies with wearable sensors are typically conducted on only a limited number of subjects (Silfee et al., 2018). Data mining studies using wearable sensor data have focused on exploring associations in the wearable sensor data recorded from only some tens of subjects (Sathyanarayana et al., 2016; Williams and Cook, 2017), although wearable sensors can be used as means to collect large amounts of data (Peake, Kerr and Sullivan, 2018; Swift et al., 2018). So far only a few studies have been published employing RWD from wearable sensors used by thousands to tens of thousands of subjects (van Dyck et al., 2015; de Lima et al., 2017; Lee et al., 2018).

Additionally, there are a few studies making secondary use of RWD from the wearable sensors of a large number of subjects (Helander, Wansink and Chieh, 2016; Sperrin et al., 2016; Althoff et al., 2017). On the whole, the RWD provided for secondary use in academic research comes from the manufacturers of the devices (Helander, Wansink and Chieh, 2016; Sperrin et al., 2016; Althoff et al., 2017). The use of such RWD has mostly focused on descriptive and exploratory analyses (Helander, Wansink and Chieh, 2016; Sperrin et al., 2016; Althoff et al., 2017). These descriptive and exploratory analyses typically illustrate a number of variables, such as the demographics of the study population, the outcome measure, observed PA, changes in weight or sleep patterns, the most important covariates and confounding factors, and the associations between these (Sperrin et al., 2016; Althoff et al., 2017; Fagherazzi et al., 2017). These studies have already thrown up some interesting associations. For example, in large-scale RWD studies employing thousands of subjects, the increased frequency of self-weighing has been associated with decreased weight (Sperrin et al., 2016), and holidays are associated with weight gain (Helander, Wansink and Chieh, 2016). Although it is acknowledged that the individuals involved in self-monitoring RWD studies do not represent the general population (Sperrin et al., 2016), their data may still give valuable insights into good public healthcare practice (Helander, Wansink and Chieh, 2016). Hopefully, the recently published framework of best practices for analyzing large-scale wearable sensor data that also highlights the opportunities and insights we can gain from such data, inspires both the wearable sensor companies as well as researchers to take the advantage of wearable sensor data to improve public health (Hicks et al., 2019).

Currently, consumer wearables typically provide only simple descriptive statistics from the monitored data but given the rapid development of artificial intelligence for the feedback systems, wearable devices are also becoming more personalized (Piwek et al., 2016; Sawka and Friedl, 2018). More personalized and comprehensive feedback can be also achieved by combining the data gathered from multiple wearable sensors with clinical health data (Gay and Leijdekkers, 2015; Wright et al., 2017). The ultimate, if rather hypothetical, aim for the usage of wearable devices is to provide continuous real-time assessment and prediction of the subject's health status based on sophisticated algorithms and data (Sawka and Friedl, 2018). To achieve the full potential of wearable devices in both research and practice, the wearable devices and their algorithms need to be designed in multi-disciplinary teams including clinicians, data scientists, engineers, behavioral scientists, and interface designers (Patel, Asch and Volpp, 2015; Piwek et al., 2016; Wright et al., 2017; Sawka and Friedl, 2018).

3 AIMS OF THE STUDY

The aim of this thesis work was to apply RWD analysis methodologies for a large-scale real-world HRV dataset to quantify PA and the sleep behaviors in Finnish employees. The specific objectives of the thesis are:

1) from the perspective of real-world data analysis:
    a. to apply methodologies applicable for the analysis of real-world health monitoring data characterized by uncontrolled real-life monitoring settings, unbalanced sampling, and unknown confounding factors (Publications I, III, IV and V)
    b. to evaluate the feasibility of a real-world health monitoring data analysis being able to provide results consistent with previous research, including traditional scientific research studies conducted in controlled settings with a limited number of subjects (Publications III, IV, V)
    c. to provide data-driven hypotheses to be further studied in traditional controlled research studies (Publications II, IV)
2) from the perspective of assessing health behaviors and their associations with physiological functions:
    a. to quantify physical activity bouts from the estimated continuous VO_2 based on the RR-interval recordings (Publications I, II, III, V)
    b. to demonstrate the observations for temporal (month and weekday) patterns of physical activity behavior (Publication I)
    c. to show the impact of PA quantification methods on the estimated PA levels in BMI and age groups, by gender (Publication II)
    d. to assess the association of individual lifestyles and daily activities with ANS regulation during sleep (Publication III)
    e. to quantify the association between acute alcohol intake as well as its interactions with subject's background characteristics, especially age and physical fitness, and ANS regulation during sleep (Publication IV)
    f. to estimate the association between physical activity behavior and the ANS regulation during sleep (Publications III, V)

4 MATERIALS AND METHODS

4.1 Description of a real-world heart rate variability monitoring dataset

This thesis presents research based on the secondary use of a database, gathered originally by Firstbeat Technologies Oy (Jyväskylä, Finland) for their own operational purposes. Firstbeat Technologies Oy is a Finnish company providing and developing analytics for beat-to-beat RR-interval recordings to extract HR and HRV parameters for the subsequent analysis of stress, recovery, physical activity and sleep (Kettunen and Saalasti, 2005b; Firstbeat Technologies Oy, 2014). The analysis algorithms for the RR-interval data developed by Firstbeat Technologies Oy are proprietary and patented (Kettunen and Saalasti, 2005a; Kettunen and Saalasti, 2005b).

The data in the database originates from participants who have performed beat-to-beat RR-interval recordings with a wearable ECG-based device (Firstbeat Bodyguard or Firstbeat Bodyguard 2, Firstbeat Technologies Oy, Jyväskylä, Finland) during their daily life, typically continuously for three days in succession (two workdays and one day off from work). Besides filling in a pre-test questionnaire, the participants are also asked to keep a journal about their work and sleep times, their alcohol consumption and other significant events during the recording period (e.g. exercising, stressful events). Primarily, the database includes recordings from employees who have voluntarily participated in the measurement as a part of a preventive health and wellness program in occupational healthcare provided by the employer (Firsbeat Lifestyle Assessment). The original analysis of the results of the RR-interval recordings was used to provide the participants with an objective assessment of their well-being and health behavior status. The participants are typically healthy, as the RR-interval recording were not to be carried out on individuals with, for example, chronic heart rhythm disturbances, high blood pressure (≥180/100 mmHg), diabetes with autonomic neuropathy, cardiac pacemaker or transplant, fever or other acute disease, a BMI over 40 kg/m2, and any medication influencing HR, HRV or PA levels. (Mutikainen et al., 2014).

Firstbeat Technologies Oy has gathered the data onto an anonymized, constantly growing database. In this study, two different datasets from two different extractions of the datasets at two different time points have been employed. The first dataset, extracted from the servers in October 2013 covered a total of 50,844 measurement days from 18,736 subjects. The second dataset was extracted from the servers in June 2015, and this contained a total of 147,733 measurement days from 52,273 subjects since 2007.

The principles used for the preparation and preprocessing of the datasets have been extensively reported earlier (Pietilä, 2014) but are briefly summarized here. The preprocessing of the data started with a validation of the measurement durations and measurement results as well as removing any duplicated measurements. It was found out that a considerable number of the three-day measurements were broken into multiple measurement segments as the wearable device stopped the recording e.g. if the device was unattached for long enough or if the battery ran empty. Thus, all measurement segments from each subject were compiled by taking into account the start and end times of the measurement segments and the different sampling of the analysis results of the beat-to-beat RR-interval recordings e.g. HR and HRV, the physiological state detection and the oxygen uptake. After compiling all the measurement segments of the subject, the measurement periods were divided into measurement days by taking advantage of the self-reported sleep times. A measurement day was set to start at the start time of the recording (for a new measurement) or at the time of wake-up for the day (for an ongoing measurement). A measurement day was set to end at the time of wake-up in the next morning (for an ongoing measurement) or at the end of the measurement (for the last measurement day). After the measurement days were extracted from the original data and the measurement days were checked to have no inconsistencies, a format for the dataset to be used in the analysis was prepared. Figure 8 shows the information included in the preprocessed dataset: each measurement day consists of the subject's background information, analysis results of the beat-to-beat RR-interval recordings, their work and sleep times together with their alcohol intake as recorded in the participants' electronic journals.

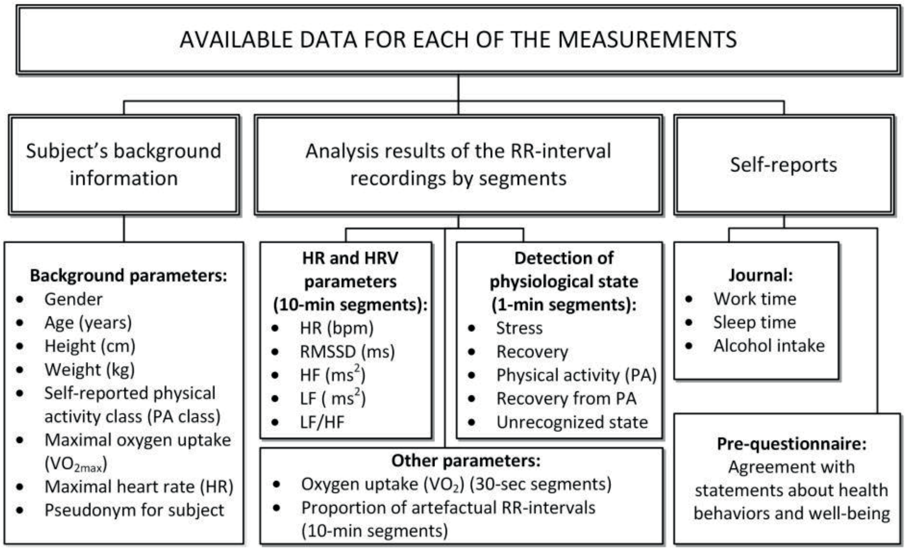


Figure 8. The available data for each of the measurement days in the extracted dataset.

The anonymized background characteristics include the participant's age, gender, height and weight. Their body-mass-indices (BMIs) were calculated from the participants' self-reported height and weight. In addition, the participants were asked to self-report their PA behavior by selecting the appropriate physical activity class (PA class) (Appendix A). The participants' cardiorespiratory fitness i.e. the maximal oxygen uptake (VO_{2max}) was calculated from their background characteristics based on the formulas by Jackson et al. (1990):

$$For\ men: VO_{2max} = 67.350 + 1.921 \times PAclass - 0.381 \times age - 0.754 \times BMI \quad (2)$$

$$For\ women: VO_{2max} = 56.363 + 1.921 \times PAclass - 0.381 \times age - 0.754 \times BMI \quad (3)$$

The maximal HR (HR_{max}) was estimated from age as (Weisman and Zeballos, 2002):

$$HR_{max} = 210 - 0.65 \times age \quad (4)$$

If a higher HR was detected during the recordings, this was used as maximal HR. Before the measurements, the participants were asked to complete a pre-questionnaire about their health behaviors and well-being (Appendix B). The datasets also included the information from the participants journals, which included their self-reported work and sleep times, and their alcohol consumption for each measurement day.

4.2 Analysis of heart rate variability monitoring data for physiological state detection

The RR-interval recordings in the database were performed with either Firstbeat Bodyguard or Firstbeat Bodyguard 2 device (Firstbeat Technologies Oy, Jyväskylä, Finland). The device is attached to the skin with two disposable electrodes: one just below the collarbone on the right side of the body, and the other on the rib cage on the left side of the body. The device detects the beat-by-beat heartbeats with a sampling frequency of 1000 Hz. The recorded RR-intervals have been reported to have a mean absolute error (MAE) of 4.5 ms (Parak and Korhonen, 2013). For the recorded RR-interval data, Firstbeat Technologies Oy has developed a powerful artefact correction algorithm for RR-intervals that employs interpolation methods (Saalasti, Seppänen and Kuusela, 2004). After the artefact correction, the MAE of the RR-intervals has been reported to have decreased to 2.3 ms (Parak and Korhonen, 2013). The information about the proportion of erroneous RR-intervals in each 10-minute segment of the recording was also saved to the dataset.

The artefact-corrected, equidistantly resampled and band-pass (0.03–1.2 Hz) filtered RR-intervals are then used to extract the HR, HRV and VO_2 time series data (Firsbeat Technologies Oy, 2014). For the dataset extraction, the mean values of HR and all HRV parameters were calculated for 10-minute non-overlapping signal segments. The RMSSD was calculated based on 5-minute RR-interval segments, and the frequency-domain HRV parameters of LF and HF were obtained using short-time Fourier transform (Martinmäki et al., 2006; Firstbeat Technologies Oy, 2014). The respiration rate was extracted based on the rhythmic changes in the RR-interval time series, and the VO_2 in 30-second segments was estimated based on HR data and the subject's HR_{max} and VO_{2max} (Saalasti, 2003; Kettunen and Saalasti, 2005a; Firstbeat Technologies Oy, 2005). This VO_2 estimation method employing a neural network model has been shown to correlate strongly (correlation coefficient: ≥ 0.75) with the measured VO_2 in physical activities ranging from low to vigorous in intensity (Smolander et al., 2011; Robertson et al., 2015).

Firstbeat Technologies Oy has also developed an algorithm for classifying the subject's physiological state of stress, recovery and physical activity based on the HR, HRV and VO_2 parameters (Kettunen and Saalasti, 2005b; Firstbeat Technologies Oy, 2014). The physiological state classification algorithm is based on the differences in the physiological responses as shown by the HR, HRV and VO_2 parameters (Kettunen and Saalasti, 2005b). PA is primarily assessed from the increased HR and VO2, while recovery and stress are assessed by obtaining the HR and HRV parameters related to the SNS and PNS modulation of the ANS (Kettunen and Saalasti, 2005b). PA increases cardiac activity and/or metabolic rate, and these correspond to increased HR and/or VO_2. In addition, acceleration data can be used to further improve the detection of PA (Kettunen and Saalasti, 2005b). Recovery is associated with predominant PNS regulation of the ANS seen as decreased HR and increased HRV, and especially HF power (Kettunen and Saalasti, 2005b). Stress is associated with increased SNS regulation of the ANS, which corresponds to increased HR and decreased HRV but no increase in VO_2 (Kettunen and Saalasti, 2005b). The PA, stress and recovery reactions are assessed by taking into account the individual ranges for the physiological reactions e.g. minimal and maximal HR (Firstbeat Technologies Oy, 2014). The indices for PA, stress and recovery are calculated from the relevant HR, HRV and VO_2 parameters, and the presence of PA, stress and recovery is simply detected by using thresholds for the indices (Kettunen and Saalasti, 2005b). The intensity of the PA can be estimated from the VO_2 level (Kettunen and Saalasti, 2005b). Based on the intensity, the PA may be categorized as either light or intense, using different thresholds (Kettunen and Saalasti, 2005b). The indices of stress and recovery may also be used for assessing the intensity of the stress and recovery reactions, respectively (Kettunen and Saalasti, 2005b).

The validity of this physiological state classification algorithm has been evaluated in different studies. The VO_2 estimation based on the RR-intervals correlates strongly (correlation coefficient: ≥ 0.75) with the measured VO_2 in physical activities of varying intensities (Smolander et al., 2011; Robertson et al., 2015). The VO_2 estimation method is regarded as suitable for field studies due to its practical utility and good accuracy (Smolander et al., 2011; Robertson et al., 2015).

The cortisol levels after awakening have been associated with the quantity of stress and recovery reactions detected with this algorithm during sleep (Rusko et al., 2006). These quantities can be measured in minutes or in percentage terms, so a higher number of recovery minutes or percentage and a lower number of stress minutes or percentage detected during sleep is reflected in a lower cortisol level after awakening (Rusko et al., 2006). Based on the physiological state classification algorithm in daily living settings, the quantity of recovery has been reported to be higher during sleep than waking while the quantity of stress is higher during waking than sleep (Kinnunen et al., 2006; Uusitalo et al., 2011).

In addition, using the physiological state classification algorithm, the intensity of detected stress reactions and the amount of recovery during sleep in RR-interval recordings during daily life seem to be associated with subjective stress levels (Föhr et al., 2015). A higher amount of stress classified during work and the daytime has been associated with self-reported occupational burnout symptoms (Teisala et al., 2014), and a decreased quantity of recovery has been associated with higher self-reported effort at work (Uusitalo et al., 2011). Mental strain has also been associated with the intensity of stress reactions in a within-subject study design (Kinnunen et al., 2006).

Using this algorithm, it has also been reported that body composition, PA level and cardiorespiratory fitness also affect the levels of stress and recovery (Teisala et al., 2014). In addition, one's subjective feeling of stress may not always accord with the physiological state classification of stress, and the physiological reaction may be similar regardless of whether the subjective experience of the stress event is negative or positive (Oksman, Ermes and Tikkamäki, 2016; Kaikkonen, Lindholm and Lusa, 2017).

4.3 Summary of the methodologies for Publications

The use of the database described in section 4.1 for the research has been approved by the Ethics Committee of Tampere University Hospital (reference number R13160), valid for all the Publications of this thesis. Table 3 summarizes the study objectives, designs and methods of the Publications. The detailed description of the methodologies used in the Publications are given in sections 4.4 and 4.5.

Table 3. A summary table of the objective, variables of interest and methodologies for the Publications.

Publication	Objective	Study design(s)	Variables of interest	Analysis outcomes and methodologies
I	Temporal (month and weekday) variations in cardiorespiratory-fitness-enhancing PA minutes ($MVPA_{MET}$10min)	5,124 subjects, cross-sectional	$MVPA_{MET}$10min weekday, month age, gender, BMI, PA class	PA minutes for weekdays and months were adjusted for the subjects' demographics using LR to compensate the non-balanced sampling in the data.
II	The impact of PA quantification method using *absolute* ($MVPA_{MET}$) and *relative* to subject's fitness level ($MVPA_{VO2R}$) criteria, and subject demographics on the estimated mean PA minutes per day	23,224 subjects, within-subject & cross-sectional	$MVPA_{MET}$, $MVPA_{VO2R}$ age, gender, BMI	PA minutes by *absolute* and *relative* criteria were compared within subjects using the Wilcoxon signed rank test, and between subjects in age and BMI groups using the Kruskal-Wallis test, separately for men and women.
III	The association of individual lifestyles and daily activities (e.g. alcohol intake, PA, sleep time) with the ANS regulation during sleep (recovery minutes)	6,228 subjects, cross-sectional	recovery minutes alcohol, BMI, PA class, age $LightPA_{VO2max}$, PA_{VO2max}	The associations of lifestyles and daily activities with the ANS regulation during sleep and their importance were explored using RF.
IV	The association between acute alcohol intake as well as its interactions with subject background characteristics, and the ANS regulation during sleep (HR, RMSSD, recovery% and recovery index)	4,098 subjects, repeated-measures within-subject	HR, RMSSD recovery%, recovery index hour of sleep age, gender, BMI, PA class	The association between alcohol intake and the hour-by-hour ANS regulation during sleep was studied with two-way repeated-measures ANOVA. The interactions of subject's demographics and alcohol intake with the ANS regulation during sleep was studied with LR.
V	The association of PA behavior (weekly $MVPA_{MET}$10min) with the ANS regulation during sleep (stress balance, recovery index)	16,275 subjects, cross-sectional	weekly $MVPA_{MET}$10min stress balance, recovery index age, gender, BMI	The association between the PA behavior and the ANS regulation during sleep was assessed using linear and Tobit regressions.

4.4 Study designs and participants

In this thesis, the observational dataset of HRV monitoring results was mainly used for cross-sectional studies (Publications I, II, III, V) but the three-day recording also enabled the within-subject study design to be utilized (Publication IV). Two anonymized datasets were used in Publications I–V. Publications I and III employed the dataset extracted in October 2013, which included 50,844 measurement days from 18,736 subjects, while Publications II, IV and V used the dataset extracted in June 2015 consisting of 147,733 measurement days from 52,273 subjects. In Publication I, only a random sample covering one third of the original dataset was used, but in Publications II–V all the available data was used. In fact, the analysis from Publication I was repeated using the whole dataset, and the results were similar, although this data is not shown.

Because the data was for secondary use, the eligibility criteria were set a posteriori for both the data and the participants, which ensured the quality of the data and the subsequent analyses. Some of the eligibility criteria were the same for each publication, while some were specific to the research question and data analysis scheme. Simplified flow charts of the participants and monitoring days included in the analysis for Publications I–V are shown in Figure 9.

The general eligibility criteria used in all the Publications I–V considered both the background characteristics of the subjects and the quality of the data. Only subjects who were 18–65 years old and had a BMI of 18.5–40 kg/m^2 at the time of the measurements were considered for the analysis. With regard to data quality, the proportion of erroneous RR-intervals and no break longer than 30 minutes during the recording was accepted. In addition, any missing data in the variables of interest, such as sleep times or alcohol intake, were also used as exclusion criteria.

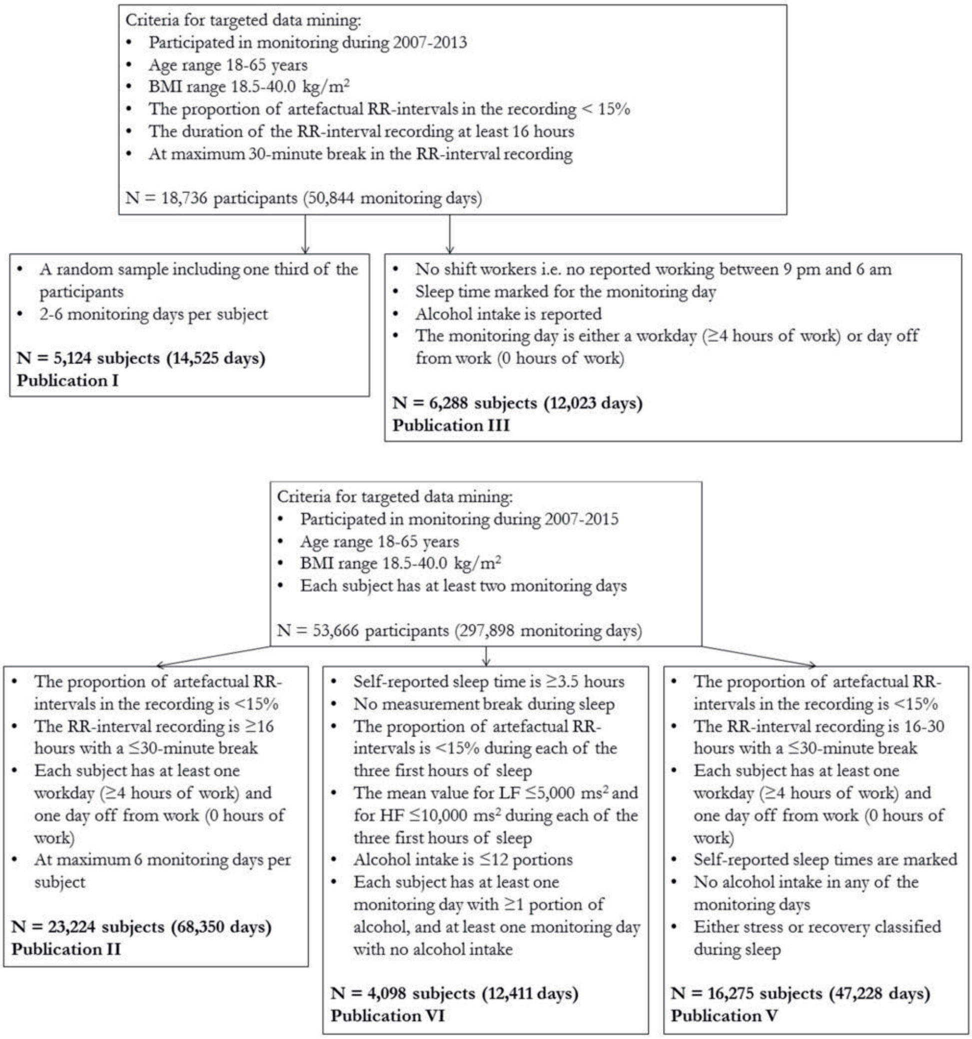


Figure 9. The eligibility criteria for the participants and recordings used in the Publications.

Any further eligibility criteria for the participants depended on the research question and the study design. To estimate the overall mean PA minutes per day in Publication II, each participant was required to have an eligible recording for at least one workday and for at least one day off from work. The same applied to Publication V, where each participant had to have at least one workday and one leisure day recorded so that the weekly minutes of PA could be assessed. In Publication IV, the analysis used a within-subject study design and thus, each of the eligible participants had to have at least one measurement day with no alcohol intake and at least one measurement day with at least one alcohol unit marked. In Publication III, the shift workers were excluded from the analysis by applying an exclusion criterion for participants who reported working between 9 pm and 6 am.

The requirements for the RR-interval recordings were used to ensure the quality of the data. As the analysis results of the RR-interval recordings were used from the whole measurement day, the measurement day was included in the analysis only if its duration was at least 16 hours (Publications I, II, III, V). In Publication IV, thresholds were set for the LF and HF parameters to ensure the quality of the data. In Publication V, the measurement days with alcohol intake were removed to control for the effect of alcohol intake on ANS regulation during sleep.

4.5 Quantification of physical activity, autonomic nervous system regulation and self-reported behaviors

4.5.1 Physical activity

There are various aspects to PA (Garber et al., 2011) but this thesis focuses on cardiorespiratory PA and PA behavior. Intensity is one of the key elements in describing cardiorespiratory PA (Howley, 2001; Garber et al., 2011). For cardiorespiratory exercise, the intensity of the PA should be at least moderate, but there are several methods to estimate PA intensity (Garber et al., 2011). In this thesis, the PA bouts have been quantified in terms of PA intensity based on the estimated VO_2 time series data. Furthermore, the intensity of the PA has been used to study the PA volume, which is expressed as the number of PA minutes. In short, the three different criteria for assessing the PA bouts used in this thesis are:

1) "PA based on MET levels": The *absolute* intensity of the PA was assessed by the MET i.e. the oxygen uptake (VO_2) divided by the resting metabolic rate, assumed to be 3.5 ml/kg/min: MET=VO_2/3.5 (ml/kg/min). For *absolute* intensity PA, the recommended threshold for moderate-to-vigorous intensity PA is ≥3 METs (Garber et al., 2011; Piercy et al., 2018). In the results of this thesis, the number of minutes during which the *absolute* intensity of the PA is moderate-to-vigorous ($MVPA_{MET}$) are reported.
2) "PA based on VO_2R": The *relative* intensity of the PA describes the intensity of PA *relative* to the subject's physical fitness and it was assessed by the % VO_2R (Howley, 2001; Garber et al., 2011). The recommended threshold for

moderate-to-vigorous PA is ≥40% VO_2R (Howley, 2001; Garber et al., 2011). Using this criteria, the number of minutes during which the *relative* intensity of the PA is moderate-to-vigorous ($MVPA_{VO2R}$) are reported in the results of this thesis.

3) "PA based on VO_{2max}": The algorithm for detecting the physiological state from the RR-intervals, developed by Firstbeat Technologies Oy (Firstbeat Technologies Oy, 2014), has been employed in the quantification of PA in this thesis. The detection of PA was based on the intensity of PA, assessed relatively by the cardiorespiratory fitness measured as the VO_{2max} of the subject. A VO_2 level at 20–30% of VO_{2max} was regarded as light PA, while a VO_2 level at >30% of VO_{2max} was regarded as physical exercise. The amount of time spent doing light PA ($LightPA_{VO2max}$) or physical exercise (PA_{VO2max}) is used in the results of the thesis.

PA bouts with at least moderate *absolute* intensity and a duration of at least 10-minutes were regarded as bouts of PA that enhance cardiorespiratory fitness, as recommended by Garber et al. (2011). To adapt this recommendation for continuous measurement, the PA bouts were defined as being at least moderate *absolute* intensity (≥3 METs) for at least 10-minutes, although allowance was made for the MET-level to be lower than moderate for up to 1 minute. Using this definition, the cardiorespiratory PA bouts were quantified. The number of minutes regarded as cardiorespiratory-fitness-enhancing PA bouts ($MVPA_{MET}$10min) was also used as a measure of PA in this thesis.

The number of minutes per week of cardiorespiratory-fitness-enhancing PA (weekly $MVPA_{MET}$10min) was used as an objective assessment of the participants' PA behavior. For each workday and day off from work, the mean $MVPA_{MET}$10min was calculated so that the number of PA bouts exceeding the vigorous *absolute* intensity (≥6 METs) were doubled:

$$MVPA_{MET}10\,\mathrm{min} = moderate\ PA\ minutes + 2 \times vigorous\ PA\ minutes. \quad (5)$$

The weekly $MVPA_{MET}$10min was further estimated as the weighted sum of the mean PA minutes from workdays and from leisure days (Mutikainen et al., 2014):

$$\begin{aligned} MVPA_{MET}10\,\mathrm{min} = {} & 5 \times mean\ workday\ MVPA_{MET}10min + \\ & 2 \times mean\ leisure\ day\ MVPA_{MET}10min. \end{aligned} \quad (6)$$

Based on the weekly $MVPA_{MET}10min$, the subjects were categorized into four PA groups: inactive (weekly $MVPA_{MET}10min$ of 0 minutes), low activity (weekly $MVPA_{MET}10min$ of 0–<150 min), medium activity (weekly $MVPA_{MET}10min$ of 150–300 min), and high activity (weekly $MVPA_{MET}10min$ of >300 min).

The $MVPA_{MET}$ and $MVPA_{VO2R}$ were used in Publication II to assess the impact of the PA quantification method on the estimated levels of PA. The $LightPA_{VO2max}$ and PA_{VO2max} were used for studying the association that light daytime PA and physical exercise had with the next night's sleep (Publication III). The $MVPA_{MET}10min$ was used to study temporal variations in the PA (Publication I), and the association that daytime cardiorespiratory-fitness-enhancing PA had with the next night's sleep (Publication III). The PA groups based on the subjects' weekly $MVPA_{MET}10min$ were used in Publication V to study the association between PA behavior and sleep.

In addition to the objective PA measures, the self-reported PA class was employed as a subjective estimate of the subject's PA behavior and physical fitness. The PA class was assessed by a questionnaire in which the participants were asked to self-evaluate their weekly PA by its typical intensity, duration and frequency (Appendix A). This PA class has 10 categories, of which categories 0–3 are inactive, categories 4–6 are moderately active and categories 7–10 are athletic subjects. The self-reported PA class was employed as a continuous variable in Publications I, III and IV. In Publication I the PA class was used to control for the difference in PA behavior and physical fitness of the subjects. In Publication III it was used to study the long-term effect of PA behavior and physical fitness on sleep. In Publication IV it was used to show the interaction of alcohol intake, PA behavior and physical fitness on the ANS regulation during sleep.

4.5.2 Autonomic nervous system regulation during sleep

The ANS regulation during sleep was assessed based on the analysis results of the RR-interval recordings using the traditional HR and HRV parameters as well as the stress and recovery reactions, detected by the physiological state classification algorithm (Kettunen and Saalasti, 2005b). The traditional HR and HRV parameters calculated during sleep were HR, RMSSD, LF, HF and LF/HF ratio. From the artefact corrected RR-intervals, the HR was calculated as the mean over 10-minute segments, and the RMSSD, first calculated using 5-minute segments, was also averaged for 10-minute segments. The LF, HF and LF/HF ratio are based on the power in the frequency spectrum estimated using short-time Fourier transform

(Martinmäki et al., 2006). The LF, HF and LF/HF ratio were also calculated as the mean for 10-minute segments. The lower the value is for the HR and the higher it is for the HRV, the more predominant is the PNS in ANS regulation during sleep.

The amount of recovery reaction during sleep was assessed by using the analysis results for the detection of the physiological state. The proportion and the number of minutes with recovery reactions detected during sleep were used as a simple measure of recovery in the analysis, and have been used also in the previous studies of the field (e.g. Kinnunen et al., 2006; Rusko et al., 2006; Uusitalo et al., 2011). Another measure for recover was stress balance that has also been used in the previous studies (e.g. Teisala et al., 2014; Föhr et al., 2015). Stress balance indicates the proportion of stress and recovery reactions during sleep, and it is calculated as based on the number of minutes detected to be recovery and stress (Firstbeat Technologies Oy, Jyväskylä, Finland):

$$Stress\ balance = \frac{Recovery\ minutes - Stress\ minutes}{Recovery\ minutes + Stress\ minutes} \qquad (7)$$

Consequently, the stress balance can vary from -1 to 1, and values close to 1 show a higher proportion of recovery reactions than stress reactions during sleep. Both higher stress balance values and more recovery reactions detected during sleep are interpreted as more recovery during sleep, and thus higher PNS predominance in the ANS regulation during sleep.

In addition, the intensity of the recovery reactions, called the recovery index, was also employed (Kettunen and Saalasti, 2005b). The recovery index describes the magnitude of the PNS modulation, and it is high when the individual HR is low and the individual HRV is high (Firstbeat Technologies Oy, 2014). Thus, a higher recovery index value is interpreted as higher PNS predominance in ANS regulation.

The ANS regulation during sleep was studied in Publications III, IV and V. In Publication III, the amount of recovery during sleep was used as the outcome for studying the associations of individual lifestyles and daily activities with sleep. The amount of recovery during sleep was calculated as the sum of the minutes during which recovery reactions were detected (recovery minutes). In Publication IV, the association between acute alcohol intake and the ANS regulation during sleep was studied using the traditional HR and HRV parameters as well as the amount and intensity of recovery during the first three hours of sleep. The amount of recovery was assessed as a percentage of sleep time, i.e. the proportion of time with detected recovery reactions over the total sleep time (recovery percentage). In Publication IV, the outcomes were calculated as the hourly mean values for each of the first three hours of sleep, and as the average during the first three hours. In Publication V, the ANS regulation during sleep was assessed using the stress balance and the intensity of recovery.

4.5.3 Self-reported behaviors

Information about the self-reported behaviors was taken from the diaries the participants kept during the RR-interval recordings. The work and sleep times were the first values to be extracted. The work times were used to distinguish between workdays and days off (leisure days). A workday included at least 4 hours of work, and a day off had 0 hours of work (Publications II, III and V). The sleep times were used for assessing sleeping (Publications III, IV and V), the duration of the sleep (Publication III), and the sleep onset time (Publication III and IV). For the analysis in Publication IV, the sleep onset time was set to be 30 minutes after the self-reported sleep onset time in an attempt to ensure that the subjects really were sleeping during the time analyzed as sleep.

The other major self-reported behavior in the diaries was the alcohol intake. The participants reported the amount of alcohol consumed during a measurement day in standardized units equal to 12 grams of ethanol. A typical standardized unit of alcohol is, for example, a bottle (33cl) of beer or cider, or one glass (12cl) of wine. In Publications III and IV, alcohol intake was quantified as the amount of ethanol per kilogram of the subject's weight in order to standardize the alcohol intake for body weight. In Publication IV, the alcohol intake of the subjects was also categorized into low (≤0.25 g/kg), moderate (>0.25–0.75 g/kg) and high (>0.75 g/kg) dose.

4.6 Statistical data analysis

All the data processing and statistical data analyses were conducted using MATLAB (The MathWorks Inc., Natick, MA) versions R2013b (Publication I) and R2015b (Publication II), and R (The R Foundation for Statistical Computing, Vienna, Austria) version 3.2.2 (Publications II, III, IV, V). The level of significance in statistical analysis was set to <0.05 in Publications I, III, IV and V and to <0.001 in Publication II due to the large size of the dataset. However, with all the results of this research, the emphasis should be put on the effect sizes, due to the issues with significance level in large-scale datasets, as discussed in section 2.4.3.

4.6.1 Statistical hypotheses testing

Traditionally, research studies such as RCTs focus on testing a statistical hypothesis that has been set a priori (Berger et al., 2009). In statistical hypothesis testing, the statistical hypothesis is a statement about the data that is either accepted or rejected based on a statistical test (Vidakovic, 2011; Lin and Jiang, 2013). In statistical hypothesis testing, the null hypothesis (H_0) and the alternative hypothesis (H_1) are formulated (Vidakovic, 2011). The idea in statistical hypothesis testing is to use a statistical test to study whether or not the data supports the H_0 and thus, to conclude whether or not the H_0 should be accepted or rejected. In other words, a hypothesis test is applied for an appropriate test statistic in order to estimate the probability of obtaining equally or more extreme test statistics in the observed data under H_0, commonly known as the p-value (Lin and Jiang, 2013). Thus, H_1 is typically set as the statement that it is hoped can be established with the statistical hypothesis test, and H_0 is set as the negation of H_1 (Vidakovic, 2011).

In the case of statistical hypothesis testing for two samples, the null hypothesis for the equality of the means in two populations, for example, would be formulated as H_0: $\mu_x = \mu_Y$ and the alternative two-sided hypothesis would be formulated as H_1: $\mu_x \neq \mu_Y$ (Lin and Jiang, 2013). Depending on the situation, one-sided alternative hypotheses may also be used, i.e. H_1: $\mu_x > \mu_Y$ or H_1: $\mu_x < \mu_Y$ (Vidakovic, 2011; Lin and Jiang, 2013).

If the statistical hypothesis testing is for more than two samples, the one-way analysis of variance (ANOVA) methodology is used, which is a generalization of the two-sample test of equality of the means for several populations (Vidakovic, 2011; Lin and Jiang, 2013). Consequently, in a one-way ANOVA, the H_0: $\mu_1 = \mu_2 = \ldots = \mu_n$ and the H_1: $\mu_i \neq \mu_j$ for at least one (i,j) pair (Vidakovic, 2011). The ANOVA methodology can also be used to study the effect of two factors on the outcome measure (Vidakovic, 2011; Lin and Jiang, 2013). In a two-way ANOVA, a total of three pairs of H_0 and H_1 are tested: one for the effect of each of the two factors and one for the interaction of the two factors (Vidakovic, 2011). In all three pairs, the null hypothesis is for equality and the alternative hypothesis is for inequality (Vidakovic, 2011).

In addition to studying the differences of the test statistics between the populations, statistical hypothesis testing may be used with a repeated-measures design (Vidakovic, 2011). In a repeated-measures design, the statistical hypothesis testing is typically performed on data originating from the same subjects under different conditions. Thus, a repeated-measures design controls for variability between the subjects but also cause dependence between the data points (Vidakovic, 2011). Two-sample statistical tests can be applied by employing the paired version of the statistical tests, and ANOVA is the generalization of the statistical hypotheses testing in the case of more than two conditions (Vidakovic, 2011).

In Publication III, because the PA volumes were not normally distributed, statistical hypothesis testing was employed to compare the $MVPA_{MET}$ and $MVPA_{VO2R}$ using two-sided and non-parametric statistical tests. $MVPA_{MET}$ and $MVPA_{VO2R}$ were compared separately for men and women, within each age and BMI category, using the Wilcoxon signed rank test. This is a non-parametric paired statistical test for the differences of means (Vidakovic, 2011). In addition, $MVPA_{MET}$ and $MVPA_{VO2R}$ were compared between the age and BMI categories using the Kruskal-Wallis test, a statistical test designed for the comparison of medians between at least two samples (Vidakovic, 2011). The Kruskal-Wallis test is analogous to a one-way ANOVA but without the assumption of independent and normally distributed populations (Vidakovic, 2011).

In Publication IV, a within-subject, two-way, repeated-measures ANOVA was used to study the difference in the shape of the hour-by-hour pattern of HRV parameters between the days with and without alcohol intake. The two-way ANOVA was applied separately on the three different alcohol-dose groups (low, moderate and high alcohol intake).

4.6.2 Linear regression model

A linear regression model (LR) can be used for studying the linear associations between the predictor variables and the response variable, and one purpose of the LR can be to predict the value of the response variable based on the value of the predictor variable (Lin and Jiang, 2013). LR can be defined as $\mathbf{Y} = \boldsymbol{\beta}\mathbf{X}+\boldsymbol{\varepsilon}$, where $\mathbf{Y}$ is the vector of the response variable, $\boldsymbol{\beta}$ is the vector with constant term and regression coefficients for the predictors ($\mathbf{X}$), and $\boldsymbol{\varepsilon}$ is the vector of error terms that are assumed to be independent, have a mean of zero and a variance that does not vary with the predictors ($\mathbf{X}$) (Vidakovic, 2011). In addition to simple predictor variables, the interaction of predictor variables can also be included in the LR (Vidakovic, 2011). Typically, the regression coefficients for the LR are estimated using the least square fit for the observed response variable and the observed predictors (Lin and Jiang, 2013). In other words, the regression coefficients are estimated so that the mean of the squared difference between the observed response variable and the estimated response variable values i.e. the predictors multiplied with the regression coefficients, is minimized (Lin and Jiang, 2013).

Statistical hypothesis testing can also be applied to the regression coefficients to determine whether they are different from zero i.e. H_0: $\beta = 0$ and H_1: $\beta \neq 0$ (Lin and Jiang, 2013). The goodness of the regression fit for the observed predictors and the observed response variable can be assessed with the coefficient of determination (R^2) (Lin and Jiang, 2013). This can be calculated as the sum of the squared differences between the regression model outcomes ($\hat{y}_i$) and the mean of the observed response variable ($\bar{y}$) over the sum of the squared differences between the observed response variable (y_i) and the mean of the observed response variable (Lin and Jiang, 2013):

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \quad (8)$$

The LR assumes that the error terms are independent and normal with zero mean and a constant variance (Lin and Jiang, 2013). In addition, the LR assumes that there is no multi-collinearity between the predictors, i.e. the predictors are not correlated (Lin and Jiang, 2013). These assumptions about the error terms and multi-collinearity should be checked after the model has been fitted.

In Publication I, LR was used to adjust for the association between the subject's background characteristics and the volume of cardiorespiratory-fitness-enhancing PA ($MVPA_{MET}10min$). The LR was fitted between the $MVPA_{MET}10min$ (response variable) and the subject's background characteristics of age, gender, PA class and BMI (predictors). Thereafter, the residuals, which are the error terms from the fitted LR, were added to the mean of the observed volumes of PA from which the mean of the residuals was subtracted. In other words, the LR model was used to remove the linear relationship between the predictors and the response variable.

In Publication IV, LR was employed to study the association between alcohol intake and the average ANS regulation during the first three hours of sleep, by also taking into account the subject's background characteristics. In addition, the LR was used to study the interaction between the alcohol intake and the subject's demographics had on the ANS regulation.

In Publication V, the LR was used to study the association of the BMI and weekly $MVPA_{MET}10min$ with the ANS regulation during sleep. In addition, the subject's age was incorporated as a covariate. The LRs were fitted separately for men and women, and the stress balance and recovery index were used as response variables for the LR models to describe the ANS regulation. Because the assumption of the LR regarding the normal distribution of the error terms was not fulfilled for the recovery index regression model, the Box-Cox transformation was applied on the response variable (Osborne, 2010). Because the stress balance values are limited between -1 and 1, the Tobit regression model was applied (Austin, Escobar and Kopec, 2000). The Tobit regression model is used, especially in economics, in the presence of censored data where the true response is not observed due to ceiling effects (Tobin, 1958; Austin, Escobar and Kopec, 2000).

4.6.3 Random forest method

Random forest (RF) is a tree-based machine-learning method that can be used for classification and regression (Hastie, Tibshirani and Friedman, 2009). For regression problems, the RF is a large ensemble of full-grown de-correlated regression trees (Hastie, Tibshirani and Friedman, 2009). The regression trees are built on a bootstrapped sample of the data, and as the trees are full-grown, they have relatively low bias (Hastie, Tibshirani and Friedman, 2009). The binary splits in the regression trees are generated by employing the variable producing the best split from a randomly selected subset of variables (Hastie, Tibshirani and Friedman, 2009). The random subset of variables for the splitting candidates reduces the correlation between the trees in the RF, and it is recommended to randomly select one third of the variables for each split (Hastie, Tibshirani and Friedman, 2009). In regression trees, the best split is defined based on the smallest mean squared error (MSE) (Hastie, Tibshirani and Friedman, 2009). The output of the RF is the average of the regression trees, which reduces the variance of the RF (Hastie, Tibshirani and Friedman, 2009).

As the trees in the RF are built on bootstrapped samples, not all samples are used to grow all the trees and the out-of-bag (OOB) samples can be set apart (Hastie, Tibshirani and Friedman, 2009). The OOB samples can be used for estimating the goodness of the model from the OOB error by predicting the RF model outcome for the OOB samples by only using the trees built without the OOB samples (Hastie, Tibshirani and Friedman, 2009). As the OOB error represents the goodness of the RF's fit, the OOB error can also be used for assessing the appropriate number of trees for the RF (Hastie, Tibshirani and Friedman, 2009).

RF has become one of the most popular machine-learning methods (Hastie, Tibshirani and Friedman, 2009; Genuer, Poggi and Tuleau-Malot, 2010). RFs have low bias and variance, and they can be used for modelling complex, non-linear associations (Hastie, Tibshirani and Friedman, 2009). One benefit of RF is the metrics for the importance of the variables, which estimate the variables' prediction strength (Hastie, Tibshirani and Friedman, 2009). The variable importances may be calculated with the so-called permutation method, using the OOB samples (Genuer, Poggi and Tuleau-Malot, 2010). The accuracy of the model is assessed by comparing the original OOB samples with the randomly permuted values for the variables of the OOB samples (Hastie, Tibshirani and Friedman, 2009; Genuer, Poggi and Tuleau-Malot, 2010). The variable's importance is the average decrease in accuracy over the trees caused by the permutation (Hastie, Tibshirani and Friedman, 2009; Genuer, Poggi and Tuleau-Malot, 2010). The variable importances help to identify the most influential predictors in the RF model and the variable importances can also be applied for the feature selection (Genuer, Poggi and Tuleau-Malot, 2010; Gregorutti, Michel and Saint-Pierre, 2017).

It must be noted that the variable importance measures should be interpreted with caution when there are correlated variables in the data (Gregorutti, Michel and Saint-Pierre, 2017). In the case of correlated variables, the feature selection may benefit from the recursive feature elimination (RFE) approach (Gregorutti, Michel and Saint-Pierre, 2017). In RFE, the feature with the lowest variable importance is recursively removed from the model, and the new RF model and new values for the variable importances are calculated (Gregorutti, Michel and Saint-Pierre, 2017). The set of variables yielding the lowest prediction error in the RF may be used to maximize the accuracy of the RF model for prediction (Genuer, Poggi and Tuleau-Malot, 2010).

The variable importances, however, cannot be used for approximating how the predictors are associated with the outcome. For the interpretation of the RF model it is useful to estimate the association between the most important predictors and the outcome using partial dependence plots (Friedman, 2001). These plots show the dependence between a variable and the outcome by marginalizing the values of other variables (Friedman, 2001; Greenwell, 2017). In practice, partial dependence between a variable and an outcome can be studied by replacing the variable values with a constant while keeping the other variable values as they were and calculating the average of the model outputs (Greenwell, 2017). This procedure averages out the effect of other variables on the outcome, and for the partial dependence plot, the procedure is repeated multiple times with different constant values for the variable (Greenwell, 2017).

RF modelling was used in Publication III to investigate the association of individual lifestyles and daily activities with the ANS regulation during sleep. In Publication III, the analyses were exploratory, so RF modelling was chosen as it requires minimal assumptions and the partial plots can show any non-linearity in the associations. The RFE was used to find the most important variables and to minimize the prediction error of the RF model. The RFE procedure was repeated 10 times to obtain a consensus about the set of predictors minimizing the prediction error. The RF model with the minimal prediction error was used to study the importances of the variables and the associations between the most important predictors and the outcome using partial dependence plots.

5 SUMMARY OF THE RESULTS

5.1 Preprocessing of the datasets and conducting the analysis

Two anonymized datasets extracted from the original database were preprocessed and prepared for the Publications I–V. The first dataset extraction in October 2013 included 50,844 measurement days from 18,736 subjects and was used in Publications I and III. The second dataset extraction in June 2015 included 147,733 measurement days from 52,273 subjects and was used in Publication II, IV and IV. The data preprocessing and preparation work was automated by programming. First, the overall preprocessing principles as described in section 4.1 were implemented step-by-step with programs. The programs were run on the dataset and whenever there was an exceptional measurement case that caused an error in the program, the case was inspected and the program logic was adjusted for any similar cases. After each programming step, the validity of the results was ensured by sanity checks and inspecting different random cases in detail. Altogether, the preprocessing and preparation of the datasets took 1–2 months of work.

After the preprocessing and preparation of the dataset started the work to outline and test the methodologies for the data analysis. The measurements were collected in uncontrolled real-world environment over an extended period of time and thus, the feasibility of the different methodologies used in the literature to quantify behaviors e.g. PA from the MET data were assessed at first. Moreover, the measurements were observational and not collected specifically for any given research purpose. Thus, feasible and meaningful research questions for the available measurement data had to be thought of, as typical in secondary use of data. The iterative development process of the methodologies for the quantification of behaviors and data analysis involved experts from different fields of science, especially from sports and exercise medicine, occupational health as well as health technology.

5.2 Demographics of the subjects

The subjects of the free-living HRV monitoring data used in this thesis represent a real-world sample of Finnish employees. A wide range of both blue- and white-collar workers voluntarily participated in the HRV monitoring. The participants were assumed to be healthy, as the RR-interval recording was not supposed to be performed in conditions of acute or chronic disease, as mentioned in section 4.1.

The demographics of the subjects analyzed in the Publications are shown in Table 4. Approximately 45% of the participants were male, and on average the participants were middle-aged, slightly overweight and did regular PA 2–3 times per week for a total duration of around 2 hours. The BMI and the PA of the subjects are in line with typical values reported for Finnish employees (Vartiainen et al., 2009; Helldán and Helakorpi, 2015; Lahti et al., 2016).

Table 4. The demographics of the subjects employed in the Publications' data analyses.

Publication	Number of subjects	Number of male subjects	Age in years (Mean±SD)	BMI in kg/m^2 (Mean±SD)	PA class (Mean±SD)
I	5,124	2,422 (47.3%)	44.0±9.9	26.1±4.0	4.9±1.9
II	23,224	10,201 (43.9%)	44.7±9.8	26.0±4.0	4.8±1.8
III	6,228	2,788 (44.3%)	45.6±9.6	26.4±4.1	4.8±1.9
IV	4,098	1,811 (44.2%)	45.1±9.6	26.0±4.0	4.8±1.8
V	16,275	6,863 (42.2%)	44.8±9.9	26.0±4.1	4.8±1.8

SD = standard deviation
PA class range: 0–10

5.3 Observations of physical activity behavior

5.3.1 Temporal patterns in physical activity behavior (Publication I)

Figure 10 shows the mean of observed (A) and background-controlled (B) $MVPA_{MET}10min$ for weekdays and months. Because the duration of the measurements was limited to three days per subject and the same subjects were not measured for each weekday and month, the observed $MVPA_{MET}10min$ are confounded by the non-balanced sampling. In other words, differences originating from other factors than only weekdays and months are likely to be present in the observed $MVPA_{MET}10min$ (Figure 10 A).

LR was employed to compensate for the effect of non-balanced sampling on the observed $MVPA_{MET}10min$. In the LR, the observed $MVPA_{MET}10min$ was used as the dependent variable and the background characteristics were used as independent variables. The LR showed that the background characteristics of age, gender, BMI and PA class were all statistically significant ($p<0.05$ for all) and accounted for a total of 15.9% of the variance in the observed $MVPA_{MET}$10min. The LR showed the observed $MVPA_{MET}$10min decreased with increasing age or BMI, and increased for male gender and increasing PA class.

In general, when controlling for the subjects' background characteristics, any variation in the estimate of $MVPA_{MET}$10min between weekdays and months decreased (Figure 10 B). Furthermore, when adjusted for the background characteristics, the estimated $MVPA_{MET}$10min changed more for the monthly mean than the weekday one.

The overall mean of daily $MVPA_{MET}$10min was 23.3 minutes but it was clearly higher for the weekends (28.3 minutes) than for the weekdays (21.2 minutes). Saturday seemed to have the highest $MVPA_{MET}$10min, while Friday and Thursday seemed to have the lowest (Figure 10 B). Looked at monthly, the $MVPA_{MET}$10min means were the highest for January, February and August, and lowest in June and September (Figure 10 B). In general, high $MVPA_{MET}$10min levels were observed at the weekends at the beginning of the year, while low $MVPA_{MET}$10min levels were observed on weekdays (except for Wednesday) in the autumn (September–November).

The number of measurements was clearly lowest (N=912) for Wednesday, but this can be explained by the test protocol which stipulated that the three-day recording period included two workdays and one day off from work, e.g. Thursday, Friday and Saturday, or Sunday, Monday and Tuesday (Figure 10). A similar explanation applies for the monthly measurements, which were clearly lowest (N=161) in July when the majority of Finnish workers have their summer holidays (Figure 10). As expected, a lower number of observations increases the confidence intervals (CIs) (Lin, Lucas and Shmueli, 2013). In July, the 95% CI for the mean $MVPA_{MET}$10min was 5.6 minutes while the 95% CIs for the other monthly averages ranged from 1.3 to 3.0 minutes. For the mean $MVPA_{MET}$10min on Wednesdays, the 95% CI was 1.9 minutes while on the other weekdays the 95% CI was between 1.1 and 1.6 minutes. Consequently, due to the lower number of observations, the estimates for the $MVPA_{MET}$10min on Wednesdays and in July are likely to be less accurate than the estimates of the $MVPA_{MET}$10min at other times.

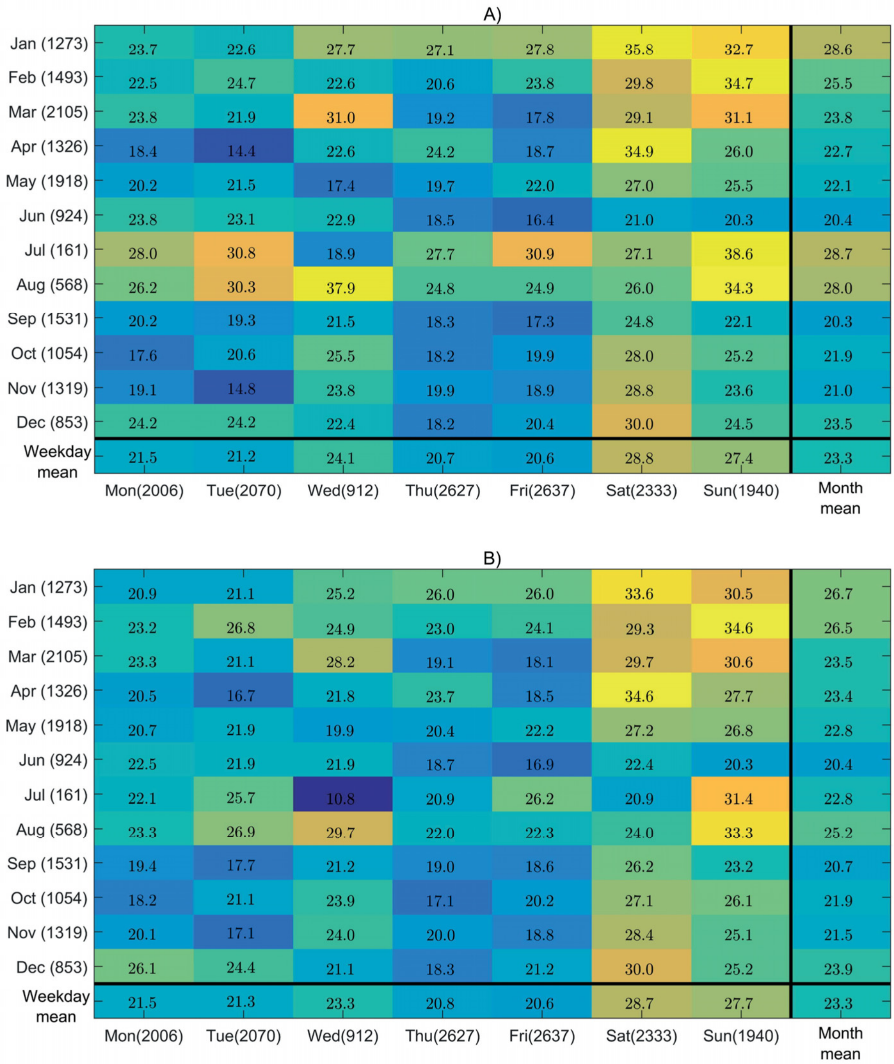


Figure 10. The observed (A) and background-controlled (B) mean number of minutes of cardiorespiratory-fitness-enhancing physical activity bouts ($MVPA_{MET}10min$) for weekdays and months. The number of observations for weekdays and months are shown in the brackets.

5.3.2 Impact of the physical activity quantification method on the estimated levels of physical activity (Publication II)

The impact of the PA quantification method on the estimated levels of PA was studied by comparing the subjects' amounts of PA fulfilling the criteria for VO_2 in *absolute* terms ($MVPA_{MET}$) and *relative* to the subject's fitness level ($MVPA_{VO2R}$). Figure 11 shows the mean with 95% confidence interval (CI) values for $MVPA_{MET}$ and $MVPA_{VO2R}$ in men and women, stratified by age and BMI.

In men, significantly more $MVPA_{MET}$ than $MVPA_{VO2R}$ was observed in all age and BMI categories ($P<0.001$ for all, Wilcoxon signed rank test) (Figure 11). $MVPA_{MET}$ was highest for 18–30 year-old men and lowest for 51–65 year-old men. The difference in $MVPA_{MET}$ between the 18–30 and 51–65 year-old men was substantial: on average 43.5 minutes per day. This means that the observed $MVPA_{MET}$ for 18–30 year-old men was almost twice the observed $MVPA_{MET}$ for 51–65 year-old men. In BMI groups, $MVPA_{MET}$ was highest for normal weight and lowest for obese men with a substantial difference of 25.7 minutes per day. $MVPA_{VO2R}$ was higher for 18–30 year-old men than for 51–65 year-old men, and also for normal-weight men over obese ones. The difference in $MVPA_{VO2R}$ between the age and BMI groups was statistically significant ($P<0.001$ for all, Kruskal-Wallis test) but small, only a few minutes per day.

The results are much the same for women. The $MVPA_{MET}$ decreased substantially ($P<0.001$ for all, Kruskal-Wallis test) with increasing age and BMI (Figure 11). As with the men, the differences in $MVPA_{VO2R}$ between the age and BMI groups for the women were statistically significant ($P<0.001$ for all, Kruskal-Wallis test) but small, only a few minutes per day. In fact, in obese women the $MVPA_{VO2R}$ was higher than the $MVPA_{MET}$ ($P<0.001$, Wilcoxon signed rank test) (Figure 11).

The differences between the age and BMI groups in the estimated PA using the *absolute* ($MVPA_{MET}$) and *relative* to subject's fitness level ($MVPA_{VO2R}$) values may be partly explained by the difference in the subjects' physical fitness, which was estimated by calculating VO_{2max} (Jackson et al., 1990). The subjects' VO_{2max} was estimated based on their background characteristics of gender, age, PA class and BMI using a formula where increasing age and/or increasing BMI and/or being female decrease VO_{2max} (Jackson et al., 1990). This implies that young age, low BMI, high PA-class and male gender are associated with higher VO_{2max}, which corresponds to higher physical fitness. The higher the VO_{2max}, the higher the VO_2 required to achieve the limit of 40% VO2R.

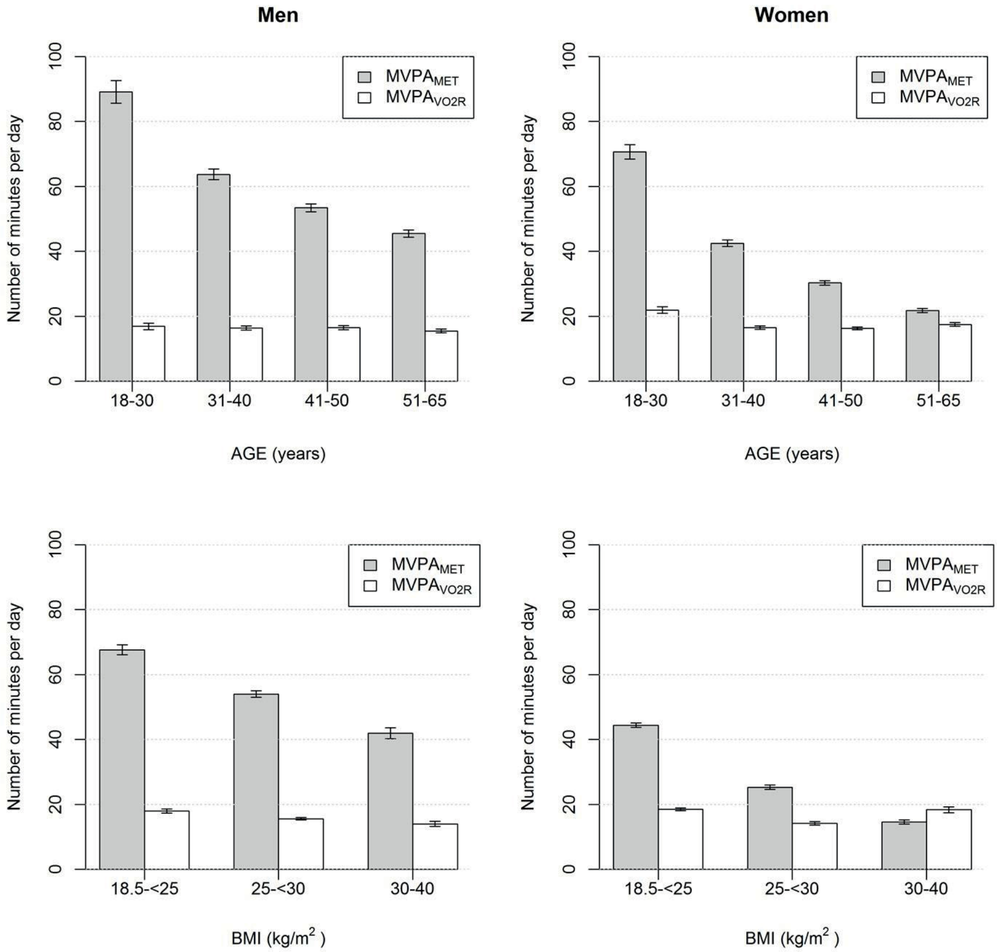


Figure 11. The mean and 95% CI for the number of moderate-to-vigorous intensity physical activity quantified with criteria of *absolute* (MVPA$_{MET}$) and *relative* to subject's fitness level (MVPA$_{VO2R}$) by age and BMI groups, separately for men and women.

5.4 Associations between behaviors and autonomic nervous system regulation during sleep

5.4.1 Associations of individual lifestyles and daily behaviors with the autonomic nervous system regulation during sleep (Publication III)

The impact of individual lifestyles and daily activities on the ANS regulation during sleep was assessed by the importance of the variables in the RF model shown in Figure 12. Alcohol intake during the day was found to be the most important predictor for recovery time during sleep. The second most important predictor was the duration of sleep, and the next most important predictors described the PA during the day. The subjects' background characteristics of age and gender had relatively low importance for predicting the amount of recovery during sleep. In other words, the recovery minutes were relatively similar between the subjects, which indicates that the recovery minutes are analyzed by taking into account the individual characteristics of HRV.

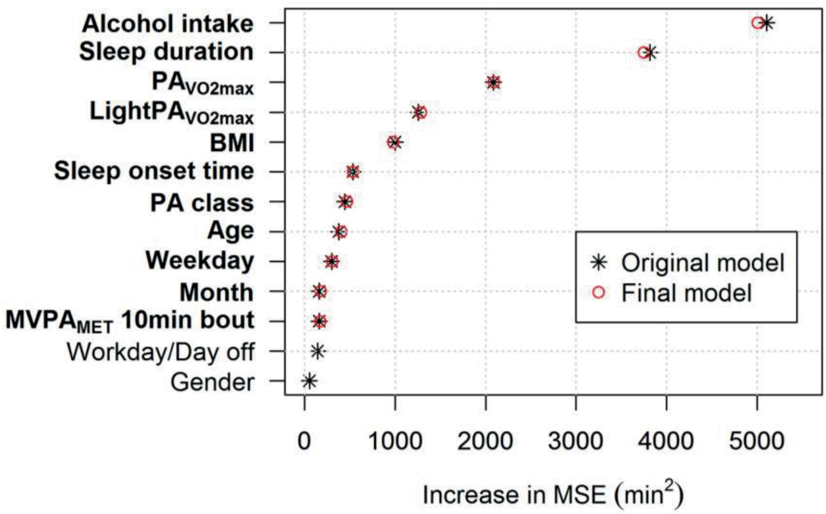


Figure 12. The importance of the predictors for the number of recovery minutes during sleep. The importance of the predictor is expressed as the observed average increase in the mean-squared error (MSE) over all trees in the RF model when permuting the predictor variable. The importances are shown for the original model including all the available predictors with an asterisk, and with a circle for the final model including the predictors with which the minimum MSE is achieved.

Partial dependence plots were constructed for the most important predictors (Figure 13). For all age groups, the decrease in the recovery time with alcohol intake seemed to be dose-dependent: the higher the alcohol intake, the lower the number of recovery minutes (Figure 13 A). For an 80-kg person, one standardized unit of alcohol (0.15 g/kg) did not remarkably decrease the recovery time during sleep, but six units of alcohol (0.9 kg/g) halved the recovery time during sleep compared to a night without alcohol intake. As expected, the recovery time during sleep increased linearly with the duration of sleep as more sleep meant more time for recovery (Figure 13 B). On average, each hour of sleep increased the recovery time by half an hour.

The partial dependencies of $LightPA_{VO2max}$ and PA_{VO2max} were studied in the partial plots in relation to the PA class (Figure 13 C and D). In general, PA during the day seemed to decrease the recovery time during sleep but the subjects with high PA class that were engaged in regular PA and who had good physical fitness had clearly more recovery time during sleep than the inactive and unfit subjects. In other words, good physical fitness seemed to increase the recovery time during sleep but PA during the day seemed to challenge the recovery processes of the body and decrease the recovery time during sleep. Thus, physical exercise on a daily basis might have compromised recovery during the next night's sleep, but the positive effect of PA on ANS regulation during sleep is seen after a delay in the form of increased physical fitness.

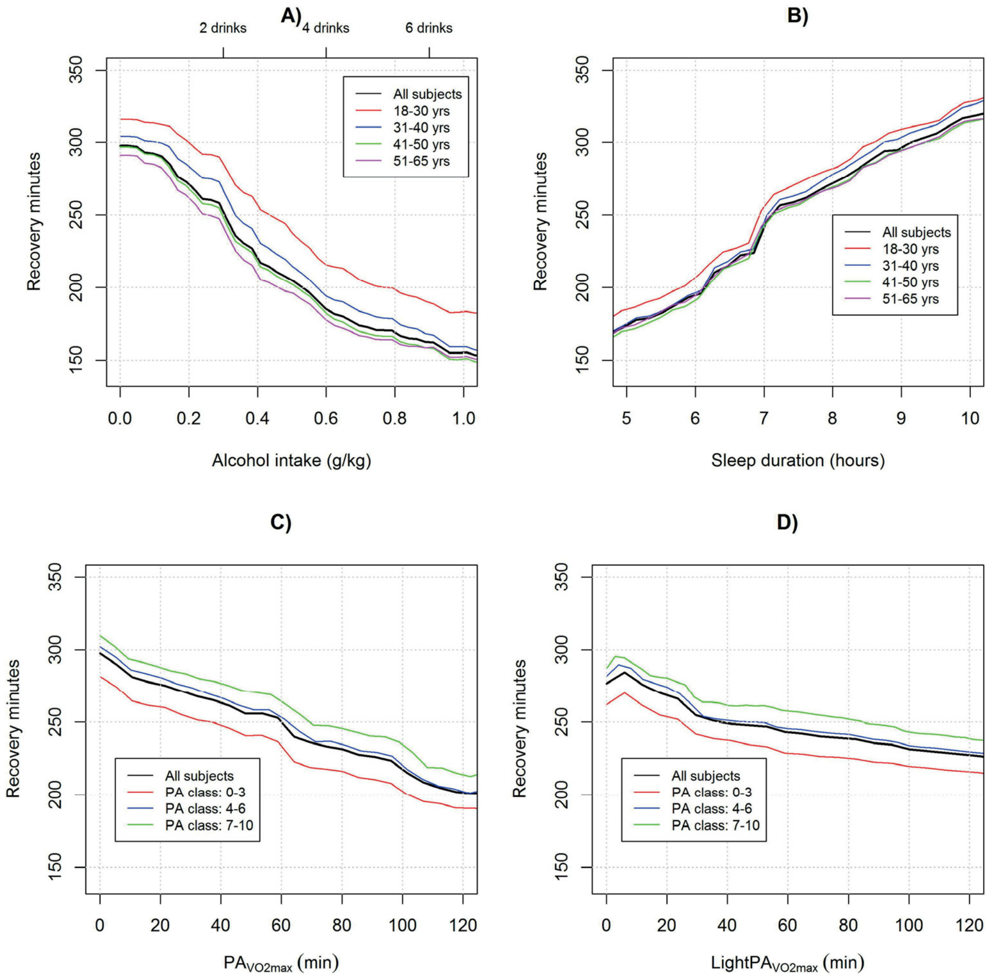


Figure 13. The partial dependence plots for recovery minutes during sleep with respect to A) alcohol intake (drinks marked for an 80-kg person), B) sleep duration, C) minutes of physical activity (PA$_{VO2max}$), and D) minutes of light physical activity (LightPA$_{VO2max}$). For alcohol intake and sleep duration, the partial dependence is shown for all (black line) and for the age groups of 18–30 years (red), 31–40 years (blue), 41–50 years (green) and 51–65 years (magenta) subjects. For PA$_{VO2max}$ and LightPA$_{VO2max}$, the partial dependence is shown for all (black line) and for physical fitness (PA class) groups. Inactive, PA class of 0–3 (red), moderately active, PA class of 4–6 (blue), and active, PA class of 7–10 (green) subjects.

5.4.2 Association between the acute alcohol intake and the autonomic nervous system regulation during sleep (Publication IV)

To further investigate the association between alcohol intake and the ANS regulation observed in Publication III (section 5.4.1) a repeated-measures study design was applied for the subjects who were measured for at least one night with alcohol intake and one night without alcohol intake. The ANS regulation during sleep was assessed with HR, RMSSD and recovery measures, and the alcohol intake was used in the analysis both as a categorized variable (low, moderate and high alcohol intake) and as a continuous one.

On average, low (≤0.25 g/kg), moderate (>0.25–0.75 g/kg) and high (>0.75 g/kg) alcohol intake increased HR by 1.4, 4.0 and 8.7 bpm, respectively. The same categorized variables decreased RMSSD by 2.0, 5.7 and 12.9 ms, recovery percentage by 9.3, 24.0 and 39.2 percentage units, and recovery index by 7.1, 20.8 and 40.2, respectively. Figure 14 shows the mean and 99% CI for the HR, RMSSD, recovery percentage and recovery index during the first three hours of sleep for nights with and without alcohol intake. For all dose groups, differences in ANS regulation parameters were observed between days with and without alcohol intake ($P<0.001$ for all outcomes; two-way repeated-measures ANOVA). In addition, the hourly parameters differed significantly from each other in all dose groups ($P<0.001$ for all). For high dose groups, the hour-by-hour patterns of all HRV parameters ($P<0.001$ for all) were different for the subjects between the days with and without alcohol intake. This was particularly noticeable in the recovery measures: the values increased hour-by-hour in the case of no alcohol intake, but not in the case of high alcohol intake. For the moderate-dose group, the hour-by-hour patterns between days with and without alcohol were only different for the recovery percentage ($P<0.01$). For the low-dose group, the hour-by-hour patterns of HR ($P<0.001$) and recovery index ($P=0.01$) differed between days with and without alcohol.

The interactions between the alcohol intake and the subjects' background characteristics, especially age and physical fitness, on the ANS regulation during sleep were studied using LRs. Alcohol intake increased HR and decreased RMSSD significantly more in younger than in older subjects. For example, a high alcohol dose (0.75 g/kg) increased HR by 1.8 bpm more in a 30 year-old subject than in a 60 year-old one. It also decreased RMSSD by 6.2 ms more for a 30-year-old than for a 60-year-old subject. On average, a high alcohol dose decreased RMSSD by 4.7 ms for a 60-year-old subject, while in a 30-year-old subject the decrease in RMSSD was more than twice this much (10.9 ms). Physical fitness, however, did not have a statistically significant interaction with the alcohol intake in predicting the HRV parameters. In other words, good physical fitness did not seem to protect the subject from the adverse effects of alcohol intake on the ANS regulation.

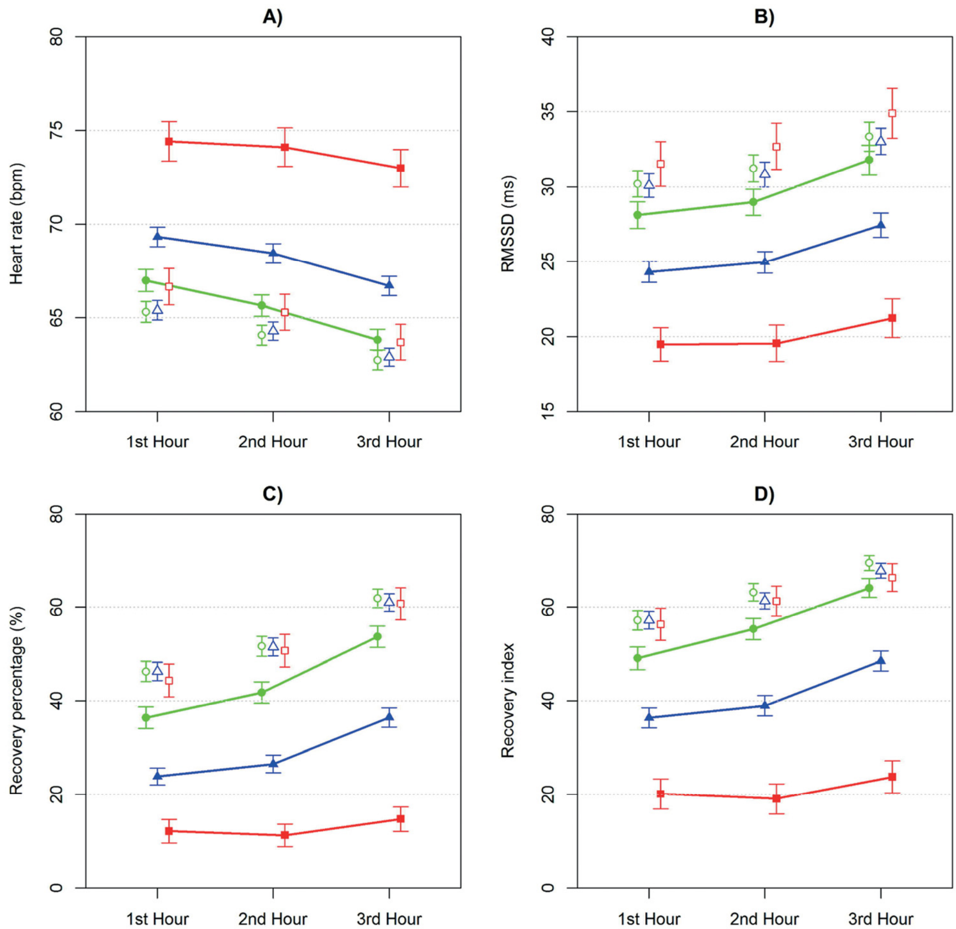


Figure 14. The association of acute alcohol intake with A) heart rate (HR), B) root mean square of the successive differences (RMSSD), C) recovery percentage, and D) recovery index in the first three hours of sleep. The mean and 99% CI values for each hour are marked with solid green circles for the low-alcohol dose group (≤0.25 g/kg), solid blue triangles for medium-alcohol dose group (>0.25–0.75 g/kg) and solid red squares for high-alcohol dose group (>0.75 g/kg) subjects. The corresponding unfilled markers show the mean and 99% CI values for each hour for the subjects without alcohol intake.

The differences observed between the days with and without alcohol intake in the recovery measures did have a strong partial correlation with the differences in the HR (for recovery percentage: r=0.7, P<0.001; for recovery index: r=0.63, P<0.001) and RMSSD (for recovery percentage: r=0.51, P<0.001; for recovery index: r=0.49, P<0.001). However, in contrast to the LRs employing traditional HRV measures, the differences in the recovery percentage and recovery index between the days with and without alcohol were independent of age. This indicates that changes in the ANS regulation assessed with the recovery measures are in line with the traditional HR and HRV measures, but they are independent of age.

5.4.3 Association between the physical activity behavior and the autonomic nervous system regulation during sleep (Publication V)

Figure 15 shows the age-adjusted mean and 99% CI values for stress balance and recovery index for subjects grouped by their BMI and their weekly $MVPA_{MET}$10min. The obese subjects had the lowest stress balance and recovery index, regardless of their PA group (weekly $MVPA_{MET}$10min). Comparisons of stress balance and recovery index between the PA groups were more complicated. The differences in the stress balance and recovery index between the PA groups were smallest for normal-weight subjects. In general, for overweight and obese subjects, the medium- and high-activity PA groups had lower stress balance and recovery index than the inactive group, with the exception of obese high-activity men, who had a higher recovery index than obese inactive men.

In the LR models for stress balance, the medium- and high-activity PA groups were associated with lower stress balance in both men and women, as was increasing age. Furthermore, higher BMI was associated with decreased stress balance in men. Regarding the recovery index, increased BMI and age were associated with a decreased recovery index in both men and women. However, the medium- or high-activity PA groups did not seem to have a statistically significant association with the recovery index.

The amount and intensity of recovery reactions during sleep were higher for subjects with lower BMI, suggesting that their likely better body composition and physical fitness have a positive effect on recovery during sleep. On the other hand, the results for stress balance also showed that the subjects who fulfill the recommendation of 150 minutes of weekly $MVPA_{MET}$10min (Garber et al., 2011) have less recovery during sleep than the inactive subjects. This may be explained by the study design: the weekly $MVPA_{MET}$10min was estimated on the same days that the sleep was analyzed. The subjects in higher PA groups probably had more PA on the measurement days than the subjects in the lower PA groups. It has been reported that nocturnal HR increases and HRV decreases after PA during the day (Hynynen et al., 2010). Thus, the subjects in the higher PA groups might have shown a decreased amount and intensity of recovery because the recovery was detected based on the HR and HRV parameters (Kettunen and Saalasti, 2005b). In the LR models for recovery index and stress balance, the effect size and the variance explained were higher for the BMI than for the PA groups. Thus, the BMI, which likely reflects the body's composition and physical fitness, seemed to explain the amount and intensity of the recovery reactions during sleep better than the PA group that reflects the amount of PA during the measurement days.

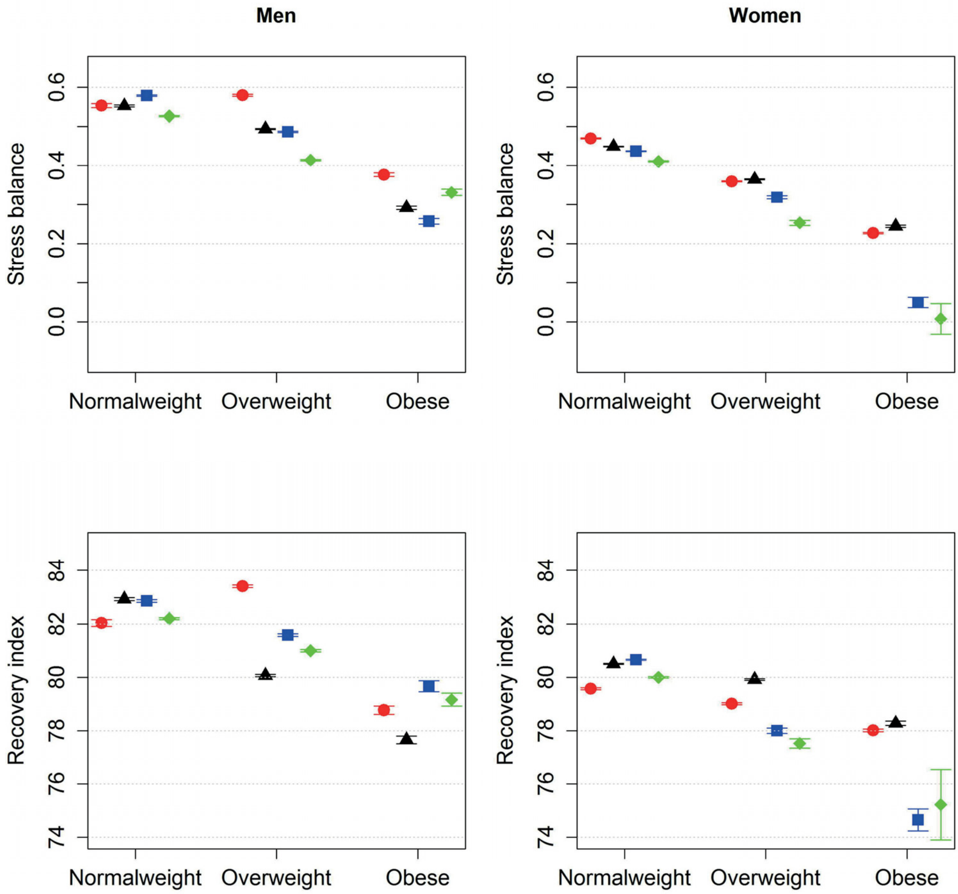


Figure 15. The mean and 99% CI values for the age-controlled stress balance and recovery index in men and women, stratified by BMI group (normal weight, overweight and obese) and objectively measured weekly amount of cardiorespiratory-fitness-enhancing physical activity (weekly $MVPA_{MET}$10min). The red circles represent inactive (weekly $MVPA_{MET}$10min: 0 min), black triangles represent low-activity (weekly $MVPA_{MET}$10min: >0-<150 min), blue squares represent medium-activity (weekly $MVPA_{MET}$10min: 150–300 min), and green diamonds represent high-activity (weekly $MVPA_{MET}$10min: >300 min) subjects.

6 DISCUSSION

6.1 Main findings

The research presented in this thesis investigated PA and sleep behaviors using statistical and machine-learning methods for a unique real-world dataset encompassing HRV recordings and self-reports from a heterogeneous sample including tens of thousands of Finnish employees. To facilitate the studying of meaningful scientific research questions using the dataset, it was required to preprocess the data and to develop methodologies for assessing PA and sleep behaviors from the recordings. After all, it was feasible to use the dataset for the studies but the effort required to produce meaningful scientific results from the monitored data highlights the complexity involved whenever trying to generate insights from real-world data.

Regarding the assessment of PA behaviors, the results of this thesis showed that there were substantial differences in PA behavior from day to day and from month to month. Moreover, the PA quantification method seemed to have a significant impact on the estimated PA levels. With respect to sleep, the results of this thesis indicate that acute alcohol intake is the most important daily behavior associated with the ANS regulation during sleep. Acute alcohol intake seemed to dose-dependently reduce the recovery during sleep and increase the sympathetic regulation in the ANS. In fact, this effect seemed to be more pronounced in younger than in older subjects, which needs to be further investigated in controlled settings. Another important daily activity that affected the ANS regulation during sleep was PA behavior. The PA during the day seemed to challenge the recovery process during sleep, and the beneficial effect of PA on the ANS regulation during sleep was only observed after a delay, in the form of increased physical fitness.

The findings of this thesis were in line with previous studies, and the large-scale real-world data enabled also studying new aspects of PA and sleep behaviors that provided some data-driven hypotheses to be further investigated in controlled studies. Consequently, this thesis demonstrates that research for health behaviors using real-world data can be feasible and can provide valuable practical insights into health and well-being.

6.1.1 Physical activity behavior

Physical activity behavior can be assessed with various methods, and in this thesis the PA behavior was assessed with both self-reporting and objective measurements. In the data used for this study, men and women seemed to have, on average, similar activity levels based on their self-reported PA class that corresponded to having PA 2–3 times per week for a total duration of around 2 hours. Thus, the self-reported PA of the subjects in this study was in line with Finnish employees' self-reports in random sample surveys (Helldán and Helakorpi, 2015).

For the objective measurements, PA was quantified using *absolute* intensity criteria of ≥3 METs, and using *relative* intensity criteria of ≥40 %VO_2R. The number of PA minutes identified using the *absolute* intensity criteria in this thesis were similar for younger adults but clearly lower for older adults when compared to the hip-worn accelerometer-based moderate-to-vigorous PA results reported for a sub-sample of population-based study on Finnish adults (Husu et al., 2016). This difference is likely explained by the differences in the measurement techniques and algorithms employed in the studies. The moderate-to-vigorous PA results from accelerometer were reported based on 6-second epochs (Husu et al., 2016), while for the results of this thesis 1-minute intervals were employed. Thus, the accelerometer-based PA results also include shorter activity bouts that probably would only have a negligible effect on HR.

On the other hand, the number of PA minutes identified using the *relative* intensity criteria in this thesis were similar with the number of moderate-to-vigorous PA minutes reported in international studies (Matthews et al., 2002; Colley et al., 2011), as well as for Finnish adults when using only ≥5-minute bouts from the acceleration data (Husu et al., 2016). When using the threshold of at least 5-minutes for the acceleration in the hip-worn accelerometer-based PA analysis, more likely only the exercise-like PA was taken into account and probably, thus the PA minutes were similar when to the PA observed in this study when using the *relative* intensity criteria. However, it should be noted that these two studies were completely distinct and a comparative study would be needed to truly understand the differences in the PA results between the two measurement modalities.

The PA minutes fulfilling the *absolute* intensity criteria ($MVPA_{MET}$ and weekly $MVPA_{MET}10min$) observed in this thesis were higher for men than for women, which is in line with the previous results obtained, although not statistically significant, from Finnish adults (Husu et al., 2016) as well as internationally (Colley et al., 2011; Tucker, Welk and Beyler, 2011). On the other hand, the PA minutes fulfilling the *relative* intensity criteria ($MVPA_{VO2R}$) for women were similar to, or even higher than, those for men. Similarly, the objectively measured PA employing *absolute* intensity criteria ($MVPA_{MET}$) was significantly higher for younger than for older subjects, and for normal-weight subjects than for overweight or obese ones. The similar differences in the PA levels of younger and older subjects have also been reported in previous studies on Finnish adult population (Husu et al., 2016; Wennmann et al., 2019) as well as internationally (Colley et al., 2011; Tucker, Welk and Beyler, 2011). However, the differences between the age and BMI groups decreased remarkably when the PA was quantified using the *relative* intensity criteria ($MVPA_{VO2R}$).

The differences for PA assessed with *absolute* and *relative* intensity criteria may be partly explained by the fitness level of the subjects. For high-fitness individuals, fulfilling the *relative* (to the subject's fitness) intensity criteria for the oxygen uptake may require more effort than fulfilling the *absolute* intensity criteria for the oxygen uptake. In other words, the high-fitness individuals have a greater VO_{2max} and thus, for high-fitness individuals the VO_2 required to achieve the level of 3 METs is lower than the VO_2 required to achieve the level of 40% VO_2R. On the other hand, low-fitness individuals with a lower VO_{2max} may exceed the limit for *relative* PA more easily than they do for the *absolute* PA level. In other words, for very low-fitness individuals the VO_2 required to achieve the level of 40% VO_2R may be lower than the VO_2 required to achieve the level of 3 METs. In addition, it should be noted that the physical activities during which the high- and low-fitness subjects fulfill the *relative* intensity criteria are probably quite different. For example, for low-fitness individuals, physical activities such as a brisk walk may be enough to achieve the *relative* moderate-to-vigorous intensity (40% VO_2R) level but the high-fitness individuals may need to have more demanding exercise, e.g. jogging, to achieve a similar *relative* PA intensity. Thus, the PA quantification methods employed for detecting PA behavior, even from any objective measurements, have a significant impact on the study results and their interpretation. The *relative* intensity criteria for PA may be the appropriate method, especially, in PA counseling as it is takes into account the subject's fitness level and thus assesses PA in personalized manner.

In the results of this thesis, PA behavior assessed objectively with the amount of cardiorespiratory-fitness-enhancing PA per day ($MVPA_{MET}$10min) seemed to indicate a pattern where the amount of PA varies throughout the year, from day to day and from month to month. On a day-to-day basis, the highest amounts of PA were observed during weekends, and the lowest amounts on Fridays. On a monthly basis, the amount of PA was highest at the beginning of the year and lowest during the autumn. High amounts of PA were observed, especially, at weekends at the beginning of the year.

Also a previous population-based study on Finnish adult population employing wrist-worn accelerometers suggested that PA level is higher during weekends than during weekdays (Wennmann et al., 2019). However, previous studies have shown higher amounts of PA for summer than for winter month in the northern hemisphere (Shephard and Aoyagi, 2009). The previous studies have focused on the general PA level estimated with self-evaluation and pedometers (Shephard and Aoyagi, 2009) while for the analysis of this thesis the available HRV-based recordings were only for three days per subject and thus did not fulfill the requirement of at least one-week recording for habitual PA assessment (Matthews et al., 2002). Consequently, the differences in the methods used to estimate PA behavior seem to be a possible explanation for the conflict in the results considering the seasonal effect in PA levels in previous studies and this thesis.

6.1.2 Associations between behaviors and autonomic nervous system regulation during sleep

The ANS regulation during sleep was found to be affected by several factors related to individual lifestyles and behaviors. Acute alcohol intake was the most important behavior having associated with the ANS regulation during sleep, expressed by the traditional HRV parameters of HR and RMSSD, as well as with the novel indices for the amount and intensity of recovery reactions. The dose-dependent alcohol intake modulated ANS regulation by increasing the sympathetic modulation and decreasing the parasympathetic modulation, which is in line with previous studies (Romanowicz et al., 2011; Sagawa et al., 2011; Ralevski, Petrakis and Altemus, 2018). The changes in the ANS regulation for the first hours of sleep were already observable after only a low alcohol dose (1–2 standardized units of alcohol). The effects of alcohol intake on ANS regulation were similar for men and women, and for physically active and inactive subjects. However, age does seem to play a part as the effects of alcohol intake on the ANS regulation were found to be stronger in younger subjects than in older subjects.

Previous studies had not analyzed the effects of acute alcohol intake on the ANS regulation across different subgroups of the participants (Romanowicz et al., 2011;

Ralevski, Petrakis and Altemus, 2018). Thus, the data-driven hypotheses about the interactions between alcohol intake and gender, age and physical activity behavior on the ANS regulation provided in this thesis should be further studied in controlled settings to verify these hypotheses. Any further studies should also investigate the effect of acute alcohol intake on the ANS by taking into account the timing of alcohol intake, as this data was unfortunately not available in the study material for this thesis.

Another important behavior associated with ANS regulation during sleep was PA. The results showed that the parasympathetic regulation of the ANS during sleep increased with overall physical fitness, but it diminished after a day with PA, as previous studies have also suggested (Hynynen et al., 2010). Thus, any positive effect that PA has on the ANS regulation is not immediate, but will appear after a delay in the form of increased physical fitness. The increased physical fitness level may be indicated by high self-reported physical fitness, regular PA, or by the subject's normal BMI. On the other hand, the disturbing effect of PA on the ANS regulation during sleep implied that days of rest are important for avoiding accumulated negative effects of PA on ANS regulation during sleep.

Due to the clear and significant effects that an individual's daily behaviors have on the ANS regulation during sleep, any studies relying on HRV measurements taken during sleep should also take into account, or control, the daily behaviors of the subjects. In practice, wearable monitoring devices could be used to inform the users about the effects that their daily behaviors may have on sleep, but also to show the long-term effects of such behaviors.

6.1.3 Use of real-world wearable health monitoring data

This thesis shows that use of real-world wearable health monitoring data for scientific research is feasible and enables studying new aspects of health behaviors. The results of this thesis were obtained from an observational real-world health monitoring dataset from a large and heterogeneous sample of Finnish employees who had voluntarily conducted the recordings of beat-to-beat RR-interval data in uncontrolled free-living settings. To ensure the quality and representativeness of the data, the inclusion criteria were set a posteriori for both the measurements and the subjects. The demographics of the subjects were similar to the demographics reported for Finnish employees in other studies (Vartiainen et al., 2009; Lahti et al., 2016). Thus, the results from the analysis of the dataset can be assumed to be generalizable for vast majority of Finnish workers.

In general, the results from observational RWD should be interpreted carefully. Due to the large sample sizes, the associations between the studied variables are often found to have high statistical significance. Thus, the effect sizes of the associations should be emphasized when drawing conclusions from the results. In addition, the potential causal relationships should not be over-interpreted due to the observational nature of the data, and mechanisms for any potential causal relationships can only be hypothesized. Nevertheless, RCTs with large sample populations are not usually feasible and thus, observational studies showing real-world evidence of various associations may be regarded as a valuable data source.

In this thesis, the results obtained with the real-world health monitoring data were in line with previous studies, most of which were controlled research studies. For example, the fact that alcohol intake was dose-dependently associated with diminished parasympathetic regulation of the ANS during sleep confirmed the results of a previous controlled research study in real-world settings (Sagawa et al., 2011). In addition, the results between the different studies described in this thesis are consistent with each other, even though the studies employed different study designs and methods for the quantification and analysis of the variables. For example, the associations between acute alcohol intake and the ANS regulation during sleep were similar in both cross-sectional and repeated-measures study designs, and by analyzing the novel HRV-based recovery indices as well as traditional HR and HRV parameters.

The large sample of RWD also enabled the exploration of previously unstudied aspects of sleep recovery, which generated data-driven hypotheses for further studies. For example, alcohol intake seemed to affect the ANS regulation during sleep more for younger than for older subjects, and regular PA did not seem to protect the subjects from the adverse effects of alcohol intake on the ANS regulation during sleep. These issues have not been studied in previous controlled studies due to the fact that the sample populations are usually heterogeneous and small (Romanowicz et al., 2011; Sagawa et al., 2011; Ralevski, Petrakis and Altemus, 2018). Thus, these hypotheses require further study with controlled settings.

6.2 Strengths and limitations

Both the strengths and limitations of the research arise, primarily, from the large-scale observational real-world dataset used in the analysis. It should be emphasized that the real-world dataset used in this research is globally unique, containing as it does a blend of continuous 72-hour physiological measurement and self-reported data collected in real-life settings from a large and heterogeneous sample. Any use of this kind of real-world health monitoring data collected from a large number of subjects for scientific research is still very rare, which also puts emphasis on the importance of the insights into health and well-being that can be generated based on the results presented in this thesis.

Giving that the analyzed dataset contained real-life recordings from a large and heterogeneous sample of Finnish employees makes the results of the thesis especially useful for the understanding of employees' health and well-being in Finland or in the Nordics. The sample of the subjects in the dataset is, however, not random and the measurements were voluntary. Thus, Finnish workers who were not interested in their wellbeing or health may be slightly under-represented in the data. In addition, the participants all appeared to be healthy, as the test protocol stipulated that the measurements were not to be performed if the subject had some chronic or acute disease. Another possible weakness is that although both manual and non-manual workers are represented in the dataset, the subjects' professions or the socioeconomic statuses were not recorded.

The self-reported data used in the analysis included, for example, the sleep times, alcohol intake and background characteristics. When BMI is determined using self-reported weight and height, the BMI is typically underestimated (Gorber et al., 2007). However, the BMI of the participants seemed to be in line with population-based estimates (Vartiainen et al., 2009). As in general with self-reported data, the number of alcohol doses, especially for high alcohol doses, may be underestimated and inaccurately reported. To counter this, very high reported alcohol intakes were excluded from the analysis. In addition, self-reported sleep times may have inaccuracies, and thus a 30-minute time window for falling asleep was used in these analyses.

The duration of the measurements were limited to three days per subject and self-reported data was not exhaustive due to the feasibility of the data collection. The measurement period was limited to three days per subject for practical reasons, such as the battery lifetime of the wearable device and the skin irritation the electrodes may cause when used for extended time periods. Limiting the recordings for three days may have caused inaccuracies when estimating the typical PA behavior or ANS regulation of the subjects. Moreover, the limited recordings may compromise the generalizability of the results, as the normal 'true' behaviors may not be captured with this short recording. On the other hand, significantly longer measurements, e.g.

three weeks, had probably decreased the subjects adherence to self-report precisely their daily activities including working and sleeping times as well as alcohol intake. In addition, longer measurements may have compromised the feasibility of the data analysis. Due to the limited self-reported data, habits of smoking, drinking or caffeine consumption as well as information about the timing of alcohol intake were not reported and thus, could not be addressed in the analysis. As these variables have been reported to affect HRV measurements, their unavailability is a limitation for the results of this thesis.

The physiological measurement employed in the data collection was the three-day beat-to-beat RR-interval recording performed in uncontrolled free-living settings where artefacts in the RR-interval recording are unavoidable. Thus, the recorded RR-intervals were first run through the proprietary artefact correction algorithm that has shown to be powerful also in free-living settings (Saalasti, Seppänen and Kuusela, 2004; Parak and Korhonen, 2013). Thereafter, from artefact-corrected RR-intervals the PA, stress and recovery parameters were extracted using also proprietary algorithms (Kettunen and Saalasti, 2005a; Kettunen and Saalasti 2005b). The analyses of PA were based on the estimated VO_2 from the beat-to-beat RR-interval recordings and the subject's background characteristics, so no direct measurement for VO_2 was available. The proprietary VO_2 estimation method used in this study has been shown to be accurate enough for field studies, although none of the validation studies for the method dealt with a population representative of the subjects in this study (Smolander et al., 2011; Robertson et al., 2015). As with the VO_2 estimation, the detection of recovery and stress reactions and their intensity was based on a proprietary algorithm. Previous studies using this algorithm have shown that stress and recovery are associated with subjective stress, as well as with morning cortisol levels (Kinnunen et al, 2006; Rusko et al., 2006; Uusitalo et al., 2011; Teisala et al., 2014; Föhr et al., 2015). However, a stressful event detected with the proprietary algorithm may not agree with the subject's personal view of the situation (Oksman, Ermes and Tikkamäki, 2016; Kaikkonen, Lindholm and Lusa, 2017). In this thesis, the ANS regulation during sleep was quantified with the traditional HR and HRV parameters as well as with the novel recovery parameters, and the results show that these recovery parameters are in accordance with the HR and HRV parameters. However, given that the artefacts in the wearable HRV monitoring are unavoidable and cannot be perfectly corrected with any artefact-correction algorithm, the HR and HRV parameters generated from RR-intervals recorded in real-life settings should be considered as practical real-life estimates instead of exact clinical values of HR and HRV.

The measurements were conducted in uncontrolled free-living settings, which from study design perspective inevitably means that there will be a number of confounding factors in the dataset. Statistical methods were used to control for any such confounding factors in the cross-sectional study designs, and the within-subject study design described in Publication IV also provided control for any unknown confounding factors. Statistical methods were also employed to compensate for the non-balanced sampling of the dataset. However, the use of statistical methods to compensate for the confounding factors and for the non-balanced sampling may not be ideal, so the results should be interpreted with caution. The results of this thesis could not be validated using the traditional scientific approach due to the limitations of the available reference data. Collecting reference data would have required conducting RCTs with detailed data collection and long-term physiological measurements on a large number of subjects, which would not have been feasible.

6.3 Contribution to science and practice

This thesis focused on analyses of a real-world health monitoring dataset, and demonstrated that the results obtained from controlled research studies also apply in uncontrolled, free-living settings. Furthermore, the large-scale dataset enabled the study of aspects of PA and sleep behaviors that could not have been studied in previous controlled research studies due to their small sample size. Thus, the results of this thesis can be seen as an example of the complementary roles that RWD and RCT studies have in providing further insights into health and well-being.

Those associations found in the RWD studies that have not previously been the topic of clinical studies can be used to generate data-driven hypotheses for future RCTs. For example, the results of this thesis have generated data-driven hypotheses related to PA and ANS regulation during sleep. PA during the day may actually compromise recovery during the following night's sleep. However, PA does have a beneficial effect on recovery during sleep, but only in the longer term. The benefits of PA can eventually be seen in the form of increased levels of physical fitness, which aids recovery during sleep. Regarding ANS regulation during sleep, alcohol intake seemed to have a stronger effect on the younger than older subjects, but a similar effect in male and female subjects and in sedentary and physically active subjects. These interactions between alcohol intake and the subject's background characteristics should be further studied with RCTs on large and heterogeneous samples (Romanowicz et al., 2011; Ralevski, Petrakis and Altemus, 2018). Regarding PA quantification, the results of this large-scale RWD study have also thrown up an interesting point about the way the subjects' PA can be quantified. The results of this study indicate that if the PA estimation takes into account the subject's current level of physical fitness, this may be a better way of motivating low-fit individuals to increase their PA. This data-driven hypothesis should also be confirmed in further research studies.

With regard to interpreting the results of research studies in general, this thesis highlights the significant effect that the methods and context of the studies can have, and these should always be taken into account. For example, in PA research studies, the PA quantification method may strongly affect the estimates of the subjects' PA and thus treat differently high-fitness and low-fitness individuals. Furthermore, in HRV research studies it is important to consider the context of the HRV measurement, as ANS regulation seems to be affected by the daily activities and behaviors.

In practice, a feasible and relevant PA quantification method and a good knowledge of the temporal patterns in PA behavior can make PA counselling and interventions more personal, and therefore more effective. When trying to promote behavioral changes in PA, the most motivating PA quantification method should be selected. For example, the information about the varying temporal patterns of PA can be utilized, perhaps by trying to increase the amount and intensity of the subject's PA on those days when they have naturally higher PA levels, or vice versa, by engaging and promoting activities with PA on the days the subjects naturally have low PA levels.

In addition, this thesis has demonstrated some of the effects that lifestyles and daily activities can have on ANS regulation during sleep. This knowledge may be used to motivate individuals to be more aware of the effect their behaviors can have on their well-being. This is the first step towards changing one's behavior and lifestyle. Regarding the effect that alcohol intake has on the ANS regulation during sleep, the results from the novel indices used in this RWD study agree with the HR and HRV parameters regarding the amount and intensity of recovery, once they have been normalized against the subject's background characteristics. For example, the recovery indices showed decreased values after only a low alcohol intake, and this information could be used to inform and convince people that even a low alcohol intake affects sleep. In other words, the recovery indices could be used as practical tools for the general public to visualize and demonstrate the effects that daily activities and behaviors can have on their sleep.

As a conclusion, the results of this thesis can benefit both scientific research as well as practice in the field of health and well-being. The results of this thesis comprehensively focused on studying PA and sleep behaviors using a real-world HRV dataset, but this rich dataset also enables studying many further scientific studies. Specifically, this dataset could be utilized to provide further real-world evidence for health and well-being about, for example, sedentary behavior and occupational health. In larger perspective, this thesis illustrates how real-world wearable health monitoring data can be put in good use to facilitate scientific results and provide practical insights and hopefully, inspires wider use of real-world health monitoring data for research.

7 SUMMARY AND CONCLUSIONS

The thesis shows the feasibility to use real-world continuous wearable health monitoring data for research. Data preprocessing, appropriate methodologies for quantifying health behaviors and well-being as well as statistical methods for the analysis are the key to provide valid observations from the real-world data. The results of this thesis, specifically, give unique observational real-world evidence about the physical activity behaviors and the associations of daily activities with sleep in Finnish employees. The amount of large-scale real-world health monitoring data is globally constantly increasing and its analysis opens up new possibilities for research as well as enables studying aspects of health behaviors and well-being that cannot be investigated in traditional research studies. On the contrary, real-world health data can also facilitates to generate valuable data-driven hypotheses to be further investigated in controlled research studies. For practice, the insights from the real-world data are essential in promoting people's health and well-being with more personalized and targeted health interventions and tools.

REFERENCES

ABHISHEKH, H.A., NISARGA, P., KISAN, R., MEGHANA, A., CHANDRAN, S., RAJU, T. and SATHYAPRABHA, T.N., 2013. Influence of age and gender on autonomic regulation of heart. *Journal of clinical monitoring and computing,* **27**(3), pp. 259-264.

ACHTEN, J. and JEUKENDRUP, A.E., 2003. Heart rate monitoring. *Sports medicine,* **33**(7), pp. 517-538.

AINSLIE, P.N., REILLY, T. and WESTERTERP, K.R., 2003. Estimating Human Energy Expenditure: A Review of Techniques with Particular Reference to Doubly Labelled Water. *Sports Medicine*, **33**(9), pp. 683-698.

ALLEN, J., 2007. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement,* **28**(3), pp. R1-R39.

ALTINI, M., PENDERS, J., VULLERS, R. and AMFT, O., 2013. Combining wearable accelerometer and physiological data for activity and energy expenditure estimation, *Proceedings of the 4th Conference on Wireless Health* 2013, ACM, pp. 1.

ALTHOFF, T., SOSIC, R., HICKS, J., KING, A., DELP, S. and LESKOVEC, J., 2017. Large-scale physical activity data reveal worldwide activity inequality. *Nature,* **547**(7663), pp. 336-336.

AN, H., JONES, G.C., KANG, S., WELK, G.J. and LEE, J., 2017. How valid are wearable physical activity trackers for measuring steps? *European Journal of Sport Science,* **17**(3), pp. 360-368.

AUSTIN, P.C., ESCOBAR, M. and KOPEC, J.A., 2000. The Use of the Tobit Model for Analyzing Measures of Health Status. *Quality of Life Research,* **9**(8), pp. 901-910.

BARUTCU, I., ESEN, A.M., KAYA, D., TURKMEN, M., KARAKAYA, O., MELEK, M., ESEN, O.B. and BASARAN, Y., 2005. Cigarette smoking and heart rate variability: dynamic influence of parasympathetic and sympathetic maneuvers. *Annals of noninvasive electrocardiology,* **10**(3), pp. 324-329.

BELTRAME, T., AMELARD, R., WONG, A. and HUGHSON, R., 2017. Prediction of oxygen uptake dynamics by machine learning analysis of wearable sensors during activities of daily living. *Scientific reports,* **7**, pp. 45738.

BENICHOU, T., PEREIRA, B., MERMILLOD, M., TAUVERON, I., PFABIGAN, D., MAQDASY, S. and DUTHEIL, F., 2018. Heart rate variability in type 2 diabetes mellitus: A systematic review and meta–analysis. *PloS one,* **13**(4), pp. e0195166.

BERGER, M.L., MAMDANI, M., ATKINS, D. and JOHNSON, M.L., 2009. Good Research Practices for Comparative Effectiveness Research: Defining, Reporting and Interpreting Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part I. *Value in Health*, **12**(8), pp. 1044-1052.

BERGER, M.L., SOX, H., WILLKE, R.J., BRIXNER, D.L., EICHLER, H., GOETTSCH, W., MADIGAN, D., MAKADY, A., SCHNEEWEISS, S., TARRICONE, R., WANG, S.V., WATKINS, J. and DANIEL MULLINS, C., 2017. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiology and drug safety,* **26**(9), pp. 1033-1039.

BERNTSON, G.G., CACIOPPO, J.T. and QUIGLEY, K.S., 1991. Autonomic determinism: the modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological review,* **98**(4), pp. 459-487.

BERNTSON, G.G. and CACIOPPO, J.T., 2007. Heart Rate Variability: Stress and Psychiatric Conditions. In: M. MALIK and A.J. CAMM, eds, *Dynamic Electrocardiography.*

BIGGER, J.T.,JR, FLEISS, J.L., STEINMAN, R.C., ROLNITZKY, L.M., KLEIGER, R.E. and ROTTMAN, J.N., 1992. Correlations among time and frequency domain measures of heart period variability two weeks after acute myocardial infarction. *The American Journal of Cardiology,* **69**(9), pp. 891-898.

BIKKINA, M., LARSON, M.G. and LEVY, D., 1992. Prognostic implications of asymptomatic ventricular arrhythmias: the Framingham Heart Study. *Annals of Internal Medicine,* **117**(12), pp. 990-996.

BILLMAN, G.E., 2013. The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Frontiers in physiology,* **4**:26, doi: 10.3389/fphys.2013.00026.

BILLMAN, G.E., 2011. Heart rate variability–a historical perspective. *Frontiers in physiology,* **2**:86, doi: 10.3389/fphys.2011.00086.

BILLMAN, G.E., HUIKURI, H.V., SACHA, J. and TRIMMEL, K., 2015. An introduction to heart rate variability: methodological considerations and clinical applications. *Frontiers in physiology,* **6**:55, doi: 10.3389/fphys.2015.00055.

BONNEMEIER, H., WIEGAND, U.K., BRANDES, A., KLUGE, N., KATUS, H.A., RICHARDT, G. and POTRATZ, J., 2003. Circadian profile of cardiac autonomic nervous modulation in healthy subjects: differing effects of aging and gender on heart rate variability. *Journal of cardiovascular electrophysiology,* **14**(8), pp. 791-799.

BOOTH, C.M. and TANNOCK, I.F., 2014. Randomised controlled trials and population-based observational research: Partners in the evolution of medical evidence. *British journal of cancer,* **110**(3), pp. 551-555.

BORRESEN, J. and LAMBERT, M.I., 2008. Autonomic control of heart rate during and after exercise. *Sports medicine,* **38**(8), pp. 633-646.

BRAGE, S., WEDDERKOPP, N., FRANKS, P.W., BO ANDERSEN, L. and FROBERG, K., 2003. Reexamination of validity and reliability of the CSA monitor in walking and running. *Medicine and Science in Sports and Exercise,* **35**(8), pp. 1447-1454.

BRANDENBERGER, G., BUCHHEIT, M., EHRHART, J., SIMON, C. and PIQUARD, F., 2005. Is slow wave sleep an appropriate recording condition for heart rate variability analysis? *Autonomic Neuroscience,* **121**(1-2), pp. 81-86.

BRAVI, A., GREEN, G., HERRY, C., WRIGHT, H.E., LONGTIN, A., KENNY, G.P. and SEELY, A.J., 2013. Do physiological and pathological stresses produce different changes in heart rate variability? *Frontiers in physiology,* **4**:197, doi: 10.3389/fphys.2013.00197.

BROOK, R.D. and JULIUS, S., 2000. Autonomic imbalance, hypertension, and cardiovascular risk. *American journal of hypertension,* **13**(S4), pp. 112S-122S.

BROWN, T.E., BEIGHTOL, L.A., KOH, J. and ECKBERG, D.L., 1993. Important influence of respiration on human R-R interval power spectra is largely ignored. *Journal of applied physiology (Bethesda, Md.: 1985),* **75**(5), pp. 2310-2317.

BUCCELLETTI, E., GILARDI, E., SCAINI, E., GALIUTO, L., PERSIANI, R., BIONDI, A., BASILE, F. and SILVERI, N.G., 2009. Heart rate variability and myocardial infarction: systematic literature review and metanalysis. *Eur Rev Med Pharmacol Sci,* **13**(4), pp. 299-307.

BUCHHEIT, M. and GINDRE, C., 2006. Cardiac parasympathetic regulation: respective associations with cardiorespiratory fitness and training load. *American Journal of Physiology-Heart and Circulatory Physiology,* **291**(1), pp. H451-H458.

BUCHHEIT, M., SIMON, C., CHARLOUX, A., DOUTRELEAU, S., PIQUARD, F. and BRANDENBERGER, G., 2005. Heart rate variability and intensity of habitual physical activity in middle-aged persons. *Medicine & Science in Sports & Exercise,* **37**(9), pp. 1530-1534.

BUŠEK, P., VAŇKOVÁ, J., OPAVSKÝ, J., SALINGER, J. and NEVŠÍMALOVÁ, S., 2005. Spectral analysis of heart rate variability in sleep. *Physiol res,* **54**(4), pp. 369-376.

BYRNE, N.M., HILLS, A.P., HUNTER, G.R., WEINSIER, R.L. and SCHUTZ, Y., 2005. Metabolic equivalent: one size does not fit all. *Journal of applied physiology,* **99**(3), pp. 1112-1119.

BÄR, K., BOETTGER, M.K., KOSCHKE, M., SCHULZ, S., CHOKKA, P., YERAGANI, V.K. and VOSS, A., 2007. Non-linear complexity measures of heart rate variability in acute schizophrenia. *Clinical neurophysiology,* **118**(9), pp. 2009-2015.

CAMPOS, I.D., PINTO, V., SOUSA, N. and PEREIRA, V.H., 2018. A brain within the heart: A review on the intracardiac nervous system. *Journal of Molecular and Cellular Cardiology,* **119**, pp. 1-9.

CASE, M.A., BURWICK, H.A., VOLPP, K.G. and PATEL, M.S., 2015. Accuracy of Smartphone Applications and Wearable Devices for Tracking Physical Activity Data. *JAMA,* **313**(6), pp. 625-626.

CASPERSEN, C.J., POWELL, K.E. and CHRISTENSON, G.M., 1985. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public health reports (Washington, D.C.: 1974),* **100**(2), pp. 126-131.

CHALMERS, J.A., QUINTANA, D.S., ABBOTT, M.J. and KEMP, A.H., 2014. Anxiety disorders are associated with reduced heart rate variability: a meta-analysis. *Frontiers in psychiatry,* **5**:80, doi: 10.3389/fpsyt.2014.00080.

CHOI, A. and SHIN, H., 2018. Quantitative analysis of the effect of an ectopic beat on the heart rate variability in the resting condition. *Frontiers in physiology,* **9**:922, doi: 10.3389/fphys.2018.00922.

COHEN, M.A. and TAYLOR, J.A., 2002. Short-term cardiovascular oscillations in man: measuring and modelling the physiologies. *The Journal of physiology,* **542**(3), pp. 669-683.

COLLEY, R.C., GARRIGUET, D., JANSSEN, I., CRAIG, C.L., CLARKE, J. and TREMBLAY, M.S., 2011. Physical activity of Canadian adults: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. *Health reports / Statistics Canada,* **22**(1), pp. 7-14.

COLTEN, H.R. and ALTEVOGT, B.M., 2006. Sleep physiology. In: H.R. COLTEN and B.M. ALTEVOGT, eds, *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. National Academies Press (US).

COX, E., PHD, MARTIN, BRADLEY C., PHARMD, PHD, VAN STAA, TJEERD, PHD, MD, MSC, MA, GARBE, EDELTRAUT, MD, PHD, SIEBERT, UWE, MD, MPH, MSC, SCD and JOHNSON, M.L., PHD, 2009. Good Research Practices for Comparative Effectiveness Research: Approaches to Mitigate Bias and Confounding in the Design of Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. *Value in Health,* **12**(8), pp. 1053-1061.

CULLINAN, W.E., 2009. Stress Response. In: M.D. BINDER, N. HIROKAWA and U. WINDHORST, eds, *Encyclopedia of Neuroscience*. Springer, pp. 3861-3865.

DE ALMEIDA MENDES, M., DA SILVA, I.C., RAMIRES, V.V., REICHERT, F.F., MARTINS, R.C. and TOMASI, E., 2018. Calibration of raw accelerometer data to measure physical activity: A systematic review. *Gait & posture,* **61**, pp. 98-110.

DE LIMA, A., HAHN, T., EVERS, L., DE VRIES, N., COHEN, E., AFEK, M., BATAILLE, L., DAESCHLER, M., CLAES, K., BOROOJERDI, B., TERRICABRAS, D., LITTLE, M., BALDUS, H., BLOEM, B. and FABER, M., 2017. Feasibility of large-scale deployment of multiple wearable sensors in Parkinson's disease. *PLOS ONE,* **12**(12), pp. e0189161.

DENVER, J.W., REED, S.F. and PORGES, S.W., 2007. Methodological issues in the quantification of respiratory sinus arrhythmia. *Biological psychology,* **74**(2), pp. 286-294.

DIMITROPOULOS, G., TAHRANI, A.A. and STEVENS, M.J., 2014. Cardiac autonomic neuropathy in patients with diabetes mellitus. *World journal of diabetes,* **5**(1), pp. 17-39.

DRAGHICI, A.E. and TAYLOR, J.A., 2016. The physiological basis and measurement of heart rate variability in humans. *Journal of physiological anthropology,* **35**:22, doi: 10.1186/s40101-016-0113-7.

DREYER, N.A. and GARNER, S., 2009. Registries for Robust Evidence. *JAMA,* **302**(7), pp. 790-791.

ERNST, G., 2017. Heart-Rate Variability-More than Heart Beats? *Frontiers in public health,* **5**:240, doi: 10.3389/fpubh.2017.00240.

EWING, D.J., BORSEY, D.Q., TRAVIS, P., BELLAVERE, F., NEILSON, J.M. and CLARKE, B.F., 1983. Abnormalities of ambulatory 24-hour heart rate in diabetes mellitus. *Diabetes,* **32**(2), pp. 101-105.

FAGARD, R.H., PARDAENS, K. and STAESSEN, J.A., 1999. Influence of demographic, anthropometric and lifestyle characteristics on heart rate and its variability in the population. *Journal of hypertension,* **17**(11), pp. 1589-1599.

FAGARD, R.H., 2001. A population-based study on the determinants of heart rate and heart rate variability in the frequency domain. *Verhandelingen - Koninklijke Academie voor Geneeskunde van Belgie,* **63**(1), pp. 57-89; discussion 90-1.

FAGHERAZZI, G., EL FATOUHI, D., BELLICHA, A., EL GAREH, A., AFFRET, A., DOW, C., DELRIEU, L., VEGREVILLE, M., NORMAND, A., OPPERT, J. and SEVERI, G., 2017. An International Study on the Determinants of Poor Sleep Amongst 15,000 Users of Connected Devices. *Journal of Medical Internet Research,* **19**(10), pp. e363.

FIRSTBEAT TECHNOLOGIES OY, 2014. Stress and Recovery Analysis Method Based on 24-hour Heart Rate Variability. *Firstbeat Technologies Oy, web page.* Available (accessed 04.06.2019): https://assets.firstbeat.com/firstbeat/uploads/2015/11/Stress-and-recovery_white-paper_20145.pdf. Archived by WebCite® at http://www.webcitation.org/78sqh6WOP.

FIRSTBEAT TECHNOLOGIES, OY, 2005. VO_2 Estimation Method Based on Heart Rate Measurement. *Firstbeat Technologies Oy, web page.* Available (accessed 04.06.2019): https://assets.firstbeat.com/firstbeat/uploads/2015/11/white_paper_vo2_estimation.pdf. Archived by WebCite® at http://www.webcitation.org/78ss2yplL.

FORD, E.S., CUNNINGHAM, T.J., GILES, W.H. and CROFT, J.B., 2015. Trends in insomnia and excessive daytime sleepiness among US adults from 2002 to 2012. *Sleep medicine,* **16**(3), pp. 372-378.

FRIEDMAN, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics,* **29**(5), pp. 1189-1232.

FURNESS, J.B., 2009. Autonomic Nervous System. In: L.R. SQUIRE, ed, *Encyclopedia of Neuroscience.* Oxford: Academic Press, pp. 833-835.

FÖHR, T., TOLVANEN, A., MYLLYMÄKI, T., JÄRVELÄ-REIJONEN, E., RANTALA, S., KORPELA, R., PEUHKURI, K., KOLEHMAINEN, M., PUTTONEN, S. and LAPPALAINEN, R., 2015. Subjective stress, objective heart rate variability-based stress, and recovery on workdays among overweight and psychologically distressed individuals: a cross-sectional study. *Journal of Occupational Medicine and Toxicology*, **10**:39, doi: 10.1186/s12995-015-0081-6.

GABELLA, G., 2012. Autonomic nervous system. In: *Encyclopedia of Life Sciences (Ed.).* Wiley Online Library, doi:10.1002/9780470015902.a0000081.pub2.

GARBER, C.E., BLISSMER, B., DESCHENES, M.R., FRANKLIN, B.A., LAMONTE, M.J., LEE, I., NIEMAN, D.C., SWAIN, D.P. and AMER COLL SPORTS MED, 2011. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Medicine and Science in Sports and Exercise,* **43**(7), pp. 1334-1359.

GARRISON, L.P., NEUMANN, P.J., ERICKSON, P., MARSHALL, D. and MULLINS, C.D., 2007. Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report. *Value in Health,* **10**(5), pp. 326-335.

GAY, V. and LEIJDEKKERS, P., 2015. Bringing health and fitness data together for connected health care: Mobile apps as enablers of interoperability. *Journal of Medical Internet Research,* **17**(11), pp. E260.

GENUER, R., POGGI, J. and TULEAU-MALOT, C., 2010. Variable selection using random forests. *Pattern Recognition Letters,* **31**(14), pp. 2225-2236.

GEORGIOU, K., LARENTZAKIS, A.V., KHAMIS, N.N., ALSUHAIBANI, G.I., ALASKA, Y.A. and GIALLAFOS, E.J., 2018. Can Wearable Devices Accurately Measure Heart Rate Variability? A Systematic Review. *Folia medica,* **60**(1), pp. 7-20.

GEURTS, S.A. and SONNENTAG, S., 2006. Recovery as an explanatory mechanism in the relation between acute stress reactions and chronic health impairment. *Scandinavian journal of work, environment & health,* **32**(6), pp. 482-492.

GOLDSMITH, R.L., BIGGER, J.T. JR, BLOOMFIELD, D.M. and STEINMAN, R.C., 1997. Physical fitness as a determinant of vagal modulation. *Medicine and Science in Sports and Exercise,* **29**(6), pp. 812-817.

GOLDSTEIN, D.S., BENTHO, O., PARK, M. and SHARABI, Y., 2011. Low-frequency power of heart rate variability is not a measure of cardiac sympathetic tone but may be a measure of modulation of cardiac autonomic outflows by baroreflexes. *Experimental physiology,* **96**(12), pp. 1255-1261.

GORBER, S.C., TREMBLAY, M., MOHER, D. and GORBER, B., 2007. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obesity reviews,* **8**(4), pp. 307-326.

GREENBERG, R.S. and KLEINBAUM, D.G., 1985. Mathematical modeling strategies for the analysis of epidemiologic research. *Annual Review of Public Health,* **6**, pp. 223-245.

GREENWELL, B.M., 2017. pdp: An R Package for Constructing Partial Dependence Plots. *R Journal,* **9**(1), pp. 421-436.

GREGORUTTI, B., MICHEL, B. and SAINT-PIERRE, P., 2017. Correlation and variable importance in random forests. *Statistics and Computing,* **27**(3), pp. 659-678.

GRETEBECK, R.J. and MONTOYE, H.J., 1992. Variability of some objective measures of physical activity. *Medicine and Science in Sports and Exercise,* **24**(10), pp. 1167-1172.

GUTHOLD, R., ONO, T., STRONG, K.L., CHATTERJI, S. and MORABIA, A., 2008. Worldwide Variability in Physical Inactivity. A 51-Country Survey. *American Journal of Preventive Medicine,* **34**(6), pp. 486-494.

HALL, M., VASKO, R., BUYSSE, D., OMBAO, H., CHEN, Q., CASHMERE, J.D., KUPFER, D. and THAYER, J.F., 2004. Acute Stress Affects Heart Rate Variability during Sleep. *Psychosomatic medicine,* **66**(1), pp. 56-62.

HAN, J., 2009. Data Mining. In: L. LIU and M.T. ÖZSU, eds, *Encyclopedia of Database Systems.* Boston: Springer, pp. 595-598.

HARRIS, J.A. and BENEDICT, F.G., 1918. A Biometric Study of Human Basal Metabolism. *Proceedings of the National Academy of Sciences of the United States of America,* **4**(12), pp. 370-373.

HASKELL, W.L., BLAIR, S.N. and HILL, J.O., 2009. Physical activity: health outcomes and importance for public health policy. *Preventive medicine,* **49**(4), pp. 280-282.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J., 2009. The elements of statistical learning: data mining, inference, and prediction. 2nd edn. *New York: Springer.*

HAUTALA, A.J., KIVINIEMI, A.M. and TULPPO, M.P., 2009. Individual responses to aerobic exercise: the role of the autonomic nervous system. *Neuroscience & Biobehavioral Reviews,* **33**(2), pp. 107-115.

HAYANO, J., SAKAKIBARA, Y., YAMADA, A., YAMADA, M., MUKAI, S., FUJINAMI, T., YOKOYAMA, K., WATANABE, Y. and TAKATA, K., 1991. Accuracy of assessment of cardiac vagal tone by heart rate variability in normal subjects. *The American Journal of Cardiology,* **67**(2), pp. 199-204.

HEATHERS, J.A., 2014. Everything Hertz: methodological issues in short-term frequency-domain HRV. *Frontiers in physiology,* **5**:177, doi: 10.3389/fphys.2014.00177.

HELANDER, E., VUORINEN, A., WANSINK, B. and KORHONEN, I., 2014. Are breaks in daily self-weighing associated with weight gain? *PloS one,* **9**(11), pp. e113164.

HELANDER, E., WANSINK, B. and CHIEH, A., 2016. Weight Gain over the Holidays in Three Countries. *NEW ENGLAND JOURNAL OF MEDICINE,* **375**(12), pp. 1200-1202.

HELLDÁN, A. and HELAKORPI, S., 2015. *Health Behavior and Health among the Finnish Adult Population, Spring 2014.* Tampere: National Institute for Health and Welfare.

HERNANDO, D., ROCA, S., SANCHO, J., ALESANCO, Á. and BAILÓN, R., 2018. Validation of the apple watch for heart rate variability measurements during relax and mental stress in healthy subjects. *Sensors,* **18**(8), pp. 2619.

HICKS, J.L., ALTHOFF, T., KUHAR, P., BOSTJANCIC, B., KING, A.C., LESKOVEC, J. and DELP, S.L., 2019. Best practices for analyzing large-scale health data from wearables and smartphone apps. *NPJ digital medicine,* **2**(1), pp. 1-12.

HILL, L., SIEBENBROCK, A., SOLLERS, J. and THAYER, J.F., 2009. Are all measures created equal? Heart rate variability and respiration. *Biomedical sciences instrumentation,* **45**, pp. 71-76.

HILLEBRAND, S., GAST, K.B., DE MUTSERT, R., SWENNE, C.A., JUKEMA, J.W., MIDDELDORP, S., ROSENDAAL, F.R. and DEKKERS, O.M., 2013. Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: meta-analysis and dose–response meta-regression. *Europace,* **15**(5), pp. 742-749.

HILLS, A.P., MOKHTAR, N. and BYRNE, N.M., 2014. Assessment of physical activity and energy expenditure: an overview of objective measures. *Frontiers in nutrition,* **1**:5, doi: 10.3389/fnut.2014.00005.

HINTERBERGER, H., 2009. Exploratory Data Analysis. In: L. LIU and M.T. ÖZSU, eds, *Encyclopedia of Database Systems.* Boston: Springer, pp. 1080-1080.

HOULE, M.S. and BILLMAN, G.E., 1999. Low-frequency component of the heart rate variability spectrum: a poor marker of sympathetic activity. *American Journal of Physiology-Heart and Circulatory Physiology,* **276**(1), pp. H215-H223.

HOWLEY, E.T., 2001. Type of activity: resistance, aerobic and leisure versus occupational physical activity. *Medicine & Science in Sports & Exercise,* **33**(6), pp. S364-S369.

HUIKURI, H.V., PERKIÖMÄKI, J.S., MAESTRI, R. and PINNA, G.D., 2009. Clinical impact of evaluation of cardiovascular control by novel methods of heart rate dynamics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* **367**(1892), pp. 1223-1238.

HUIKURI, H.V. and STEIN, P.K., 2012. Clinical application of heart rate variability after acute myocardial infarction. *Frontiers in physiology,* **3**:41, doi: 10.3389/fphys.2012.00041.

HUSU, P., SUNI, J., VÄHÄ-YPYÄ, H., SIEVÄNEN, H., TOKOLA, K., VALKEINEN, H., MÄKI-OPAS, T. and VASANKARI, T., 2016. Objectively measured sedentary behavior and physical activity in a sample of Finnish adults: a cross-sectional study. *BMC public health,* **16**(1), pp. 920.

HYNYNEN, E., VESTERINEN, V., RUSKO, H. and NUMMELA, A., 2010. Effects of moderate and heavy endurance exercise on nocturnal HRV. *International Journal of Sports Medicine,* **31**(06), pp. 428-432.

IRWIN, M.R., VALLADARES, E.M., MOTIVALA, S., THAYER, J.F. and EHLERS, C.L., 2006. Association between nocturnal vagal tone and sleep depth, sleep quality, and fatigue in alcohol dependence. *Psychosomatic medicine,* **68**(1), pp. 159-166.

JACKSON, A.S., BLAIR, S.N., MAHAR, M.T., WIER, L.T., ROSS, R.M. and STUTEVILLE, J.E., 1990. Prediction of functional aerobic capacity without exercise testing. *Medicine and Science in Sports and Exercise,* **22**(6), pp. 863-870.

JAESCHKE, L., STEINBRECHER, A., JERAN, S., KONIGORSKI, S. and PISCHON, T., 2018. Variability and reliability study of overall physical activity and activity intensity levels using 24 h-accelerometry-assessed data. *BMC public health,* **18**(1), pp. 530.

JARCZOK, M., JARCZOK, M.N., MAUSS, D., KOENIG, J., LI, J., HERR, R.M. and THAYER, J.F., 2013. Autonomic nervous system activity and workplace stressors—A systematic review. *Neuroscience and biobehavioral reviews,* **37**(8), pp. 1810-1823.

JOHNSON, M.L., CROWN, W., MARTIN, B.C., DORMUTH, C.R. and SIEBERT, U., 2009. Good Research Practices for Comparative Effectiveness Research: Analytic Methods to Improve Causal Inference from Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III. *Value in Health,* **12**(8), pp. 1062-1073.

JONES, B.E., 2009. Sleep-Wake Autonomic Regulation. In: M.D. BINDER, N. HIROKAWA and U. WINDHORST, eds, *Encyclopedia of Neuroscience.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 3732-3736.

KAIKKONEN, P., LINDHOLM, H. and LUSA, S., 2017. Physiological Load and Psychological Stress During a 24-hour Work Shift Among Finnish Firefighters. *Journal of occupational and environmental medicine,* **59**(1), pp. 41-46.

KARAKAYA, O., BARUTCU, I., KAYA, D., ESEN, A.M., SAGLAM, M., MELEK, M., ONRAT, E., TURKMEN, M., ESEN, O.B. and KAYMAZ, C., 2007. Acute effect of cigarette smoking on heart rate variability. *Angiology,* **58**(5), pp. 620-624.

KAREMAKER, J.M., 2017. An introduction into autonomic nervous function. *Physiological Measurement,* **38**(5), pp. R89-R118.

KARLSSON, M., HÖRNSTEN, R., RYDBERG, A. and WIKLUND, U., 2012. Automatic filtering of outliers in RR intervals before analysis of heart rate variability in Holter recordings: a comparison with carefully edited data. *Biomedical engineering online,* **11**:2, doi:10.1186/1475-925X-11-2.

KEMP, A.H., QUINTANA, D.S., GRAY, M.A., FELMINGHAM, K.L., BROWN, K. and GATT, J.M., 2010. Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis. *Biological psychiatry,* **67**(11), pp. 1067-1074.

KEMP, A.H., QUINTANA, D.S., QUINN, C.R., HOPKINSON, P. and HARRIS, A.W., 2014. Major depressive disorder with melancholia displays robust alterations in resting state heart rate and its variability: implications for future morbidity and mortality. *Frontiers in psychology,* **5**:1387, doi:10.1016/j.biopsych.2009.12.012.

KETTUNEN, J. and SAALASTI, S., 2005A. *Procedure for deriving reliable information on respiratory activity from heart period measurement.* A61B 5/02.

KETTUNEN, J. and SAALASTI, S., 2005B. *Procedure for detection of stress by segmentation and analyzing a heart beat signal.* A61B 5/0402.

KEYTEL, L., GOEDECKE, J., NOAKES, T., HIILOSKORPI, H., LAUKKANEN, R., VAN DER MERWE, L. and LAMBERT, E., 2005. Prediction of energy expenditure from heart rate monitoring during submaximal exercise. *Journal of sports sciences,* **23**(3), pp. 289-297.

KHOSLA, S., WHITE, R., MEDINA, J., OUWENS, M., EMMAS, C., KODER, T., MALE, G. and LEONARD, S., 2018. Real world evidence (RWE) – a disruptive innovation or the quiet evolution of medical evidence generation? *F1000Research,* **7**:111, doi: 10.12688/f1000research.13585.1.

KIM, H., LEE, S. and KIM, J.H., 2018. Real-world evidence versus randomized controlled trial: clinical research based on electronic medical records. Journal of Korean medical science, 33(34), e213.

KINNUNEN, M., RUSKO, H., FELDT, T., KINNUNEN, U., JUUTI, T., MYLLYMÄKI, T., LAINE, K., HAKKARAINEN, P. and LOUHEVAARA, V., 2006. Stress and relaxation based on heart rate variability: Associations with self-reported mental strain and differences between waking hours and sleep. In: K.L. SAARELA, C. NYGÅRD and S. LUSA, eds, *Promotion of Well-Being in Modern Society.* Tampere, Finland: pp. 136-139.

KLEIGER, R.E., MILLER, J.P., BIGGER JR, J.T. and MOSS, A.J., 1987. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *The American Journal of Cardiology,* **59**(4), pp. 256-262.

KLEIGER, R.E., STEIN, P.K. and BIGGER, J.T. JR., 2005. Heart rate variability: measurement and clinical utility. *Annals of Noninvasive Electrocardiology,* **10**(1), pp. 88-101.

KOENIG, J., JARCZOK, M., WARTH, M., ELLIS, R., BACH, C., HILLECKE, T. and THAYER, J.F., 2014. Body mass index is related to autonomic nervous system activity as measured by heart rate variability—a replication using short term measurements. *The journal of nutrition, health & aging,* **18**(3), pp. 300-302.

KOENIG, J. and THAYER, J.F., 2016. Sex differences in healthy human heart rate variability: a meta-analysis. *Neuroscience & Biobehavioral Reviews,* **64**, pp. 288-310.

KOIZUMI, K., TERUI, N., KOLLAI, M. and BROOKS, C.M., 1982. Functional significance of coactivation of vagal and sympathetic cardiac nerves. *Proceedings of the National Academy of Sciences of the United States of America,* **79**(6), pp. 2116-2120.

KOIZUMI, K., 2009. Homeostasis. In: M.D. BINDER, N. HIROKAWA and U. WINDHORST, eds, *Encyclopedia of Neuroscience.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1860-1862.

KORHONEN, I., PÄRKKÄ, J. and VAN GILS, M., 2003. Health monitoring in the home of the future. *IEEE Engineering in medicine and biology magazine,* **22**(3), pp. 66-73.

KOZEY, S., LYDEN, K., STAUDENMAYER, J. and FREEDSON, P., 2010. Errors in MET estimates of physical activities using 3.5 ml.kg -1. min-1 as the baseline oxygen consumption. *Journal of Physical Activity and Health,* **7**(4), pp. 508-516.

KRONHOLM, E., PARTONEN, T., HÄRMÄ, M., HUBLIN, C., LALLUKKA, T., PELTONEN, M. and LAATIKAINEN, T., 2016. Prevalence of insomnia‐related symptoms continues to increase in the Finnish working‐age population. *Journal of sleep research,* **25**(4), pp. 454-457.

KUDAT, H., AKKAYA, V., SOZEN, A., SALMAN, S., DEMIREL, S., OZCAN, M., ATILGAN, D., YILMAZ, M. and GUVEN, O., 2006. Heart rate variability in diabetes patients. *Journal of international medical research,* **34**(3), pp. 291-296.

KWON, O., JEONG, J., KIM, H.B., KWON, I.H., PARK, S.Y., KIM, J.E. and CHOI, Y., 2018. Electrocardiogram Sampling Frequency Range Acceptable for Heart Rate Variability Analysis. *Healthcare informatics research,* **24**(3), pp. 198-206.

LABORDE, S., MOSLEY, E. and THAYER, J.F., 2017. Heart rate variability and cardiac vagal tone in psychophysiological research–recommendations for experiment planning, data analysis, and data reporting. *Frontiers in psychology,* **8**:213, doi: 10.3389/fpsyg.2017.00213.

LAHTI, J., SABIA, S., SINGH-MANOUX, A., KIVIMAKI, M., TATSUSE, T., YAMADA, M., SEKINE, M. and LALLUKKA, T., 2016. Leisure time physical activity and subsequent physical and mental health functioning among midlife Finnish, British and Japanese employees: a follow-up study in three occupational cohorts. *BMJ open,* **6**(1), pp. e009788-2015-009788.

LEE, I., SHIROMA, E.J., EVENSON, K.R., KAMADA, M., LACROIX, A.Z. and BURING, J.E., 2018. Accelerometer-Measured Physical Activity and Sedentary

Behavior in Relation to All-Cause Mortality: The Women's Health Study. *Circulation,* **137**(2), pp. 203-205.

LEICHT, A.S., SINCLAIR, W.H. and SPINKS, W.L., 2008. Effect of exercise mode on heart rate variability during steady state exercise. *European journal of applied physiology,* **102**(2), pp. 195-204.

LEVIN, S., JACOBS, D.R., AINSWORTH, B.E., RICHARDSON, M.T. and LEON, A.S., 1999. Intra-Individual Variation and Estimates of Usual Physical Activity. *Annals of Epidemiology,* **9**(8), pp. 481-488.

LIN, Y. and JIANG, R., 2013. Basic Statistics for Bioinformatics. In: R. JIANG, X. ZHANG and M. ZHANG, eds, *Basics of Bioinformatics.* Springer, pp. 27-68.

LIN, M., LUCAS, H.C. and SHMUELI, G., 2013. Too big to fail: Large samples and the p-value problem. *Information Systems Research,* **24**(4), pp. 906-917.

LINDHOLM, H., 2013. Physiological determinants and assessment of stress and recovery among media workers, *Helsingin yliopisto.*

LINDHOLM, H., SINISALO, J., AHLBERG, J., HIRVONEN, A., HUBLIN, C., PARTINEN, M. and SAVOLAINEN, A., 2012. Attenuation of vagal recovery during sleep and reduction of cortisol/melatonin ratio in late afternoon associate with prolonged daytime sleepiness among media workers with irregular shift work. *American Journal of Industrial Medicine,* **55**(7), pp. 643-649.

LU, K., YANG, L., SEOANE, F., ABTAHI, F., FORSMAN, M. and LINDECRANTZ, K., 2018. Fusion of Heart Rate, Respiration and Motion Measurements from a Wearable Sensor System to Enhance Energy Expenditure Estimation. *Sensors,* **18**(9), pp. 3092.

LUCINI, D., DI FEDE, G., PARATI, G. and PAGANI, M., 2005. Impact of Chronic Psychosocial Stress on Autonomic Cardiovascular Regulation in Otherwise Healthy Subjects. *Hypertension,* **46**(5), pp. 1201-1206.

LUYSTER, F.S., STROLLO, P.J., ZEE, P.C. and WALSH, J.K., 2012. Sleep: a health imperative. *Sleep,* **35**(6), pp. 727-734.

MADIGAN, D., RYAN, P.B., SCHUEMIE, M., STANG, P.E., OVERHAGE, J.M., HARTZEMA, A.G., SUCHARD, M.A., DUMOUCHEL, W. and BERLIN, J.A., 2013. Evaluating the impact of database heterogeneity on observational study results. *American Journal of Epidemiology,* **178**(4), pp. 645-651.

MAISSENHAELTER, B.E., WOOLMORE, A.L. and SCHLAG, P.M., 2018. Real-world evidence research based on big data. Der Onkologe, 24(2), pp. 91-98.

MAJUMDER, S., MONDAL, T. and DEEN, M.J., 2017. Wearable sensors for remote health monitoring. *Sensors (Switzerland),* **17**:130, doi:10.3390/s17010130.

MAKADY, A., DE BOER, A., HILLEGE, H., KLUNGEL, O. and GOETTSCH, W., 2017. What is real-world data? A review of definitions based on literature and stakeholder interviews. *Value in Health,* **20**(7), pp. 858-865.

MALIK, M., BIGGER, J.T., CAMM, A.J., KLEIGER, R.E., MALLIANI, A., MOSS, A.J. and SCHWARTZ, P.J., 1996. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European heart journal,* **17**(3), pp. 354-381.

MALLIANI, A. and MONTANO, N., 2002. Heart rate variability as a clinical tool. *Italian Heart Journal,* **3**, pp. 439-445.

MALLIANI, A., PAGANI, M., MONTANO, N. and MELA, G.S., 1998. Sympathovagal balance: a reappraisal. *Circulation,* **98**(23), pp. 2640-2643.

MARIN, T.S., KOURBELIS, C., FOOTE, J., NEWMAN, P., BROWN, A., DANIEL, M., COFFEE, N.T., NICHOLLS, S.J., GANESAN, A. and VERSACE, V.L., 2019. Examining adherence to activity monitoring devices to improve physical activity in adults with cardiovascular disease: A systematic review. *European journal of preventive cardiology,* **26**(4), pp. 382-397.

MARTINMÄKI, K., RUSKO, H., SAALASTI, S. and KETTUNEN, J., 2006. Ability of short-time Fourier transform method to detect transient changes in vagal effects on hearts: a pharmacological blocking study. *American journal of physiology. Heart and circulatory physiology,* 290(6), pp. H2582-9.

MATTHEWS, C.E., AINSWORTH, B.E., THOMPSON, R.W. and BASSETT, D.R., 2002. Sources of variance in daily physical activity levels as measured by an accelerometer. *Medicine and science in sports and exercise,* **34**(8), pp. 1376-1381.

MCCRATY, R. and SHAFFER, F., 2015. Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk. *Global advances in health and medicine,* **4**(1), pp. 46-61.

MELANSON, E.L., 2000. Resting heart rate variability in men varying in habitual physical activity. *Medicine and Science in Sports and Exercise,* **32**(11), pp. 1894-1901.

MELENOVSKY, V., SIMEK, J., SPERL, M., MALIK, J. and WICHTERLE, D., 2005. Relation between actual heart rate and autonomic effects of beta blockade in healthy men. *The American Journal of Cardiology,* **95**(8), pp. 999-1002.

MICHAEL, S., GRAHAM, K.S. and DAVIS, G.M., 2017. Cardiac autonomic responses during exercise and post-exercise recovery using heart rate variability and systolic time intervals—a review. *Frontiers in physiology,* **8**:301, doi: 10.3389/fphys.2017.00301.

MINAMI, J., ISHIMITSU, T. and MATSUOKA, H., 1999. Effects of Smoking Cessation on Blood Pressure and Heart Rate Variability in Habitual Smokers. *Hypertension,* **33**(1S Suppl), pp. 586-590.

MOLFINO, A., FIORENTINI, A., TUBANI, L., MARTUSCELLI, M., FANELLI, F.R. and LAVIANO, A., 2009. Body mass index is related to autonomic nervous system activity as measured by heart rate variability. *European journal of clinical nutrition,* **63**(10), pp. 1263-1265.

MONFREDI, O., LYASHKOV, A.E., JOHNSEN, A.B., INADA, S., SCHNEIDER, H., WANG, R., NIRMALAN, M., WISLOFF, U., MALTSEV, V.A., LAKATTA, E.G., ZHANG, H. and BOYETT, M.R., 2014. Biophysical characterization of the underappreciated and important relationship between heart rate variability and heart rate. *Hypertension (Dallas, Tex.: 1979),* **64**(6), pp. 1334-1343.

MONTANO, N., PORTA, A., COGLIATI, C., COSTANTINO, G., TOBALDINI, E., CASALI, K.R. and IELLAMO, F., 2009. Heart rate variability explored in the frequency domain: a tool to investigate the link between heart and behavior. *Neuroscience & Biobehavioral Reviews,* **33**(2), pp. 71-80.

MORENO, I.L., PASTRE, C.M., FERREIRA, C., DE ABREU, L.C., VALENTI, V.E. and VANDERLEI, L.C.M., 2013. Effects of an isotonic beverage on autonomic regulation during and after exercise. *Journal of the International Society of Sports Nutrition,* **10**:2, doi:10.1186/1550-2783-10-2.

MOTHERAL, B., BROOKS, J., CLARK, M.A., CROWN, W.H., DAVEY, P., HUTCHINS, D., MARTIN, B.C. and STANG, P., 2003. A Checklist for Retrospective Database Studies—Report of the ISPOR Task Force on Retrospective Databases. *Value in Health,* **6**(2), pp. 90-97.

MUTIKAINEN, S., HELANDER, E., PIETILÄ, J., KORHONEN, I. and KUJALA, U.M., 2014. Objectively measured physical activity in Finnish employees: a cross-sectional study. *BMJ open,* **4**(12), pp. e005927-2014-005927.

MYLLYMÄKI, T., KYRÖLÄINEN, H., SAVOLAINEN, K., HOKKA, L., JAKONEN, R., JUUTI, T., MARTINMÄKI, K., KAARTINEN, J., KINNUNEN, M. and RUSKO, H., 2011. Effects of vigorous late-night exercise on sleep quality and cardiac autonomic activity. *Journal of sleep research,* **20**(1pt2), pp. 146-153.

MYLLYMÄKI, T., RUSKO, H., SYVÄOJA, H., JUUTI, T., KINNUNEN, M. and KYRÖLÄINEN, H., 2012. Effects of exercise intensity and duration on nocturnal heart rate variability and sleep quality. *European journal of applied physiology,* **112**(3), pp. 801-809.

NANO, M., FONSECA, P., VULLINGS, R. and AARTS, R.M., 2017. Measures of cardiovascular autonomic activity in insomnia disorder: A systematic review. *PloS one,* **12**(10), pp. e0186716.

NG, M., FLEMING, T., ROBINSON, M., THOMSON, B., GRAETZ, N., MARGONO, C., MULLANY, E.C., BIRYUKOV, S., ABBAFATI, C. and ABERA, S.F., 2014. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet,* **384**(9945), pp. 766-781.

NISSINEN, S.I., MÄKIKALLIO, T.H., SEPPÄNEN, T., TAPANAINEN, J.M., SALO, M., TULPPO, M.P. and HUIKURI, H.V., 2003. Heart rate recovery after exercise as a predictor of mortality among survivors of acute myocardial infarction. *The American Journal of Cardiology,* **91**(6), pp. 711-714.

NUNAN, D., SANDERCOCK, G.R. and BRODIE, D.A., 2010. A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. *Pacing and clinical electrophysiology,* **33**(11), pp. 1407-1417.

OKSMAN, V., ERMES, M. and TIKKAMÄKI, K., 2016. Eustress-findings concerning the indication and interpretation of positive stress among entrepreneurs-a case study. *The Business & Management Review,* **7**(3), pp. 342-347.

ONGHENA, P. and VAN DER NOORTGATE, W., 2005. Confounding Variable. In: B. EVERITT and D. HOWELL, eds, *Encyclopedia of Statistics in Behavioral Science.* Chichester: John Wiley & Sons, Ltd, pp. 391-392.

OPTHOF, T., 2000. The normal range and determinants of the intrinsic heart rate in man. *Cardiovascular research,* **45**(1), pp. 177-184.

OSBORNE, J.W., 2010. Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research and Evaluation,* **15**(12), pp. 1-9.

OYINLOLA, J.O., CAMPBELL, J. and KOUSOULIS, A.A., 2016. Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidances. *BMC Health Services Research,* **16**:299, doi: 10.1186/s12913-016-1562-8.

PALLESEN, S., SIVERTSEN, B., NORDHUS, I.H. and BJORVATN, B., 2014. A 10-year trend of insomnia prevalence in the adult Norwegian population. *Sleep medicine,* **15**(2), pp. 173-179.

PAN, J. and TOMPKINS, W.J., 1985. A real-time QRS detection algorithm. *IEEE Trans.Biomed.Eng,* **32**(3), pp. 230-236.

PANG, D. and IGASAKI, T., 2018. A Combined Syntactical and Statistical Approach for R Peak Detection in Real-Time Long-Term Heart Rate Variability Analysis. Algorithms, **11**: 6, doi: 10.3390/a11060083.

PARAK, J. and KORHONEN, I., 2014. Evaluation of wearable consumer heart rate monitors based on photopletysmography, *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2014*, IEEE, pp. 3670-3673.

PARAK, J. and KORHONEN, I., 2013. Accuracy of Firstbeat Bodyguard 2 Beat-to-beat Heart Rate Monitor. *Firsbeat Technologies Oy, web page*. Available (accessed 04.06.2019): https://assets.firstbeat.com/firstbeat/uploads/2015/11/white_paper_bodyguard2_final.pdf. Archived by WebCite® at http://www.webcitation.org/78sqIv3H2.

PARAK, J., 2018. Evaluation of wearable optical heart rate monitoring sensors. *Tampere University of Technology*, **1580**.

PATEL, M.S., ASCH, D.A. and VOLPP, K.G., 2015. Wearable devices as facilitators, not drivers, of health behavior change. *JAMA,* **313**(5), pp. 459-460.

PATON, J., BOSCAN, P., PICKERING, A. and NALIVAIKO, E., 2005. The yin and yang of cardiac autonomic control: vago-sympathetic interactions revisited. *Brain Research Reviews*, 49(3), pp. 555-565.

PATTERSON, R., MCNAMARA, E., TAINIO, M., DE SÁ, T.H., SMITH, A.D., SHARP, S.J., EDWARDS, P., WOODCOCK, J., BRAGE, S. and WIJNDAELE, K., 2018. Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis. *European Journal of Epidemiology,* **33**(9), pp. 811-829.

PEAKE, J., KERR, G. and SULLIVAN, J., 2018. A Critical Review of Consumer Wearables, Mobile Applications, and Equipment for Providing Biofeedback, Monitoring Stress, and Sleep in Physically Active Populations. *Frontiers in physiology,* **9**:743, doi: 10.3389/fphys.2018.00743

PELTOLA, M., 2012. Role of editing of RR intervals in the analysis of heart rate variability. *Frontiers in physiology,* **3**:148, doi: 10.3389/fphys.2012.00148.

PELTOLA, M., 2010. Analysis of heart rate variability from 24-hour ambulatory electrocardiographic recordings. *Acta Universitatis Ouluensis D Medica,* **1087**.

PENTTILÄ, J., HELMINEN, A., JARTTI, T., KUUSELA, T., HUIKURI, H.V., TULPPO, M.P., COFFENG, R. and SCHEININ, H., 2001. Time domain, geometrical and frequency domain analysis of cardiac vagal outflow: effects of various respiratory patterns. *Clinical Physiology,* **21**(3), pp. 365-376.

PHYSICAL ACTIVITY GUIDELINES ADVISORY COMMITTEE, 2018. Physical activity guidelines advisory committee scientific report. *Washington, DC: US Department of Health and Human Services*.

PHYSICAL ACTIVITY GUIDELINES ADVISORY COMMITTEE, 2008. Physical activity guidelines advisory committee report, 2008. *Washington, DC: US Department of Health and Human Services,* **2008**, pp. A1-H14.

PIERCY, K.L., TROIANO, R.P., BALLARD, R.M., CARLSON, S.A., FULTON, J.E., GALUSKA, D.A., GEORGE, S.M. and OLSON, R.D., 2018. The physical activity guidelines for Americans. *JAMA,* **320**(19), pp. 2020-2028.

PIETILÄ, J., 2014. Preparation and Processing of a Large-Scale Heart Rate Variability and Lifestyle Assessment Results Data. *Tampere University of Technology.*

PIETILÄINEN, K.H., KAPRIO, J., BORG, P., PLASQUI, G., YKI-JÄRVINEN, H., KUJALA, U.M., ROSE, R.J., WESTERTERP, K.R. and RISSANEN, A., 2008. Physical inactivity and obesity: a vicious circle. *Obesity,* **16**(2), pp. 409-414.

PIVARNIK, J.M., REEVES, M.J. and RAFFERTY, A.P., 2003. Seasonal variation in adult leisure-time physical activity. *Medicine and Science in Sports and Exercise,* **35**(6), pp. 1004-1008.

PIWEK, L., ELLIS, D.A., ANDREWS, S. and JOINSON, A., 2016. The Rise of Consumer Health Wearables: Promises and Barriers. *PLoS Medicine,* **13**(2), pp. e1001953.

PORKKA-HEISKANEN, T., ZITTING, K. and WIGREN, H., 2013. Sleep, its regulation and possible mechanisms of sleep disturbances. *Acta physiologica,* **208**(4), pp. 311-328.

PRADHAPAN, P., TARVAINEN, M.P., NIEMINEN, T., LEHTINEN, R., NIKUS, K., LEHTIMÄKI, T., KÄHÖNEN, M. and VIIK, J., 2014. Effect of heart rate correction on pre-and post-exercise heart rate variability to predict risk of mortality—an experimental study on the FINCAVAS cohort. *Frontiers in physiology,* **5**:208, doi: 10.3389/fphys.2014.00208.

QUINTANA, D.S. and HEATHERS, J.A., 2014. Considerations in the assessment of heart rate variability in biobehavioral research. *Frontiers in psychology,* **5**:805, doi: 10.3389/fpsyg.2014.00805.

QUINTANA, D.S., MCGREGOR, I.S., GUASTELLA, A.J., MALHI, G.S. and KEMP, A.H., 2013. A meta-analysis on the impact of alcohol dependence on short-term resting-state heart rate variability: Implications for cardiovascular risk. *Alcoholism: Clinical and Experimental Research,* **37**, pp. E23-E29.

QUINTANA, D., ALVARES, G. and HEATHERS, J., 2016. Guidelines for Reporting Articles on Psychiatry and Heart rate variability (GRAPH): recommendations to advance research communication. *Translational psychiatry,* **6**(5), pp. e803.

RALEVSKI, E., PETRAKIS, I. and ALTEMUS, M., 2019. Heart rate variability in alcohol use: A review. *Pharmacology Biochemistry and Behavior,* **176**, pp. 83-92.

RENNIE, K.L., HEMINGWAY, H., KUMARI, M., BRUNNER, E., MALIK, M. and MARMOT, M., 2003. Effects of moderate and vigorous physical activity on heart rate variability in a British study of civil servants. *American Journal of Epidemiology,* **158**(2), pp. 135-143.

RENNIE, K.L., HENNINGS, S.J., MITCHELL, J. and WAREHAM, N.J., 2001. Estimating energy expenditure by heart-rate monitoring without individual calibration. *Medicine and Science in Sports and Exercise,* **33**(6), pp. 939-945.

REYES DEL PASO, GUSTAVO A, LANGEWITZ, W., MULDER, L.J., VAN ROON, A. and DUSCHEK, S., 2013. The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: a review with emphasis on a reanalysis of previous studies. *Psychophysiology,* **50**(5), pp. 477-487.

ROBERTSON, A., KING, K., RITCHIE, S., GAUTHIER, A.P., LAURENCE, M. and DORMAN, S., 2015. Validating the Use of Heart Rate Variability for Estimating Energy Expenditure. *International Journal of Human Movement and Sports Sciences,* **3**, pp. 19-26.

ROMANOWICZ, M., SCHMIDT, J.E., BOSTWICK, J.M., MRAZEK, D.A. and KARPYAK, V.M., 2011. Changes in heart rate variability associated with acute alcohol consumption: current knowledge and implications for practice and research. *Alcoholism: Clinical and Experimental Research*, **35**(6), pp. 1092-1105.

ROTHWELL, P.M., 2005. External validity of randomised controlled trials:"to whom do the results of this trial apply?". *The Lancet,* **365**(9453), pp. 82-93.

RUSKO, H., RÖNKÄ, T., UUSITALO, A., KINNUNEN, U., MAUNO, S., FELDT, T., KINNUNEN, M., MARTINMÄKI, K., HIRVONEN, A. and HYTTINEN, S., 2006. Stress and relaxation during sleep and awake time, and their associations with free salivary cortisol after awakening, *Nordic Ergonomics Society congress* 2006.

SAALASTI, S., 2003. Neural networks for heart rate time series analysis, *Jyväskylän yliopisto.*

SAALASTI, S., SEPPÄNEN, M. and KUUSELA, A., 2004. Artefact correction for heart beat interval data, *Advanced Methods for Processing Bioelectrical Signal, ProBisi Meeting 2004*.

SACHA, J., 2014. Interaction between heart rate and heart rate variability. *Annals of Noninvasive Electrocardiology,* **19**(3), pp. 207-216.

SACHA, J. and PLUTA, W., 2008. Alterations of an average heart rate change heart rate variability due to mathematical reasons. *International journal of cardiology,* **128**(3), pp. 444-447.

SAGAWA, Y., KONDO, H., MATSUBUCHI, N., TAKEMURA, T., KANAYAMA, H., KANEKO, Y., KANBAYASHI, T., HISHIKAWA, Y. and SHIMIZU, T., 2011. Alcohol has a dose-related effect on parasympathetic nerve activity during sleep. *Alcoholism: Clinical and Experimental Research,* **35**(11), pp. 2093-2100.

SALGADO, C.M., AZEVEDO, C., PROENÇA, H. and VIEIRA, S.M., 2016. Missing Data. In: MIT CRITICAL DATA, ed, *Secondary Analysis of Electronic Health Records.* Springer International Publishing, pp. 143-162.

SALO, M.A., HUIKURI, H.V. and SEPPÄNEN, T., 2001. Ectopic beats in heart rate variability analysis: effects of editing on time and frequency domain measures. *Annals of noninvasive electrocardiology,* **6**(1), pp. 5-17.

SANDERCOCK, G. and BRODIE, D., 2006. The use of heart rate variability measures to assess autonomic control during exercise. *Scandinavian Journal of Medicine & Science in Sports,* **16**(5), pp. 302-313.

SANDERCOCK, G.R., BROMLEY, P.D. and BRODIE, D.A., 2005. Effects of exercise on heart rate variability: inferences from meta-analysis. *Medicine and Science in Sports and Exercise,* **37**(3), pp. 433-439.

SATHYANARAYANA, A., JOTY, S., FERNANDEZ-LUQUE, L., OFLI, F., SRIVASTAVA, J., ELMAGARMID, A., ARORA, T. and TAHERI, S., 2016. Sleep Quality Prediction From Wearable Data Using Deep Learning. *JMIR mHealth and uHealth,* **4**(4), pp. e125.

SAWKA, M. and FRIEDL, K., 2018. Emerging Wearable Physiological Monitoring Technologies and Decision Aids for Health and Performance. *Journal of applied physiology,* **124**(2), pp. 430-431.

SCHOELLER, D.A. and VAN SANTEN, E., 1982. Measurement of energy expenditure in humans by doubly labeled water method. *Journal of applied physiology: respiratory, environmental and exercise physiology,* **53**(4), pp. 955-959.

SCHOLZ, U.J., BIANCHI, A.M., CERUTTI, S. and KUBICKI, S., 1997. Vegetative background of sleep: spectral analysis of the heart rate variability. *Physiology & Behavior,* **62**(5), pp. 1037-1043.

SHAFFER, F. and GINSBERG, J., 2017. An overview of heart rate variability metrics and norms. *Frontiers in public health,* **5**:258, doi: 10.3389/fpubh.2017.00258.

SHAFFER, F., MCCRATY, R. and ZERR, C.L., 2014. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in psychology,* **5**:1040, doi: 10.3389/fpsyg.2014.01040.

SHEPHARD, R.J. and AOYAGI, Y., 2009. Seasonal variations in physical activity and implications for human health. *European journal of applied physiology,* **107**(3), pp. 251-271.

SHEPHARD, R.J., 2003. Limits to the measurement of habitual physical activity by questionnaires Commentary. *British journal of sports medicine,* **37**(3), pp. 197-206.

SHERMAN, R.E., ANDERSON, S.A., DAL PAN, G.J., GRAY, G.W., GROSS, T., HUNTER, N.L., LAVANGE, L., MARINAC-DABIC, D., MARKS, P.W., ROBB, M.A., SHUREN, J., TEMPLE, R., WOODCOCK, J., YUE, L.Q. and CALIFF, R.M.,

2016. Real-world evidence - What is it and what can it tell us? *New England Journal of Medicine,* **375**(23), pp. 2293-2297.

SIBBALD, B. and ROLAND, M., 1998. Understanding controlled trials. Why are randomised controlled trials important? *BMJ (Clinical research ed.),* **316**(7126), pp. 201.

SIEVERINK, F., KELDERS, S.M. and VAN GEMERT-PIJNEN, J.E., 2017. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. *Journal of medical Internet research,* **19**(12), pp. e402.

SILFEE, V.J., HAUGHTON, C.F., JAKE-SCHOFFMAN, D.E., LOPEZ-CEPERO, A., MAY, C.N., SREEDHARA, M., ROSAL, M.C. and LEMON, S.C., 2018. Objective measurement of physical activity outcomes in lifestyle interventions among adults: A systematic review. *Preventive Medicine Reports,* **11**, pp. 74-80.

SINGH, N., MONEGHETTI, K.J., CHRISTLE, J.W., HADLEY, D., FROELICHER, V. and PLEWS, D., 2018. Heart Rate Variability: An Old Metric with New Meaning in the Era of Using mHealth technologies for Health and Exercise Training Guidance. Part Two: Prognosis and Training. *Arrhythmia & electrophysiology review,* **7**(4), pp. 247-255.

SMITH, G.D. and EBRAHIM, S., 2002. Data dredging, bias, or confounding. *BMJ,* **325**(7378), pp. 1437-1438.

SMOLANDER, J., AJOVIITA, M., JUUTI, T., NUMMELA, A. and RUSKO, H., 2011. Estimating oxygen consumption from heart rate and heart rate variability without individual calibration. *Clinical Physiology and Functional Imaging,* **31**(4), pp. 266-271.

SOMERS, V.K., DYKEN, M.E., MARK, A.L. and ABBOUD, F.M., 1993. Sympathetic-nerve activity during sleep in normal subjects. *New England Journal of Medicine,* **328**(5), pp. 303-307.

SPAAK, J., TOMLINSON, G., MCGOWAN, C.L., SOLEAS, G.J., MORRIS, B.L., PICTON, P., NOTARIUS, C.F. and FLORAS, J.S., 2010. Dose-related effects of red wine and alcohol on heart rate variability. *American Journal of Physiology-Heart and Circulatory Physiology,* **298**(6), pp. H2226-H2231.

SPALLONE, V., ZIEGLER, D., FREEMAN, R., BERNARDI, L., FRONTONI, S., POP-BUSUI, R., STEVENS, M., KEMPLER, P., HILSTED, J. and TESFAYE, S., 2011. Cardiovascular autonomic neuropathy in diabetes: clinical impact, assessment, diagnosis, and management. *Diabetes/metabolism research and reviews,* **27**(7), pp. 639-653.

SPERRIN, M., RUSHTON, H., DIXON, W., NORMAND, A., VILLARD, J., CHIEH, A. and BUCHAN, I., 2016. Who Self-Weighs and What Do They Gain From It? A Retrospective Comparison Between Smart Scale Users and the General Population in England. *Journal of Medical Internet Research,* **18**(1), pp. e17.

STAMATAKIS, E., EKELUND, U., DING, D., HAMER, M., BAUMAN, A.E. and LEE, I.M., 2019. Is the time right for quantitative public health guidelines on sitting? A narrative review of sedentary behaviour research paradigms and findings. *British journal of sports medicine,* **53**(6), pp. 377-382.

STRATH, S.J., KAMINSKY, L.A., AINSWORTH, B.E., EKELUND, U., FREEDSON, P.S., GARY, R.A., RICHARDSON, C.R., SMITH, D.T., SWARTZ, A.M., 2013. Guide to the assessment of physical activity: Clinical and research applications: A scientific statement from the American Heart association. *Circulation,* **128**(20), pp. 2259-2279.

SULLIVAN, G.M. and FEINN, R., 2012. Using Effect Size-or Why the P Value Is Not Enough. *Journal of graduate medical education,* **4**(3), pp. 279-282.

SWAN, M., 2012. Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. *Journal of Sensor and Actuator Networks,* **1**(3), pp. 217-253.

SWIFT, B., JAIN, L., WHITE, C., CHANDRASEKARAN, V., BHANDARI, A., HUGHES, D.A. and JADHAV, P.R., 2018. Innovation at the Intersection of Clinical Trials and Real-World Data Science to Advance Patient Care: Innovation at the Intersection of Clinical Trials. *Clinical and Translational Science,* **11**(5), pp. 450-460.

TAYLOR, J.A., MYERS, C.W., HALLIWILL, J.R., SEIDEL, H. and ECKBERG, D.L., 2001. Sympathetic restraint of respiratory sinus arrhythmia: implications for vagal-cardiac tone assessment in humans. *American Journal of Physiology-Heart and Circulatory Physiology,* **280**(6), pp. H2804-H2814.

TEISALA, T., MUTIKAINEN, S., TOLVANEN, A., ROTTENSTEINER, M., LESKINEN, T., KAPRIO, J., KOLEHMAINEN, M., RUSKO, H. and KUJALA, U.M., 2014. Associations of physical activity, fitness, and body composition with heart rate variability–based indicators of stress and recovery on workdays: a cross-sectional study. *Journal of occupational medicine and toxicology,* **9**:16, doi:10.1186/1745-6673-9-16.

THAYER, J.F. and STERNBERG, E., 2006. Beyond heart rate variability: vagal regulation of allostatic systems. *Annals of the New York Academy of Sciences,* **1088**(1), pp. 361-372.

THAYER, J.F., YAMAMOTO, S.S. and BROSSCHOT, J.F., 2010. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. *International journal of cardiology,* **141**(2), pp. 122-131.

TOBALDINI, E., NOBILI, L., STRADA, S., CASALI, K.R., BRAGHIROLI, A. and MONTANO, N., 2013. Heart rate variability in normal and pathological sleep. *Frontiers in physiology,* **4**:294, doi: 10.3389/fphys.2013.00294.

TOBIN, J., 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica,* **26**(1), pp. 24-36. Available (accessed 04.06.2019): http://links.jstor.org/sici?sici=0012-9682%28195801%2926%3A1%3C24%3AEORFLD%3E2.0.CO%3B2-R. Archived by WebCite® at http://www.webcitation.org/78stu6cpQ.

TRINDER, J., KLEIMAN, J., CARRINGTON, M., SMITH, S., BREEN, S., TAN, N. and KIM, Y., 2001. Autonomic activity during human sleep as a function of time and sleep stage. *Journal of sleep research,* **10**(4), pp. 253-264.

TSUJI, H., VENDITTI, F.J., MANDERS, E.S., EVANS, J.C., LARSON, M.G., FELDMAN, C.L. and LEVY, D., 1996. Determinants of heart rate variability. *Journal of the American College of Cardiology,* **28**(6), pp. 1539-1546.

TUCKER, J.M., WELK, G.J. and BEYLER, N.K., 2011. Physical activity in US adults: compliance with the physical activity guidelines for Americans. *American Journal of Preventive Medicine,* **40**(4), pp. 454-461.

TUOMAINEN, P., PEUHKURINEN, K., KETTUNEN, R. and RAURAMAA, R., 2005. Regular physical exercise, heart rate variability and turbulence in a 6-year randomized controlled trial in middle-aged men: the DNASCO study. *Life Sciences,* **77**(21), pp. 2723-2734.

UITENBROEK, D.G., 1993. Seasonal variation in leisure time physical activity. *Medicine and Science in Sports and Exercise,* **25**(6), pp. 755-760.

UMETANI, K., SINGER, D.H., MCCRATY, R. and ATKINSON, M., 1998. Twenty-four hour time domain heart rate variability and heart rate: relations to age and gender over nine decades. *Journal of the American College of Cardiology,* **31**(3), pp. 593-601.

UUSITALO, A.L., LAITINEN, T., VÄISÄNEN, S.B., LÄNSIMIES, E. and RAURAMAA, R., 2004. Physical training and heart rate and blood pressure variability: a 5-yr randomized trial. *American Journal of Physiology-Heart and Circulatory Physiology,* **286**(5), pp. H1821-H1826.

UUSITALO, A., METS, T., MARTINMÄKI, K., MAUNO, S., KINNUNEN, U. and RUSKO, H., 2011. Heart rate variability related to effort at work. *Applied Ergonomics,* **42**(6), pp. 830-838.

VALKONEN-KORHONEN, M., TARVAINEN, M.P., RANTA-AHO, P., KARJALAINEN, P.A., PARTANEN, J., KARHU, J. and LEHTONEN, J., 2003. Heart rate variability in acute psychosis. *Psychophysiology,* **40**(5), pp. 716-726.

VAN DYCK, D., CERIN, E., DE BOURDEAUDHUIJ, I., HINCKSON, E., REIS, R.S., DAVEY, R., SARMIENTO, O.L., MITAS, J., TROELSEN, J., MACFARLANE, D., SALVO, D., AGUINAGA-ONTOSO, I., OWEN, N., CAIN, K.L. and SALLIS, J.F., 2015. International study of objectively measured physical activity and sedentary time with body mass index and obesity: IPEN adult study. *International journal of obesity,* **39**(2), pp. 199-207.

VANDENBROUCKE, J.P., VON ELM, E., ALTMAN, D.G., GØTZSCHE, P.C., MULROW, C.D., POCOCK, S.J., POOLE, C., SCHLESSELMAN, J.J., EGGER, M. and STROBE INITIATIVE, 2007. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *Epidemiology,* **18**(6), pp. 805-835.

VANDER, A.J., SHERMAN, J.H. and LUCIANO, D.S., 1990. *Human physiology: the mechanisms of body function.* 5th edn. New York (NY): McGraw-Hill.

VARTIAINEN, E., LAATIKAINEN, T., PELTONEN, M., JUOLEVI, A., MÄNNISTÖ, S., SUNDVALL, J., JOUSILAHTI, P., SALOMAA, V., VALSTA, L. and PUSKA, P., 2009. Thirty-five-year trends in cardiovascular risk factors in Finland. *International journal of epidemiology,* **39**(2), pp. 504-518.

VIDAKOVIC, B., 2011. Statistics for bioengineering sciences: with MATLAB and WinBUGS support. *Springer Science & Business Media.*

VINIK, A., 2012. The conductor of the autonomic orchestra. Frontiers in endocrinology, **3**: 71, doi: 10.3389/fendo.2012.00071.

VINIK, A.I., MASER, R.E., MITCHELL, B.D. and FREEMAN, R., 2003. Diabetic autonomic neuropathy. *Diabetes care,* **26**(5), pp. 1553-1579.

WAREHAM, N.J. and RENNIE, K.L., 1998. The assessment of physical activity in individuals and populations: Why try to be more precise about how physical activity is assessed? *International journal of obesity,* **22**(2), pp. S30-S38.

WEILER, D.T., VILLAJUAN, S.O., EDKINS, L., CLEARY, S. and SALEEM, J.J., 2017. Wearable Heart Rate Monitor Technology Accuracy in Research: A Comparative Study Between PPG and ECG Technology, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 2017, SAGE Publications Sage CA: Los Angeles, CA, pp. 1292-1296.

WEISMAN, I.M. and ZEBALLOS, R.J., 2002. Clinical exercise testing. *Karger Medical and Scientific Publishers.*

WENNMAN, H., PIETILÄ, A., RISSANEN, H., VALKEINEN, H., PARTONEN, T., MÄKI-OPAS, T. and BORODULIN, K., 2019. Gender, age and socioeconomic variation in 24-hour physical activity by wrist-worn accelerometers: the FinHealth 2017 Survey. *Scientific reports,* **9**(1), pp. 6534.

WEST, S.G., DUAN, N., PEQUEGNAT, W., GAIST, P., DES JARLAIS, D.C., HOLTGRAVE, D., SZAPOCZNIK, J., FISHBEIN, M., RAPKIN, B. and CLATTS, M., 2008. Alternatives to the randomized controlled trial. *American Journal of Public Health,* **98**(8), pp. 1359-1366.

WESTERTERP, K.R., 2009. Assessment of physical activity: a critical appraisal. *European journal of applied physiology,* **105**(6), pp. 823-828.

WILLIAMS, J.A. and COOK, D.J., 2017. Forecasting behavior in smart homes based on sleep and wake patterns. *Technology and Health Care,* **25**(1), pp. 89-110.

WILMORE, J.H., STANFORTH, P.R., GAGNON, J., RICE, T., MANDEL, S., LEON, A.S., RAO, D., SKINNER, J.S. and BOUCHARD, C., 2001. Heart rate and blood pressure changes with endurance training: the HERITAGE Family Study. *Medicine and Science in Sports and Exercise,* **33**(1), pp. 107-116.

WRIGHT, S.P., HALL BROWN, T.S., COLLIER, S.R. and SANDBERG, K., 2017. How consumer physical activity monitors could transform human physiology research. *American Journal of Physiology - Regulatory Integrative and Comparative Physiology,* **312**(3), pp. R358-R367.

YAMAMOTO, K., MIYACHI, M., SAITOH, T., YOSHIOKA, A. and ONODERA, S., 2001. Effects of endurance training on resting and post-exercise cardiac autonomic control. *Medicine and Science in Sports and Exercise,* **33**(9), pp. 1496-1502.

YOTSUKURA, M., KOIDE, Y., FUJII, K., TOMONO, Y., KATAYAMA, A., ANDO, H., SUZUKI, J. and ISHIKAWA, K., 1998. Heart rate variability during the first month of smoking cessation. *American Heart Journal,* **135**(6), pp. 1004-1009.

ZAFFALON, J.J.R., VIANA, A.O., DE MELO, G.E.L. and DE ANGELIS, K., 2018. The impact of sedentarism on heart rate variability (HRV) at rest and in response to mental stress in young women. *Physiological reports,* **6**(18), pp. e13873.

ZAZA, A. and LOMBARDI, F., 2001. Autonomic indexes based on the analysis of heart rate variability: a view from the sinus node. *Cardiovascular Research,* **50**(3), pp. 434-442.

ZHANG, D., SHEN, X. and QI, X., 2016. Resting heart rate and all-cause and cardiovascular mortality in the general population: a meta-analysis. *Canadian Medical Association journal,* **188**(3), pp. E53-63.

"behaviour | behavior, n.". 2019. *Oxford English Dictionary Online.* Oxford University Press.

APPENDIX A

FIRSTBEAT Technologies

CHOOSING THE ACTIVITY CLASS

Choose the activity class (one number between 0 and 10) that best describes your typical physical activity (endurance-type exercise or physical work) during the last 2 - 3 months:

Your typical physical activity level	How often are you physically active?	Weekly training amount	Activity class
No physical activity	-	-	0
Occasional light physical activity	Once every two weeks	Less than 15 min	1
		Less than 30min	2
	Once per week	~30min	3
Regular training	2-3x / week	~45min	4
		45min-1h	5
		1-3h	6
	3-5x / week	3-5h	7
		5-7h	7,5
Daily training	Almost daily	7-9h	8
		9-11h	8,5
	Daily	11-13h	9
		13-15h	9,5
		More than 15h	10

APPENDIX B

FIRSTBEAT LIFESTYLE ASSESSMENT

- ***PRE QUESTIONNAIRE***

Name

Before the Lifestyle Assessment, I would characterize my well-being and my current methods for taking care of my well-being as follows:

Questions:	Your answers:
1. I think I am physically active enough to get health benefits.	
2. I think my physical activity is intensive enough to improve my fitness.	
3. In my opinion, my eating habits are healthy.	
4. I feel that my alcohol consumption is not excessive.	
5. I feel stressed.	
6. My days include breaks that allow me to recover.	
7. I feel tired frequently.	
8. I feel that I sleep enough.	
9. I feel that I can influence the things that affect my health.	
10. In my opinion, I feel well at the moment.	

Scale of answers:

Completely agree
Partially agree
Cannot say
Partially disagree
Completely disagree

PUBLICATIONS

PUBLICATION
I

Methods to Use Big Wearable Heart Rate Data for Estimation of Physical Activity in Population Level

PIETILÄ, J., MUTIKAINEN, S., HELANDER, E., MYLLYMÄKI, T., KUJALA, U.M. and KORHONEN, I.

6th European Conference of the International Federation for Medical and Biological Engineering 2015, pp. 66-69
DOI: 10.1007/978-3-319-11128-5_17

Methods to Use Big Wearable Heart Rate Data for Estimation of Physical Activity in Population Level

J. Pietilä[1], S. Mutikainen[2], E. Helander[1], T. Myllymäki[3], U.M. Kujala[2] and I. Korhonen[1]

[1] Tampere University of Technology, Department of Signal Processing, Tampere, Finland
[2]University of Jyvaskyla, Department of Health Sciences, Jyvaskyla, Finland
[3]Firstbeat Technologies Ltd, Jyvaskyla, Finland

Abstract— **Technologies for wearable health monitoring are becoming increasingly popular and affordable. As a result, large-scale health databases from a large number of individuals are becoming available. However, analysis of these databases requires special methodology to transform available parameters into more generic ones and to manage such non-balanced data characteristics as biases and sampling issues. In this paper, we introduce a methodology for studying physical activity from big wearable heart rate (HR) data on about 5 000 working-age individuals, each measured only for a few days. Physical activity was assessed by oxygen consumption (VO2) calculated from measured HR data using a neural network model. Minute-to-minute VO2 data was used to quantify various physical activities in a measurement day, as defined according to the health promoting physical activity minutes of the American College of Sports Medicine. We set *a posteriori* inclusion criteria for the data on the subjects' personal background parameters and the quality of their HR data. The effect of different subjects being measured in different months and weekdays was removed by using a linear model. The linear model sought to estimate the physical activity minutes based on a subject's background parameters. The results show that big data collected in real-life settings and originally for non-research purposes can with appropriate data management and analysis methodology provide unique knowledge of lifestyles and behavior.**

Keywords— **physical activity, oxygen consumption, heart rate (HR) data, Firstbeat, non-research database**

I. Introduction

Today, technologies of wearable health monitoring enable easy and affordable large-scale acquisition of health data during daily life. As a result, large-scale databases containing health data from a large number of individuals are created as by-products of, e.g., health promoting interventions. These data sets may outnumber those collected for research purposes. However, their use for research purposes is not straightforward. For example, they may lack essential information, their inclusion and exclusion criteria may be hastily defined, and they may not be automatically representative. In addition, legal issues related to data privacy and security must be properly managed. Included are also several technical challenges in processing and analyzing such data sets.

This paper aims to present a methodology for quantifying physical activity from a big database of ambulatory beat-to-beat heart rate (HR) data. We introduce key challenges and a methodology for transforming the data into generic metrics and for managing non-balanced sampling of data with statistical methods.

II. Materials and methods

A. Data

The data used here was provided by Firstbeat Technologies Ltd., a Finnish company providing analytics for well-being factors such as stress, recovery, and physical activity based on ambulatory measurement of beat-to-beat HR and subsequent analysis of heart rate variability (HRV). In real-life conditions, HR data is typically collected over three days with a Firstbeat Bodyguard device (Firstbeat Technologies Ltd, Jyväskylä, Finland) and analyzed by Firstbeat HEALTH software, producing Lifestyle Assessment results.

Over the years, Firstbeat Technologies Ltd. has been gathering Lifestyle Assessment results into an anonymized database, which includes results of Finnish working-age subjects who have voluntarily participated in measurements as part of their preventive occupational health care services. In addition to the results, this database includes the subjects' personal background parameters of age, weight, height, gender, and physical activity class. The subjects choose a physical activity class on a scale from zero to ten according to their amount and intensity of physical activity.

The methods reported in this paper were developed and applied for trial purposes to about 14 000 measurement days obtained from about 5 000 subjects who had at least two and at most six days of measurement during 2007-2013.

B. Transforming HR to oxygen consumption

A common challenge in wearable sensor data is that the monitoring devices monitor parameters that may be device-specific (e.g., "Nikefuel" and physical activity) or that are only indirectly linked to the phenomenon to be studied (e.g., HR and physical activity). Therefore, it is necessary that the data be transformed into generic metrics. For example, we describe the use of HR to assess physical activity in daily life conditions as the sole input signal.

The intensity of physical activity can be most accurately measured from oxygen consumption (VO2). However, a direct VO2 measurement is not suitable for large-scale field studies because special equipment, such as respiratory gas analyzer, is required. [1] The use of HR is one of the most studied methods for indirect estimation of VO2. However, accurate estimation of VO2 from HR requires individual or at least group-level calibration [2,3]. Consequently, large-scale studies have so far employed subjective questionnaires, inaccurate actigraphy, or HR-based estimation of VO2 to assess participants' physical activity.

Firstbeat Technologies Ltd. has developed a novel neural network model which estimates VO2 from measured HR data without calibration [4]. The model uses personal background parameters (age, weight, height, gender, and physical activity class), HR, respiration frequency, and on/off dynamics to evaluate the relationship between HR and VO2. Respiration frequency is derived from measured HR and the high-frequency component of HRV, and the purpose of the on/off dynamics is to describe variations in the HR-VO2 relationship. [5] The model produces temporal VO2, which can be presented in units of metabolic equivalents (METs) by dividing temporal VO2 values by 3.5. The 3.5 ml/kg/min is considered the VO2 level at rest, and thus MET values represent a multiplication of VO2 at rest. [6]

A study on the validity of the neural network model found the modelled VO2 accurate for light and heavy intensity activities at group level and for individual values. At low intensity levels, however, the model slightly underestimated VO2 at group level. Overall, the modelled VO2 accounted for 87 % of the variability in the measured VO2 at group level and for 77–97 % at individual level. In other words, the model was at least as accurate as the methods that use a simple HR calibration test. The model was thus judged accurate enough to determine the average VO2 in field measurements. [1] Consequently, we applied this model here to transform HR into minute-to-minute (min-to-min) VO2 data without recourse to personal calibration.

C. Transforming oxygen consumption into physical activity

Physical activity can be quantified into different categories according to its intensity and duration. For instance, the American College of Sports Medicine (ACSM) has defined health promoting physical activity as that whose intensity is moderate-to-vigorous, i.e., the MET level is higher than or equal to three continuously for at least 10 minutes [7]. These classifications can be used as a basis for quantifying physical activity.

In this paper, physical activity was quantified by calculating it in the number of minutes that followed the ACSM's definition of health promoting physical activity, except that one 1-minute MET value was allowed to be lower than three in a 10-minute period. First, 1-minute MET values were calculated; i.e., for each 1-minute segment of measurement, the average of second-by-second MET values produced by the neural network model was calculated. Thereafter, the physical activity minutes were calculated as a total of 10-minute segments in which all 1-minute MET values were higher than or equal to three, except for one 1-minute MET value, which was allowed to be less than three.

D. Application of a posteriori *inclusion criteria to the data*

A key challenge in using non-research data for scientific purposes is to introduce inclusion and exclusion criteria *a posteriori* in the data. In the methodology we adopted here, inclusion criteria were set for both personal background parameters (inclusion of subjects) and data quality (inclusion of data) to produce reliable results.

The anonymized data set comprised volunteers who in all appearances were healthy. According to the instructions of Firstbeat Technologies Ltd, Lifestyle Assessment should not be performed in the following conditions: chronic rhythm disturbance, cardiac pacemaker or transplant, left bundle branch block, severe cardiac disease, very high blood pressure (≥180/100 mmHg), type 1 or 2 diabetes with autonomic neuropathy, hyperthyreosis or other disturbance of the thyroid gland leading to a resting HR of >80 bpm, severe neurological disease, fever or other acute disease, a body mass index (BMI) of >40.0 kg/m^2, and any medication affecting HR, HRV, or physical activity level. Therefore, these criteria can be considered *a priori* exclusion criteria.

In data analysis, *a posteriori* inclusion criteria for subject characteristics and measurement data are applied as seen appropriate to the study goals. In this example, subjects considered for analysis were 18–65 years old and had a BMI of 18.5–40 kg/m^2. For the data to be a reliable sample of a subject's daily physical activity, we set a daily measurement limit of a minimum of 16 hours with a measurement break no longer than 30 minutes during waking hours. Moreover, the average detected artifact percentage of HR data in a daily measurement was set to be lower than 15, even though before producing results, Firstbeat analysis includes a powerful artifact correction procedure for erroneous HR data.

E. Statistical Analysis

A key challenge in statistical analysis of large non-research databases is non-balanced sampling of data. For example, we had only a few measurement days per subject, and different subjects participated in measurements in different weekdays and months. In an optimal case, when, e.g., the minutes of physical activity differ in weekdays and months, measurements should be performed longitudinally on the same individuals. However, an appropriate statistical approach can help us use large-scale databases to address these questions. Computational modelling of data allows us to control inter-individual differences and study the effect of time in physical activity.

The physical activity minutes were found to depend on the background parameters of age, BMI, physical activity class, and gender. Thus to control the effect of these parameters on the observed physical activity minutes, a linear model was first constructed between the background parameters ($\bar{X}$) and the observed number of physical activity minutes ($\bar{Y}$):

$$\bar{Y} = \bar{\beta}\bar{X} + \bar{\varepsilon}, \tag{1}$$

where $\bar{\beta}$ is a vector of the constant term and the coefficients for the background parameters, and $\bar{\varepsilon}$ is a vector of residual terms. The coefficient values were estimated based on the data using a least-squares fit, and all the background parameters were found statistically significant; i.e., they had a p-value lower than 0.05.

To calculate the number of background-controlled physical activity minutes per subject, the baseline (b) was first calculated based on the observed number of physical activity minutes ($\bar{Y}$) and residual terms ($\bar{\varepsilon}$):

$$b = \frac{1}{N}\sum_{i=1}^{N}\bar{Y}(i) - \frac{1}{N}\sum_{i=1}^{N}\bar{\varepsilon}(i), \tag{2}$$

where N is the total number of measurements. Thereafter, the residual terms ($\bar{\varepsilon}$) were added to the baseline (b) to produce the background-controlled number of physical activity minutes ($\bar{Y}_c$):

$$\bar{Y}_c = \bar{\varepsilon} + b. \tag{3}$$

After receiving the background-controlled minutes of physical activity, i.e., after removing the effect of different subjects taking part in measurement in different days and months, mean values were calculated for the background-controlled physical activity minutes for different months and weekdays.

III. Results

A. Fulfillment of the inclusion criteria

The initial data set used for testing the above methods comprised 21 720 measurement days on 5 373 subjects. The total number of measurement days fulfilling the inclusion criteria was 14 525, gathered from a total of 5 124 subjects (2 422 males and 2 702 females). The means and standard deviations in the male and female subjects' background parameters are shown in Table 1.

Table 1 The mean and standard deviations in the background parameters of male and female subjects fulfilling the inclusion criteria.

Background parameter	Male	Female
Age (in years)	43.5 (10.1)	44.4 (9.8)
BMI (in kg/m^2)	26.7 (3.6)	25.5 (4.4)
Activity class (on scale of 0–10)	5.0 (2.0)	4.8 (1.9)

B. Everyday physical activity

VO2 analysis was successfully applied to measured HR data, and min-to-min VO2 data were extracted and transformed to daily physical activity characteristics, as suggested in the methods section. An example of HR and min-to-min VO2 data for a single individual is shown in Fig 1.

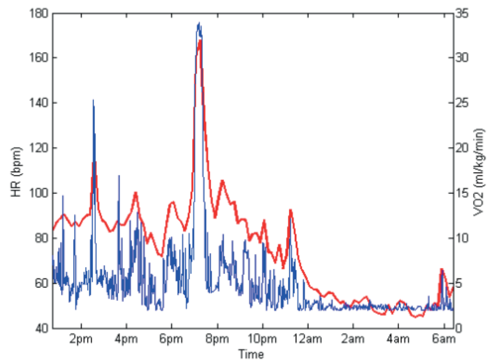


Fig 1 Example figures of HR as 10-minute averages (red line) and min-to-min VO2 data (blue line) for a single individual..

The results of the physical activity minutes were calculated as a function of weekday and month. The linear model used for removing inter-individual differences explained 15.9 % of the total variance in the observed physical activity minutes. The effect of removing the inter-individual differences is shown in Fig 2, where the colors represent the number of physical activity minutes: the warmer the color, the greater the number of physical activity minutes. The observed number of physical activity minutes without background control is shown in Fig 2 (a). Fig 2 (b) shows the distribution of the background-controlled physical activity minutes over weekdays and months.

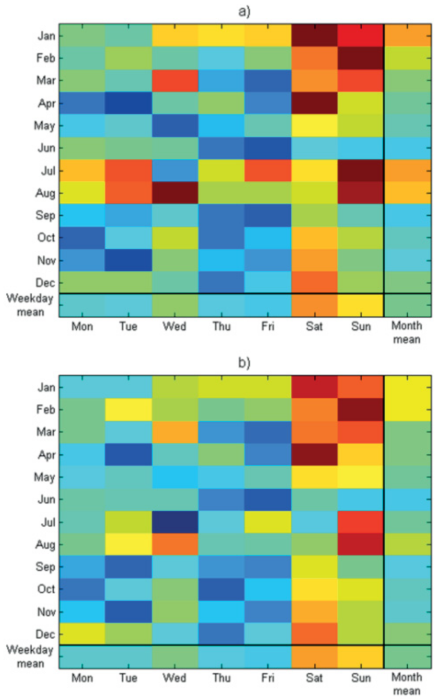


Fig. 2 Observed minutes of physical activity (figure a) and background-controlled minutes of physical activity (figure b).

As seen in Fig 2, controlling the background parameters reduced the difference in physical activity minutes between days and months. The control seemed to affect the month means of physical activity minutes but it seemed not to affect the weekday means of physical activity minutes. These results show that the approach presented in this paper can reduce the impact of inter-individual differences when a large enough sample is available.

IV. Conclusions

In this paper, we introduced a physiological database produced with beat-by-beat HR recordings, which can be used to quantify physical activity. To determine the effect of months and weekdays on physical activity, non-balanced sampling of the database was compensated for by using a linear model. However, organizing a study with reference data on this scale is very challenging and underlines the difficulty of using big data for research purposes; i.e., that results may not be able to be validated as in a traditional scientific approach and should thus be interpreted with a pinch of salt.

Use of non-research, large-scale health databases opens up new possibilities for research. Such databases often outnumber their research-originated competitors and compared to randomized controlled trials may provide some benefits such as generalizability to real life situations. However, special methodology is needed to transform such databases into standard units and to manage issues related to inclusion/exclusion criteria, representativeness, data quality, and non-balanced settings. Moreover, since these databases usually provide observational data, it is important that the results be interpreted cautiously and without over-interpreting causal relations.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Smolander J et al (2011) Estimating oxygen consumption from heart rate and heart rate variability without individual calibration. Clin Physiol Funct Imaging 31:266–271
2. Rennie KL et al (2001) Estimating energy expenditure by heart-rate monitoring without individual calibration. Med Sci Sports Exerc, 33: 939–945
3. Li R et al (1993) A critical evaluation of heart rate monitoring to assess energy expenditure in individuals. Am J Clin Nutr, 58:602–607
4. Pulkkinen A et al (2005) Energy expenditure can be accurately estimated from HR without individual laboratory calibration. The 52nd Annual Meeting of the ACSM, Nashville, Tennessee, USA.
5. Pulkkinen A et al (2004) On- and Off-dynamics and Respiration Rate Enhance the Accuracy of Heart Rate Based VO2 Estimation, The 51st Annual Meeting of the ACSM, Indianapolis, USA.
6. Norton K et al (2009) Position statement on physical activity and exercise intensity terminology. Journal of Science and Medicine in Sport, 13:496–502
7. Garber CE et al (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

PUBLICATION
II

Physical Activity: Absolute Intensity vs. Relative-to-Fitness-Level Volumes

KUJALA, U., PIETILÄ, J., MYLLYMÄKI, T., MUTIKAINEN, S.,
FÖHR, T., KORHONEN, I. and HELANDER, E

Medicine and Science in Sports and Exercise, 49(3): pp. 474-481
DOI: 10.1249/MSS.0000000000001134

Physical Activity: Absolute Intensity versus Relative-to-Fitness-Level Volumes

URHO M. KUJALA[1], JULIA PIETILÄ[2], TERO MYLLYMÄKI[3], SARA MUTIKAINEN[1], TIINA FÖHR[1], ILKKA KORHONEN[2], and ELINA HELANDER[2]

[1]*Department of Health Sciences, University of Jyväskylä, Jyväskylä, FINLAND;* [2]*Department of Signal Processing, Tampere University of Technology, Tampere, FINLAND; and* [3]*Department of Psychology, University of Jyväskylä, Jyväskylä, FINLAND*

ABSTRACT

KUJALA, U. M., J. PIETILÄ, T. MYLLYMÄKI, S. MUTIKAINEN, T. FÖHR, I. KORHONEN, and E. HELANDER. Physical Activity: Absolute Intensity versus Relative-to-Fitness-Level Volumes. *Med. Sci. Sports Exerc.*, Vol. 49, No. 3, pp. 474–481, 2017. **Purpose**: This study aimed to investigate in a real-life setting how moderate- and vigorous-intensity physical activity (PA) volumes differ according to absolute intensity recommendation and relative to individual fitness level by sex, age, and body mass index. **Methods**: A total of 23,224 Finnish employees (10,201 men and 13,023 women; ages 18–65 yr; body mass index = 18.5–40.0 kg·m^{-2}) participated in heart rate recording for 2+ d. We used heart rate and its variability, respiration rate, and on/off response information from R-R interval data calibrated by participant characteristics to objectively determine daily PA volume, as follows: daily minutes of absolute moderate (3–<6 METs) and vigorous (≥6 METs) PA and minutes relative to individual aerobic fitness for moderate (40%–<60% of oxygen uptake reserve) and vigorous (≥60%) PA. **Results**: According to absolute intensity categorization, the volume of both moderate- and vigorous-intensity PA was higher in men compared with women ($P < 0.001$), in younger compared with older participants ($P < 0.001$), and in normal weight compared with overweight or obese participants ($P < 0.001$). When the volume of PA intensity was estimated relative to individual fitness level, the differences were much smaller. Mean daily minutes of absolute vigorous-intensity PA were higher than those of relative intensity minutes in normal weight men ages 18–40 yr (17.7, 95% confidence interval [CI] = 16.9–18.6, vs 8.6, 95% CI = 8.0–9.1; $P < 0.001$), but the reverse was the case for obese women ages 41–65 yr (0.3, 95% CI = 0.2–0.4, vs 7.8, 95% CI = 7.2–8.4; $P < 0.001$). **Conclusion**: Compared with low-fit persons, high-fit persons more frequently reach an absolute target PA intensity, but reaching the target is more similar for relative intensity. **Key Words:** EXERCISE, OBJECTIVE MONITORING, HEART RATE, FITNESS

Increasing physical activity (PA) among both healthy people and individuals with chronic disease is linked to many health benefits (22,26,27). The current PA guidelines for aerobic PA (27,39) recommend at least 150 min of moderate-intensity PA (MPA) or at least 75 min of vigorous-intensity PA (VPA) per week, accumulated in bouts of at least 10 min in duration. This recommendation is based mostly on observational cohort studies that most often used self-reported questionnaire measures of leisure-time PA. Results of accelerometer-based (1,7,36) and heart rate–based (25) objective assessments of PA indicate that only a small proportion of adult populations meet the recommendation. PA bouts shorter than 10 min, often occurring during daily life and unplanned (38), are not included when investigating who fulfills this PA recommendation. However, accumulating evidence suggests that short bouts of moderate-to-vigorous PA (MVPA) are associated with reduced levels of cardiometabolic risk factors (6,12,33,38).

PA can be assessed using questionnaires or more objective monitoring methods (34). Recent advances in accelerometer-based PA monitoring techniques help yield good estimates of the intensity of certain types of PA, such as walking and running on a standard surface (37). Heart rate–based monitoring methods, however, are better for determining the intensity of different types of real-life MVPA (34), including bicycling and many work-related activities. Accelerometer-based objective monitoring methods may be more reliable in recording low to very low intensity PA compared with simple heart rate–based devices because of artifacts resulting from excitement and other stimuli unrelated to PA that still influence heart rate (34). The current aerobic PA recommendations are for MVPA characterized in absolute multiples of resting metabolic rate, MET, values. However, maximal exercise capacity in low-fit individuals, in particular among those who are older, obese, or have chronic disease, may be lower than the recommended absolute intensity level of VPA. Consequently, individuals who cannot reach the recommended intensity level do not have this type of PA recorded. Physicians and other professionals giving exercise recommendations need to understand which types of PA are achievable by physically inactive people with low fitness levels, the most important

Address for correspondence: Urho M. Kujala, M.D., Department of Health Sciences, University of Jyväskylä, PO Box 35 (LL), FI-40014 University of Jyväskylä, Finland; E-mail: urho.m.kujala@jyu.fi.
Submitted for publication May 2016.
Accepted for publication October 2016.
Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.acsm-msse.org).

0195-9131/17/4903-0474/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0000000000001134

target group for PA promotion. Although variation exists in how people feel pleasure and discomfort when they exercise at different intensity levels, in general, pleasure is reduced when the ventilatory or lactate threshold is surpassed (8).

The aim of this study was to compare the volumes of objectively monitored PA determined by recommended absolute intensity levels and by intensity levels relative to individual fitness by sex, age, and body mass index (BMI) (normal weight, overweight, and obese) among 23,224 Finnish employees during everyday life. To determine the PA volumes, we used sophisticated and validated methodology (25), including information on continuous heart rate and heart rate variability recordings.

METHODS

Study design and participants. This cross-sectional study investigated the amount of absolutely and relatively (i.e., relative to participant's maximal oxygen uptake [$\dot{V}O_{2max}$]) determined PA at different intensity levels (moderate, vigorous, and moderate-to-vigorous combined) during workdays and days off by sex, age, and BMI among a sample of 23,224 Finnish employees (10,201 men and 13,023 women; age range = 18–65 yr, BMI range = 18.5–40.0 kg·m^{-2}) who participated in real-life preventive occupational health care provided by their employers during the years 2007–2015 in Finland (Fig. 1). A wide nonselective range of nonmanual and manual labor employees was included. The employees participated in real-life continuous beat-to-beat R-R interval recordings. The majority of participants were apparently healthy because individuals with chronic disease and medications influencing heart rate did not participate in these recordings. For detailed exclusion criteria for participation in the R-R interval recordings, see Mutikainen et al. (25).

The data obtained from these R-R interval recordings were anonymously stored in a database administered by the software manufacturer (Firstbeat Technologies Ltd., Jyväskylä, Finland). According to written agreements (25), Firstbeat Technologies Ltd. extracted an anonymous data file for the present research purposes. This study was approved by the Ethics Committee of Tampere University Hospital (reference no. R13160).

PA monitoring and assessment. The ambulatory beat-to-beat R-R interval data used to calculate the intensity and amount of PA were recorded during the course of normal everyday life, usually over 3 d (typically including two workdays and 1 d off), using the Firstbeat Bodyguard device with stick-on electrodes with wires (Firstbeat Technologies Ltd.). Monitoring data were first analyzed using Firstbeat Analysis Server software (version 6.3, Firstbeat Technologies Ltd.). To be included, a participant had to have a measurement period, including at least one workday and 1 d off (Fig. 1). We included a workday or a day off in the analysis if the measurement period lasted 16–30 h·d^{-1}. Because the measurement day was determined from waking up to waking up, recordings were allowed to exceed 24 h. The information about the type of day was obtained from participant diaries; a workday had to include ≥4 working hours cumulatively, days off were without any working hours, and the days with reported work time >0 but <4 h were excluded

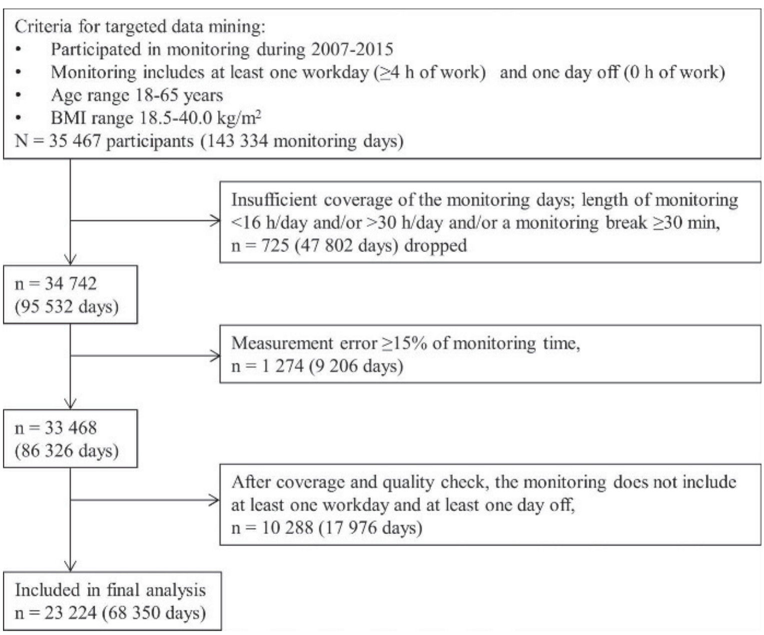


FIGURE 1—Flow of participants and measurement days included in the analysis.

from the analyses. The analyzed data consisted of successfully recorded (measurement error in recording R-R intervals detected with an automatic artifact detection and correction feature for irregular ectopic beats, and signal noise <15% and <30 min recording break) workdays and days off (Fig. 1).

Background information included age, sex, questionnaire-reported height, weight, and PA class (9), modified from Ross and Jackson (28). Then maximal heart rate ($210 - 0.65 \times$ age) (16) and $\dot{V}O_{2max}$ (men $67.350 + 1.921 \times$ PA class $- 0.381 \times$ age $- 0.754 \times$ BMI; women $56.363 + 1.921 \times$ PA class $- 0.381 \times$ age $- 0.754 \times$ BMI) (15) were estimated, and these values were further used in the estimation of oxygen uptake ($\dot{V}O_2$). If a period with a heart rate higher than the estimate was found from the recording, the maximal heart rate used for further calculations was corrected accordingly. BMI ($kg \cdot m^{-2}$) was calculated from the self-reported weight and height.

The intensity in terms of $\dot{V}O_2$ and volume of PA was first estimated based on the R-R interval recordings (10,17,18,30). The method has been validated previously; the pooled relationship (correlation) between the measured and the predicted $\dot{V}O_2$ across the different activities of daily living was 0.93, and the estimated $\dot{V}O_2$ explained 87% of the variability in the measured $\dot{V}O_2$ (32). In another validation study, Robertson et al. (29) showed that the energy expenditure estimates based on our method correlate strongly with those based on indirect calorimetry across analysis conditions ($r = 0.85$–0.98). The high validity of this method was achieved by taking into account the R-R interval–derived information about heart rate, respiration rate, and on/off response (increasing or decreasing heart rate) using neural network modeling of the data and the short-time Fourier transform method (10,17,18,30).

Participant mean $\dot{V}O_2$ for each minute was calculated from the second-by-second $\dot{V}O_2$ estimations. For the calculation of the volume of absolutely determined PA, the minute-by-minute $\dot{V}O_2$ estimations were converted to METs by dividing the $\dot{V}O_2$ values by a resting metabolic rate value of $3.5\ mL \cdot kg^{-1} \cdot min^{-1}$. On the basis of the MET values, the volume of MPA and VPA ($min \cdot d^{-1}$) at each intensity level was then calculated. These data are called MPA_{Abs} and VPA_{Abs} later in the text. The thresholds for these categories were MPA_{Abs} 3–<6 METs and VPA_{Abs} ≥6 METs (11). MVPA then refers to the sum of MPA and VPA, respectively.

The intensity of PA was also calculated relatively, i.e., in relation to estimated $\dot{V}O_{2max}$. The relative intensity was determined using the percentage of $\dot{V}O_2$ reserve (% $\dot{V}O_2R$). $\dot{V}O_2R$ is calculated by subtracting 1 MET ($3.5\ mL \cdot kg^{-1} \cdot min^{-1}$) from the $\dot{V}O_{2max}$. The % $\dot{V}O_2R$ is calculated by subtracting 1 MET from the measured $\dot{V}O_2$, dividing by the $\dot{V}O_2R$, and multiplying by 100% (14). The amount of PA ($min \cdot d^{-1}$) at different intensity levels (moderate and vigorous) was then calculated. These values are called MPA_{Rel} and VPA_{Rel} later in the text. The thresholds for these categories were MPA_{Rel} 40%–<60% $\dot{V}O_2R$ and VPA_{Rel} ≥60% $\dot{V}O_2R$ (11,14). Again, MVPA refers to the sum of MPA and VPA, respectively.

As the general use of resting metabolic rate value of $3.5\ mL \cdot kg^{-1} \cdot min^{-1}$ to calculate PA METs for individuals with differing sex, age, and BMI may cause misclassification of activities (20), we also recalculated the main results using the original Harris–Benedict formula (13). Results in our article and Supplemental Digital Content 1, http://links.lww.com/MSS/A810, are based on the generally used $3.5\ mL \cdot kg^{-1} \cdot min^{-1}$ for resting metabolic rate and those in the Supplemental Digital Content 2, http://links.lww.com/MSS/A811, on calculating the resting metabolic rates using the Harris–Benedict formula.

Statistical analysis. Data processing and statistical analysis were performed using MATLAB version R2015b (The MathWorks Inc., Natick, MA) and R version 3.2.2 (The R Foundation for Statistical Computing, Vienna, Austria).

We calculated mean, SD, and 95% confidence intervals (CI) for continuous variables. First, the total number of 1-min segments in each intensity category during each measurement day for each individual was calculated. If a participant's measurement period included two or more workdays (or days off), an average was calculated. We also calculated the mean daily absolute and relative intensity PA minutes covering both workdays and days off. Then we calculated the amount of MPA_{Abs}, MPA_{Rel}, VPA_{Abs}, and VPA_{Rel} by gender and type of day (i.e., workdays vs days off) for different age (18–30, 31–40, 41–50, and 51–65 yr) and BMI categories (normal weight, 18.5–<25.0 $kg \cdot m^{-2}$; overweight, 25.0–<30.0 $kg \cdot m^{-2}$; and obese, 30.0–40.0 $kg \cdot m^{-2}$). The absolute and relative PA volumes at different intensity levels were compared inside each age and BMI category using the Wilcoxon signed rank test. Differences in the absolute and relative PA volumes between age and BMI categories were analyzed using the Kruskal–Wallis test.

We then calculated how the determined absolute and relative intensity PA minutes overlapped (Figs. 2 and 3). In addition, we calculated at the group level the proportions between VPA_{Abs}/VPA_{Rel} in specific subgroups (Fig. 4). Because of the complexity of the relations between the absolute and the relative intensity minutes, the 95% CI values for the relations were calculated using a percentile bootstrapping method. All P values reported are two-sided, and because of the large sample size, the significance level was set to 0.001.

RESULTS

Most of the R-R interval recordings were from values taken over 3 d (13,052 participants); 7062, 1327, 936, and 847 participants had 2, 4, 5, and 6 measurement days, respectively. Altogether, the number of analyzed days was 39,904 for workdays and 28,446 for days off (Fig. 1). The mean ± SD age of the participants was 44.7 ± 9.8 yr (men = 44.4 ± 9.9 yr, women = 45.0 ± 9.8 yr), and the mean ± SD BMI was 26.0 ± 4.0 $kg \cdot m^{-2}$ (men = 26.6 ± 3.5 $kg \cdot m^{-2}$, women = 25.5 ± 4.4 $kg \cdot m^{-2}$) (Table 1).

EPIDEMIOLOGY

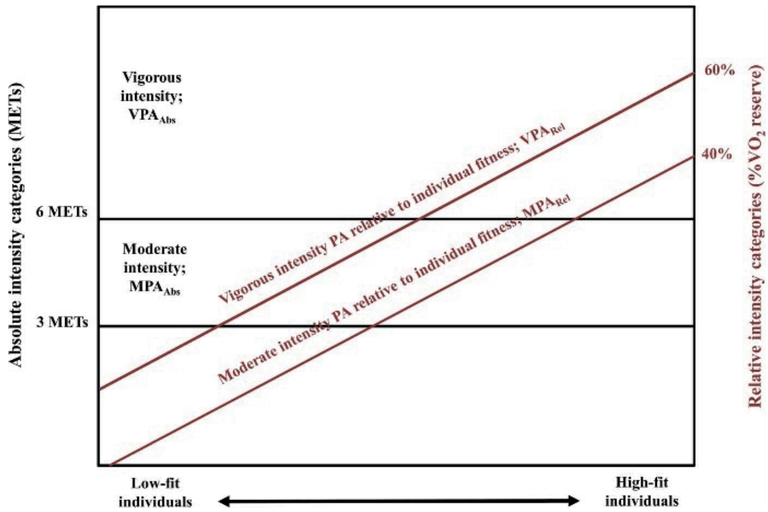


FIGURE 2—Illustration of the theoretical overlap between absolute PA intensity versus PA intensity relative to individual aerobic fitness level. % $\dot{V}O_2$ reserve, percentage of maximal oxygen uptake reserve.

Heart rates and estimated $\dot{V}O_{2max}$. Mean heart rates did not differ substantially between age-groups or between workdays and days off. However, the mean heart rates increased with increasing BMI among both men and women during both workdays and days off, covering time awake and sleeping time (see Table 1 in Supplemental Digital Content 1, Mean heart rates by age, sex, and type of day during whole recording day, http://links.lww.com/MSS/A810). For the estimated mean $\dot{V}O_{2max}$ values by sex, age, and BMI categories, see Table 2 in Supplemental Digital Content 1 (Mean estimated $\dot{V}O_{2max}$ values by sex, age and weight group, http://links.lww.com/MSS/A810).

PA volumes by sex, age, and type of day. According to absolute intensity, as expected, men had higher values for MPA and VPA minutes compared with women (Table 2). The mean values for the $MVPA_{Abs}$ minutes during workdays were 50.5 (95% CI = 49.5–51.4) for men and 33.2 (95% CI 32.6–33.8) for women ($P < 0.001$); during days off, they were 63.7 (62.5–64.9) and 34.7 (34.0–35.4) ($P < 0.001$), respectively [see Tables 3 and 4 in Supplemental Digital Content 1, Amount of absolute and relative moderate and vigorous intensity physical activity (min·d^{-1}) by age groups during workdays and days off among men; Amount of absolute and relative moderate and vigorous intensity physical activity (min·d^{-1}) by age-groups during workdays and days off among women, http://links.lww.com/MSS/A810]. In particular, VPA volumes were low for women (Table 2). However, when calculated as intensity levels relative to individual fitness, the PA volumes

EPIDEMIOLOGY

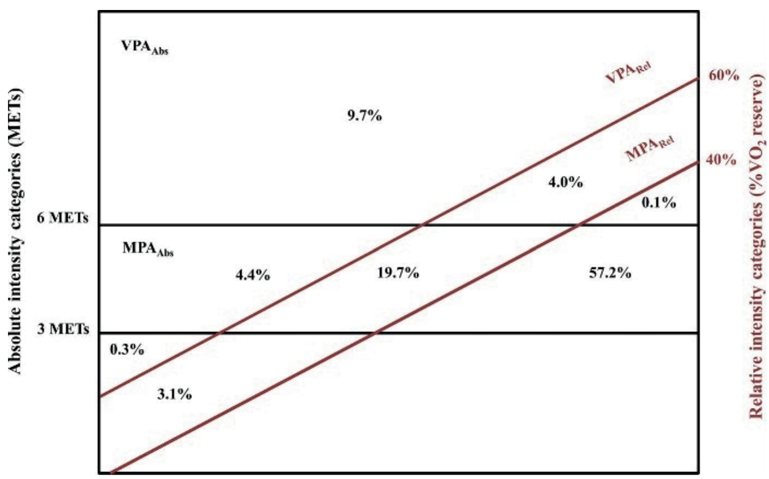


FIGURE 3—Overlap (mean percent of PA minutes falling in different intensity categories) between absolute PA intensity versus PA intensity relative to individual aerobic fitness level. For abbreviation, see Figure 2.

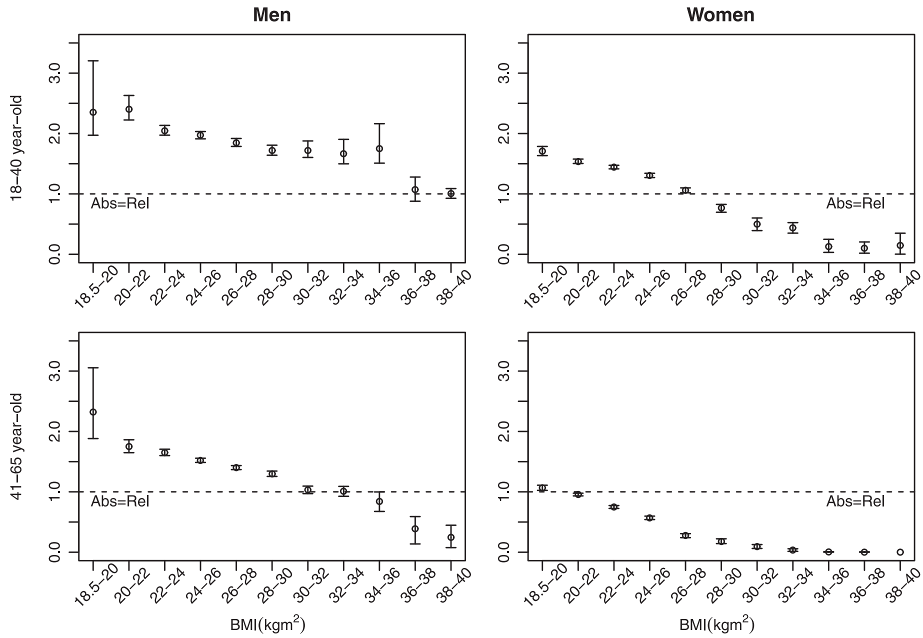


FIGURE 4—Daily VPA_{Abs}/VPA_{Rel} minute ratios among older and younger men and women by BMI categories. VPA_{Abs} = daily minutes of absolute vigorous-intensity (≥6 METs) PA. VPA_{Rel} = daily minutes of VPA relative to individual aerobic fitness (≥60% of oxygen uptake reserve). Error bars represent 95% CI.

for women were about at the same level as for men or even higher (Table 2): mean daily minutes of $MVPA_{Rel}$ were 16.2 for men and 17.3 for women.

Men older than 30 yr had higher $MVPA_{Abs}$ and $MVPA_{Rel}$ minutes during days off than during workdays, but these differences were not as strong in women and in younger men. PA volume in terms of absolute intensity decreased by age among both women and men, during both workdays and days off, but a similar strong age-related reduction of PA was not seen when the intensity was calculated relative to individual fitness level [Table 2, see also Tables 3 and 4 in Supplemental Digital Content 1, Amount of absolute and relative moderate and vigorous intensity physical activity (min·d^{-1}) by age groups during workdays and days off among men; Amount of absolute and relative moderate and vigorous intensity physical activity (min·d^{-1}) by age groups during workdays and days off among women, http://links.lww.com/MSS/A810]. Among the oldest women (51–65 yr), the amount of VPA_{Rel} was higher compared with VPA_{Abs} ($P < 0.001$).

The overlap of the minutes that fulfilled the criteria for either $MVPA_{Abs}$ or $MVPA_{Rel}$ in different absolute and relative intensity categories is shown in Figure 3. An average

EPIDEMIOLOGY

TABLE 1. Number of participants by sex, age, and weight status.

		Weight Status		
		Normal Weight	Overweight	Obese
Age Group	*n*	*n* (%)		
18–30 yr				
Men	940	524 (55.7)	349 (37.1)	67 (7.1)
Women	1148	873 (76.0)	199 (17.3)	76 (6.6)
31–40 yr				
Men	2794	1115 (39.9)	1321 (47.3)	358 (12.8)
Women	3087	1995 (64.6)	751 (24.3)	341 (11.0)
41–50 yr				
Men	3333	1048 (31.4)	1708 (51.2)	577 (17.3)
Women	4512	2318 (51.4)	1386 (30.7)	808 (17.9)
51–65 yr				
Men	3134	968 (30.9)	1646 (52.5)	520 (16.6)
Women	4276	1887 (44.1)	1511 (35.3)	878 (20.5)
Total				
Men	10,201	3655 (35.8)	5024 (49.3)	1522 (14.9)
Women	13,023	7073 (54.3)	3847 (29.5)	2103 (16.1)

Normal weight = BMI 18.5–<25.0 kg·m^{-2}.
Overweight = BMI 25.0–<30.0 kg·m^{-2}.
Obese = BMI 30.0–40.0 kg·m^{-2}.

TABLE 2. Mean[a] daily amount of absolute and relative MPA and VPA ($min \cdot d^{-1}$) by age groups and by weight status among men and women.

	MPA_{Abs}	MPA_{Rel}	VPA_{Abs}	VPA_{Rel}	$MVPA_{Abs}$	$MVPA_{Rel}$
	Mean (95% CI), min					
Activity volumes by age groups						
Men						
18–30 yr	71.3 (68.3–74.2)	9.1 (8.5–9.6)	17.9 (16.8–18.9)	7.8 (7.1–8.4)	89.1 (85.7–92.6)	16.9 (15.8–17.9)
31–40 yr	49.2 (47.9–50.4)	8.4 (8.1–8.7)	14.5 (13.9–15.1)	8.0 (7.6–8.4)	63.7 (62.1–65.3)	16.4 (15.8–17.0)
41–50 yr	41.4 (40.4–42.4)	8.9 (8.5–9.2)	12.0 (11.5–12.5)	7.7 (7.3–8.1)	53.4 (52.2–54.6)	16.5 (16.0–17.1)
51–65 yr	37.2 (36.2–38.2)	9.2 (8.9–9.5)	8.2 (7.8–8.7)	6.3 (6.0–6.7)	45.4 (44.3–46.6)	15.5 (14.9–16.1)
Women						
18–30 yr	55.1 (53.3–56.8)	12.5 (11.9–13.0)	15.6 (14.7–16.4)	9.5 (8.9–10.1)	70.6 (68.4–72.8)	21.9 (21.0–22.9)
31–40 yr	34.1 (33.3–34.9)	9.5 (9.2–9.8)	8.4 (8.0–8.8)	7.0 (6.7–7.3)	42.5 (41.5–43.5)	16.5 (16.0–17.0)
41–50 yr	25.2 (24.7–25.7)	9.5 (9.3–9.8)	5.1 (4.8–5.4)	6.8 (6.5–7.0)	30.3 (29.6–31.0)	16.3 (15.9–16.7)
51–65 yr	19.8 (19.3–20.3)	10.7 (10.4–11.0)	2.0 (1.8–2.2)	6.8 (6.5–7.1)	21.8 (21.2–22.4)	17.5 (17.0–18.1)
Activity volumes by weight status						
Men						
Normal	51.6 (50.4–52.8)	9.2 (8.9–9.5)	16.0 (15.5–16.6)	8.8 (8.4–9.2)	67.6 (66.2–69.1)	18.0 (17.5–18.6)
Overweight	43.0 (42.1–43.8)	8.6 (8.3–8.8)	11.0 (10.6–11.4)	7.0 (6.7–7.3)	54.0 (53.0–55.0)	15.6 (15.1–16.0)
Obese	35.8 (34.3–37.3)	8.8 (8.4–9.3)	6.1 (5.5–6.6)	5.2 (4.7–5.7)	41.9 (40.2–43.6)	14.0 (13.2–14.8)
Women						
Normal	35.3 (34.7–35.9)	10.4 (10.2–10.5)	9.1 (8.8–9.3)	8.2 (7.9–8.4)	44.4 (43.7–45.1)	18.5 (18.2–18.9)
Overweight	22.8 (22.2–23.4)	9.2 (9.0–9.5)	2.6 (2.4–2.7)	5.0 (4.8–5.3)	25.3 (24.7–26.0)	14.2 (13.8–14.7)
Obese	14.0 (13.4–14.6)	11.3 (10.8–11.8)	0.6 (0.4–0.7)	7.1 (6.6–7.6)	14.6 (14.0–15.2)	18.4 (17.5–19.3)
All men	45.0 (44.4–45.6)	8.9 (8.7–9.0)	12.1 (11.8–12.4)	7.4 (7.2–7.6)	57.1 (56.3–57.9)	16.2 (15.9–16.6)
All women	28.2 (27.8–28.5)	10.2 (10.0–10.3)	5.8 (5.6–5.9)	7.1 (6.9–7.2)	33.9 (33.5–34.4)	17.2 (17.0–17.5)

[a]Mean of all monitored work days and days off.
There was a statistically significant difference between absolute and relative intensity physical activity ($P < 0.001$, Wilcoxon signed rank test) in all age-groups and weight status groups and intensity categories.
Except for MPA_{Rel} in men ($P = 0.006$), there was a statistically significant difference ($P < 0.001$, Kruskal–Wallis test) in all intensity categories between age-groups.
There was a statistically significant difference ($P < 0.001$, Kruskal–Wallis test) in all activity types between weight status groups.
MPA_{Abs} = MPA, absolutely determined intensity: 3.0–<6.0 metabolic equivalents (METs); MPA_{Rel} = MPA, relatively determined intensity: 40%–<60% oxygen uptake reserve ($\dot{V}O_2R$); VPA_{Abs} = VPA, absolutely determined intensity: ≥6.0 METs; VPA_{Rel} = VPA, relatively determined intensity: ≥60% $\dot{V}O_2R$.
Normal = BMI 18.5–<25.0 $kg \cdot m^{-2}$.
Overweight = BMI 25.0–<30.0 $kg \cdot m^{-2}$.
Obese = BMI 30.0–40.0 $kg \cdot m^{-2}$.

of 9.7% (95% CI = 9.5–9.9) of these minutes were vigorous and 19.7% (95% CI = 19.4–20.0) were moderate intensity, according to both absolute and relative criteria.

PA volumes and BMI. The amount of PA in terms of absolute intensity decreased with increasing BMI among both women and men. However, a similar strong BMI-related reduction of PA was not seen when the intensity was calculated relative to individual fitness level [Table 2, see also Tables 5 and 6 in Supplemental Digital Content 1, Amount of absolute and relative moderate- and vigorous-intensity physical activity ($min \cdot d^{-1}$) by weight status during workdays and days off among men; Amount of absolute and relative moderate- and vigorous-intensity physical activity ($min \cdot d^{-1}$) by weight status during workdays and days off among women, http://links.lww.com/MSS/A810]. In men, the volume of MPA_{Rel} was significantly lower than that of MPA_{Abs} ($P < 0.001$) in each BMI category. Except for obese men during days off, the amount of VPA_{Rel} was also significantly ($P < 0.001$) lower than that of VPA_{Abs} in each BMI category among men.

Among women, the volume of MPA_{Rel} was significantly ($P < 0.001$) lower than that of MPA_{Abs} in each BMI category. However, among overweight and obese women, the amount of VPA_{Rel} was significantly ($P < 0.001$) higher than that of VPA_{Abs}. Of note, among obese women, 93% during workdays and 95% during days off had no VPA_{Abs}, with percentages of 41% and 54% for no VPA_{Rel}, respectively.

Mean daily VPA_{Abs} minutes were higher than VPA_{Rel} minutes in younger (18–40 yr) normal weight men (17.7, 95% CI = 16.9–18.6, vs 8.6, 95% CI = 8.0–9.1; $P < 0.001$). The reverse was the case for older (41–65 yr) obese women (mean 0.3, 95% CI = 0.2–0.4, vs 7.8, 95% CI = 7.2–8.4; $P < 0.001$).

Figure 4 shows the relations between all VPA_{Rel}/VPA_{Abs} minutes among younger (18–40 yr) and older (41–65 yr) men and women split into smaller BMI categories. The number of VPA_{Rel} minutes was higher than the number of VPA_{Abs} minutes among older men with BMI more than 34 $kg \cdot m^{-2}$, younger women with BMI more than 28 $kg \cdot m^{-2}$, and older women with BMI more than 20 $kg \cdot m^{-2}$. Concerning the overlap of the absolute vs relative intensity minute categories (Fig. 3), among normal weight men, there were no minutes in the category below MPA_{Abs} and $MVPA_{Rel}$, but an average of 27.7% of the recorded PA minutes fell into these categories among obese women.

When reanalyzing the main results using resting metabolic rate calculated individually according to the Harris–Benedict formula, expectedly, the resting metabolic rates were lower than 3.5, in particular among aged, obese females [for details see Table 1 in Supplemental Digital Content 2, Mean estimated resting metabolic rates ($\dot{V}O_2$ values) by sex, age and weight groups according to Harris–Benedict formula, http://links.lww.com/MSS/A811]. In this reanalysis, the number of absolute PA minutes was higher [please, compare Table 2 in the Supplemental Digital Content 2 (Mean estimated resting metabolic rates ($\dot{V}O_2$ values) by sex, age and weight groups according to Harris–Benedict formula) vs. Table 2 in the manuscript, http://links.lww.com/MSS/A811]. The overlap of the minutes that fulfilled

EPIDEMIOLOGY

the criteria for either $MVPA_{Abs}$ or $MVPA_{Rel}$ in different absolute and relative intensity categories was lower than that in our primary analysis [see Figure 1 in Supplemental Digital Content 2 (Overlap between absolute PA intensity vs. PA intensity relative to individual aerobic fitness level) vs. Figure 3, http://links.lww.com/MSS/A811], and the number of VPA_{Rel} minutes persisted higher than the number of VPA_{Abs} minutes in particular among older women with high BMI.

DISCUSSION

Objectively measured absolute volumes of MVPA were higher in men compared with women in this study, and higher in younger compared with older and in normal weight compared with obese individuals. When the MPA and VPA volumes were categorized according to % $\dot{V}O_2R$, the differences were not as stark. The mean cardiovascular strain, when indicated with mean heart rate during all of the recording days, was higher among individuals with higher BMI.

As many of our participants were physically fit and usually healthy employed individuals, the absolute PA volumes were higher compared with relative volumes in many of the studied age and body weight categories (Table 2). The findings on the ratios between absolute versus relative PA volumes according to age and sex reflect the previously reported international (31) and Finnish (24) results on the distribution of measured population fitness. It was rather easy for high-fit individuals to reach MVPA intensity levels according to absolute criteria compared with criteria relative to individual fitness level, but the situation reversed among the unfit individuals. This phenomenon was similar during days off and workdays. Our findings are in line with data from U.S. adults showing that VPA determined in absolute terms using an accelerometry method is low among older, female, and obese individuals (36).

Because we are not aware of large population studies of overlap between absolute and relative intensity PA volumes, we cannot compare our results with other studies. In PA counseling, the intensity of PA should be tailored individually (11), but this recommendation is often disregarded in practice. When trying to find effective solutions to increase PA among physically inactive low-fit individuals, we need to take into account the individual fitness level to focus on behavior changes that are easy to adopt and sustain rather than expect people to become highly motivated to adopt and maintain effortful behaviors that include too vigorous PA. In PA counseling, appropriate PA intensity may be guided most easily by using simple terms describing exercise intensity or using Borg Rating of Perceived Exertion scale (from 6 to 20) (3), where 12–13 indicates MPA and 14 or more vigorous PA. Heart rate monitoring during exercise would be a more accurate alternative to guide and control target exercise intensity. Individuals with severe chronic disease usually need personal advice from health care professionals and related to medical clearance for exercise participation/rehabilitation the intensity levels are usually given in a way proper for each condition, such as determination of symptom-free exercise intensities among patients with heart disease. In a previous real-life lifestyle intervention, increasing PA was a good indicator of success in improving the cardiovascular and metabolic risk factor levels, including body weight reduction (23).

Strengths and limitations. Although we did not have direct oxygen uptake recordings in our large real-life data, we used a validated ambulatory method to assess the intensity of PA. This method provides more accurate estimates of the intensity of PA compared with heart rate information only (32). The use of % $\dot{V}O_2R$ is relatively valid also among obese individuals (5) and patients with heart disease (4). However, there are no specific validation studies comparing the validity of our methodology between representative BMI and age groups. Our recordings had good coverage of typical workdays and days off.

Our study was a cross-sectional study. Randomized controlled trials are needed to confirm that PA recommended with guidelines applying subjective intensity levels is more feasible in the long-term compared with those using absolute intensity levels in the important target group of low-fit formerly inactive overweight and obese individuals. Although MPA relative to individual fitness level improves fitness (35) and other cardiometabolic risk factor levels (19) among low-fit individuals, a comprehensive understanding about which PA intensity is most beneficial for health in the long term is still lacking.

CONCLUSIONS

Compared with low-fit individuals, it is easier for high-fit individuals to reach MVPA intensity levels according to absolute criteria and easier for men compared with women, younger people compared with older, and lean compared with obese individuals. When the target is set as relative intensity, the frequency of reaching the target is more similar in low- and high-fit individuals. Thus, when boosting MVPA in inactive, low-fit, and/or obese individuals, intensity guidelines relative to individual fitness may be more feasible than using recommended absolute intensity classifications. Because PA counseling is suggested to be made a priority in clinical practice (2,21), our findings should be considered when taking action in the most important target group of low-fit individuals. Also, our findings need to be taken into account when interpreting the results of population studies that have used accelerometer-based monitoring of PA volumes with absolute criteria.

This study was supported by TEKES—the Finnish Funding Agency for Technology and Innovation (grant no. 40116/14).

Myllymäki reports being employed also by Firstbeat Technologies Ltd., Jyväskylä, Finland, and Korhonen reports being employed also by PulseOn Oy, Espoo, Finland. Authors have no other conflicts of interest to declare.

The results of the present study do not constitute endorsement by the American College of Sports Medicine.

The authors declare that the results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

REFERENCES

1. Bergman P, Grjibovski AM, Hagströmer M, Bauman A, Sjöström M. Adherence to physical activity recommendations and the influence of socio-demographic correlates - a population-based cross-sectional study. *BMC Public Health*. 2008;8:367.
2. Berra K, Rippe J, Manson JE. Making physical activity counseling a priority in clinical practice: the time for action is now. *JAMA*. 2015;314(24):2617–8.
3. Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc*. 1982;14(5):377–81.
4. Brawner CA, Keteyian SJ, Ehrman JK. The relationship of heart rate reserve to $\dot{V}O_2$ reserve in patients with heart disease. *Med Sci Sports Exerc*. 2002;34(3):418–22.
5. Byrne NM, Hills A. Relationships between HR and $\dot{V}O_2$ in the obese. *Med Sci Sports Exerc*. 2002;34(9):1419–27.
6. Clarke J, Janssen I. Sporadic and bouted physical activity and the metabolic syndrome in adults. *Med Sci Sports Exerc*. 2014;46(1):76–83.
7. Colley RC, Garriguet D, Janssen I, Craig CL, Clarke J, Tremblay MS. Physical activity of Canadian adults: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. *Health Rep*. 2011;22(1):7–14.
8. Ekkekakis P, Parfitt G, Petruzzello SJ. The pleasure and displeasure people feel when they exercise at different intensities: decennial update and progress towards a tripartite rationale for exercise intensity prescription. *Sports Med*. 2011;41(8):641–71.
9. Firstbeat Technologies Ltd. Web site [Internet]. EPOC based training effect assessment, 2012. http://www.firstbeat.com/userData/firstbeat/download/white_paper_training_effect.pdf. Accessed April 6, 2016.
10. Firstbeat Technologies Ltd. Web site [Internet]. $\dot{V}O_2$ estimation method based on heart rate measurement, 2012. http://www.firstbeat.com/userData/firstbeat/download/white_paper_vo2_estimation.pdf. Accessed April 6, 2016.
11. Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine Position Stand: 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*. 2011;43(7):1334–59.
12. Glazer NL, Lyass A, Esliger DW, et al. Sustained and shorter bouts of physical activity are related to cardiovascular health. *Med Sci Sports Exerc*. 2013;45(1):109–15.
13. Harris JA, Benedict FG. A biometric study of human basal metabolism. *Proc Natl Acad Sci U S A*. 1918;4(12):370–3.
14. Howley ET. Type of activity: resistance, aerobic and leisure versus occupational physical activity. *Med Sci Sports Exerc*. 2001; 33(6 Suppl):S364–9.
15. Jackson AS, Blair SN, Mahar MT, Wier LT, Ross RM, Stuteville JE. Prediction of functional aerobic capacity without exercise testing. *Med Sci Sports Exerc*. 1990;22(6):863–70.
16. Jones NL. *Clinical Exercise Testing*. 3rd ed. Philadelphia: W.B. Saunders; 1988. p. 307.
17. Kettunen J, Saalasti S, inventors; Firstbeat Technologies Ltd., assignee. *Procedure for Deriving Reliable Information on Respiratory Activity from Heart Period Measurement*. United States Patent US 7,460,901 B2, 2008. http://worldwide.espacenet.com/publicationDetails/originalDocument?CC=US&NR=7460901B2&KC=B2&FT=D&ND=4&date=20081202&DB=EPODOC&locale=fi_FI. Accessed April 6, 2016.
18. Kettunen J, Saalasti S, inventors; Firstbeat Technologies Ltd., assignee. *Procedure for Detection of Stress by Segmentation and Analyzing Heart Beat Signal*. United States Patent US 7,330,752 B2, 2008. http://worldwide.espacenet.com/publicationDetails/originalDocument?FT=D&date=20080212&DB=EPODOC&locale=fi_FI&CC=US&NR=7330752B2&KC=B2&ND=4. Accessed April 6, 2016.
19. King AC, Haskell WL, Young DR, Oka RK, Stefanick ML. Long-term effects of varying intensities and formats of physical activity on participation rates, fitness, and lipoproteins in men and women aged 50 to 65 years. *Circulation*. 1995;91(10):2596–604.
20. Kozey S, Lyden K, Staudenmayer J, Freedson P. Errors in MET estimates of physical activities using 3.5 ml x kg(-1) x min(-1) as the baseline oxygen consumption. *J Phys Act Health*. 2010;7(4):508–16.
21. Kraus WE, Bittner V, Appel L, et al. The National Physical Activity Plan: a call to action from the American Heart Association: a science advisory from the American Heart Association. *Circulation*. 2015;131(21):1932–40.
22. Kujala UM. Evidence on the effects of exercise therapy in the treatment of chronic disease. *Br J Sports Med*. 2009;43(8):550–5.
23. Kujala UM, Jokelainen J, Oksa H, et al. Increase in physical activity and cardiometabolic risk profile change during lifestyle intervention in primary healthcare: 1-year follow-up study among individuals at high risk for type 2 diabetes. *BMJ Open*. 2011;1(2):e000292.
24. Kujala UM, Viljanen T, Taimela S, Viitasalo J. Physical activity, $\dot{V}O_2$max, and jumping height in an urban population. *Med Sci Sports Exerc*. 1994;26(7):889–95.
25. Mutikainen S, Helander E, Pietilä J, Korhonen I, Kujala UM. Objectively measured physical activity in Finnish employees: a cross-sectional study. *BMJ Open*. 2014;4(12):e005927.
26. Pedersen BK, Saltin B. Exercise as medicine—evidence for prescribing exercise as therapy in 26 different chronic diseases. *Scand J Med Sci Sports*. 2015;25(3 Suppl):1–72.
27. Physical Activity Guidelines Advisory Committee. *Physical Activity Guidelines Advisory Committee Report, 2008*. Washington (DC): U.S. Department of Health and Human Services; 2008.
28. Ross RM, Jackson AS. *Exercise Concepts, Calculations, and Computer Applications*. Carmel (IN): Benchmark Press; 1990. p. 109.
29. Robertson AH, King K, Ritchie SD, Gauthier AP, Laurence M, Dorman SC. Validating the use of heart rate variability for estimating energy expenditure. *Int J Human Movement Sports Sci*. 2015;3(2):19–26.
30. Saalasti S. Neural networks for heart rate time series analysis [dissertation]. Jyväskylä (Finland): University of Jyväskylä; 2003 https://jyx.jyu.fi/dspace/bitstream/handle/123456789/13267/951391707X.pdf?sequence=1. Accessed April 6, 2016. 192 p.
31. Shvartz E, Reibold RC. Aerobic fitness norms for males and females aged 6 to 75 years: a review. *Aviat Space Environ Med*. 1990;61(1):3–11.
32. Smolander J, Ajoviita M, Juuti T, Nummela A, Rusko H. Estimating oxygen consumption from heart rate and heart rate variability without individual calibration. *Clin Physiol Funct Imaging*. 2011;31(4):266–71.
33. Strath SJ, Holleman RG, Ronis DL, Swartz AM, Richardson CR. Objective physical activity accumulation in bouts and nonbouts and relation to markers of obesity in US adults. *Prev Chronic Dis*. 2008;5(4):A131.
34. Strath SJ, Kaminsky LA, Ainsworth BE, et al. Guide to the assessment of physical activity: clinical and research applications: a scientific statement from the American Heart Association. *Circulation*. 2013;128(20):2259–79.
35. Swain DP, Franklin BA. VO(2) reserve and the minimal intensity for improving cardiorespiratory fitness. *Med Sci Sports Exerc*. 2002;34(1):152–57.
36. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S.: adults compliance with the Physical Activity Guidelines for Americans. *Am J Prev Med*. 2011;40(4):454–61.
37. Vähä-Ypyä H, Vasankari T, Husu P, et al. Validation of cut-points for evaluating the intensity of physical activity with accelerometry-based mean amplitude deviation (MAD). *PLoS One*. 2015;10(8):e0134813.
38. White DK, Gabriel KP, Kim Y, Lewis CE, Sternfeld B. Do short spurts of physical activity benefit cardiovascular health? The CARDIA Study. *Med Sci Sports Exerc*. 2015;47(11):2353–8.
39. World Health Organization. *Global Recommendations on Physical Activity for Health*. Geneva: World Health Organization; 2010. pp. 24–27.

PUBLICATION III

Exploratory Analysis of Associations Between Individual Lifestyles and Heart Rate Variability-Based Recovery During Sleep

PIETILÄ, J., HELANDER, E., MYLLYMÄKI, T., KORHONEN, I., JIMISON, H. and PAVEL, M.

37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2339-2342
DOI: 10.1109/EMBC.2015.7318862

Exploratory analysis of associations between individual lifestyles and heart rate variability –based recovery during sleep

Julia Pietilä, Elina Helander, Tero Myllymäki, Ilkka Korhonen, *Senior Member, IEEE,* Holly Jimison and Misha Pavel, *Member, IEEE*

***Abstract*—Sleep is the most important period for recovering from daily stress and load. Assessment of the stress recovery during sleep is therefore, an important metric for care and quality of life. Heart rate variability (HRV) is a non-invasive marker of autonomic nervous system (ANS) activity, and HRV–based methods can be used to assess physiological recovery, characterized by parasympathetic domination of the ANS. HRV is affected by multiple factors of which some are unmodifiable (such as age and gender) but many are related to daily lifestyle choices (e.g. alcohol consumption, physical activity, sleeping times). The purpose of this study was to investigate the association of these aforementioned factors on HRV-based recovery during sleep on a large sample. Variable importance measures yielded by random forest were used for identifying the most relevant predictors of sleep-time recovery. The results emphasize the disturbing effects of alcohol consumption on sleep-time recovery. Good physical fitness is associated to good recovery, but acute physical activity seems to challenge or delay the recovery process for the next night. Longer sleeping time enables more recovery minutes, but the proportion of recovery (i.e. recovery efficiency) seems to peak around 7.0-7.25 hours of sleep.**

I. Introduction

Stress recovery is crucial in avoiding the adverse health consequences of prolonged stress. Even though the body can recover during waking hours, the sleep is unquestionably the most important recovery period [1]. Recovery is a complex phenomenon including both physiological and psychological aspects [2]. Physiologically, recovery is manifested by vagal dominance of the autonomic nervous system (ANS) when stress factors that cause disturbance of body's homeostasis are not present. Stress, on the other hand, causes elevated activation of the physiological systems and sympathetic domination of the ANS. [3] An indicator for the ANS balance is heart rate variability (HRV). High HRV reflects vagal dominance of the ANS, while low HRV reflects sympathetic dominance of the ANS. [4] Thus, HRV can be used as a method to assess physiological recovery and stress.

Even though HRV is widely accepted as a non-invasive measure for the ANS activity, it is affected also by several other factors [4]. HRV is associated with a person's background characteristics, such as age and gender. Men have higher HRV than women but the difference in HRV between genders decreases with age. In general, HRV decreases with age and the decrease is more pronounced in nocturnal HRV. [5] The circadian pattern of HRV shows HRV is higher during sleep than awake due to dominance of parasympathetic over sympathetic activity, but for example, psychophysiological stress can diminish HRV during sleep [6]. Moreover, better aerobic fitness has been associated with increased HRV [7], whereas alcohol intake decreases HRV acutely [4].

The purpose of this study was to examine how daily lifestyle choices and personal background characteristics are associated with recovery during next night sleep. This paper presents the methodology used for quantifying daily activities and sleep and the use of these quantified variables in the random forest (RF) analysis. A key result of this paper is, therefore, the estimation of the importance and effect of the daily activities and personal background characteristics on recovery during sleep.

II. Methods

A. Data

The data used in this study were provided by Firstbeat Technologies Ltd (Jyväskylä, Finland, www.firstbeat.fi), a Finnish company providing analytics for well-being factors such as stress, recovery, and physical activity based on ambulatory measurement of beat-to-beat heart rate and subsequent analysis of HRV. The anonymized database included measurement results of Finnish working-age subjects who had voluntarily participated in measurements as part of their preventive occupational health care services. Use of the database for research purposes has been approved by the Ethics Committee of Tampere University Hospital.

The HRV data were typically collected over three days in real-life conditions. During the measurement, the subjects were asked to keep a diary of their working and sleeping times as well as the amount of alcohol consumed. As background information, the subjects provided gender, age, weight, height, and also described their physical activity. Based on the amount and intensity of their physical activity, a physical activity class varying from 0 (very inactive) to 10 (high-level athlete) was determined.

Posterior inclusion criteria were applied. The subjects' age was limited to 18−65 years and body-mass index (BMI) to 18.5−40 kg/m2. Because shift-work has been reported to affect HRV during sleep [8] the subjects who had been working between 9 pm and 6 am were removed from the analysis. Only those measurement days having less than 15% of corrected HRV data artefact and lasting longer than

J. Pietilä, E. Helander and I. Korhonen are with the Department of Signal Processing, Tampere University of Technology, Tampere, 33720 Finland (tel: +358405675334; e-mails: julia.pietila@tut.fi, elina.helander@tut.fi, ilkka.korhonen@tut.fi).

T. Myllymäki is with Firstbeat Technologies Ltd, Jyväskylä, Finland (email: tero.myllymaki@firstbeat.fi).

H. Jimison and M. Pavel are with College of Computer & Information Science / Bouvé College of Health Sciences, Northeastern University, Boston, MA 02115 USA (emails: h.jimison@neu.edu, m.pavel@neu.edu).

16 hours were analyzed. Moreover, the measurement days had to include the diary entries about the sleeping period and the possible consumption of alcohol. The days were grouped into workdays and days off according to the duration of working period. For workdays there had to be at least four hours of work, while no working was allowed for days off.

Altogether, the subset of the database which fulfilled our inclusion criteria included 12,023 measurement days. This comprised 7,044 workdays and 4,979 days off from 6,288 subjects (2,788 men (44.3 %), age 45.6±9.6 years, BMI 26.4±4.1 kg/m^2, physical activity class 4.8±1.9)).

B. HRV analysis

The physiological states such as level of physical activity, stress and recovery were determined using an algorithm relating HRV to these factors [9]. For temporal estimation of subject's oxygen consumption (VO2) we used a novel neural network model which took into account the momentary HR level, HRV-derived respiration rate and on/off-response information as well as to age-based maximal heart rate and maximal oxygen consumption (VO2max) estimated from the collected background parameters [10].

In the physiological state classification algorithm, the degree of cardiac activity, assessed with HRV and HR parameters, was compared to concurrent physical metabolic requirements, assessed with estimated VO2. During physical activity, both physical metabolic requirements and cardiac activity were increased. The intensity of physical activity was used to separate physical activity into light physical activity and physical exercise. [9] In this paper, 20-30% of personal VO2max was considered as light physical activity and over 30% of VO2max as physical exercise.

Furthermore, stress state was determined as sympathetic dominance of the ANS with individually increased HR and decreased HRV without physical metabolic requirements. During recovery, parasympathetic activity dominates the ANS manifested by individually low HR level and high HRV. Thus, the analysis program took into account the individual basic resting HR and HRV values in the determination of the physiological states. [9]

C. Predictors of recovery during sleep

As the predictors for recovery we used a set of parameters describing the subjects' personal characteristics and daily factors. Personal characteristics were age, gender, BMI and self-reported physical activity class. Daily factors were the daily physical activities, the time of going to sleep relative to midnight, the sleep duration, weekday and month of the measurement day, and alcohol consumed.

To assess subjects' physical activity, the total number of minutes for light physical activity and physical exercise, as previously described, were calculated. Moreover, continuous physical activity periods likely to improve the subjects' aerobic fitness were considered using different criterion. The subjects were classified to those who had an aerobic fitness enhancing physical activity period during the day and those who did not. The aerobic fitness enhancing physical activity period was defined to be a physical activity period lasting continuously at least 10 minutes at 3 METs [11].

The amount of ethanol (grams) per kg was estimated by taking number of alcohol portions reported by the subject, multiplied by 12 and divided by the subject's weight. The amount of ethanol per kilogram is referred to as "alcohol" in this paper. The predictors are summarized in Table 1.

D. Random forest analysis of predictors

RF is a statistical learning technique which uses a large number of decision trees to predict the outcome from the input variables. RFs have been widely used for prediction and classification tasks due to their good performance and their ability to assess the importance of predictors. We chose to use RFs to identify the most important predictors because they require only minimal parametric assumptions about the relationships between the input and output variables and the distribution of the error [12].

The RF analysis started by generating bootstrapped samples of the data. Bootstrapped samples had the same size as the original data but they were generated from only about two thirds of the data. Thereafter, for each sample a tree was grown based on recursive binary splits on the sample. In the RF trees the binary splits were generated by finding the best binary split among a randomly chosen subset of variables. In prediction, the predicted value was obtained by averaging the predictions of all trees. [13]

The RF's goodness of fit can be estimated using the out-of-bag (OOB) samples i.e. the samples which were not used in generating the bootstrapped samples of the original data. The OOB prediction error for the RF is estimated by using the OOB samples as inputs and calculating the mean squared error (MSE) between the RF generated outputs and the actual outputs of the OOB samples. The OOB samples are also used for calculating the predictor importance of the variables. The predictor importance for a variable is estimated by permuting the variable and keeping other variables fixed and calculating the increase in MSE compared to the original (without permutation). [13]

In this study, the RF was implemented in R using package randomForest. The quantified daily lifestyle choices were used as input variables and the recovery minutes during sleep was used as an output variable. RF including 1,000 trees was found appropriate for this study. MSE was applied as splitting criterion. The minimum leaf size for the trees was set to be five and the number of predictors randomly chosen for each split was set to be one third of the total number of predictors as recommended [11].

TABLE I. PREDICTORS OF RECOVERY DURING SLEEP

Variable categories	**Variable description (Variable name)**
Background characteristics	Age, Gender, Body-mass index (BMI), Self-reported physical activity class (ActivityClass)
Physical Activity	Minutes of physical exercise (PAmins), Minutes of light physical activity (LightPAmins), Measurement day includes or does not include fitness enhancing physical activity (IsPAPeriod)
Time	Weekday, Month, Measurement is or is not a workday (IsWorkday)
Sleep	Duration of sleep (SleepDuration), Time of going to bed relative to midnight (BedTime)
Alcohol	Amount of ethanol in grams consumed during the measurement day per body mass (Alcohol)

RFs can be used for prediction even if the predictors are correlated, but some caution is needed when assessing the importance of predictors [14]. In order to cope with correlated predictors, we adapted a recursive feature elimination algorithm to repeatedly train new RF and remove less informative features. At first, all predictors were used to train a RF, the importance of predictors was evaluated and the predictor with the lowest importance was removed. Thereafter, a new RF was trained using the remaining predictors. Removing of predictors and training of RFs was continued until there was only one predictor left. For the compact model we used the set of predictors which yielded in the lowest OOB prediction error. This procedure was repeated 10 times in order to obtain a consensus.

III. RESULTS AND DISCUSSION

The importance of predictors for complete and compact models is shown in Figure 1. Only gender and day type (workday or day off) were removed as unnecessary features and the increase in MSE within other variable remains rather similar to the complete model. Alcohol was clearly the strongest predictor of recovery time followed by sleep duration and physical activity.

The direction how predictors affect cannot be seen from the variable importance plot (Fig. 1). Partial dependence plots depict the dependence between the response and a target feature by marginalizing over the values of all other features. The nonlinearities in the partial dependence plots revealed interesting details in the dependence between the response and target variables.

Fig. 2 shows the partial dependence plot of alcohol. The decrease in recovery minutes appears dose-dependent and according to our experiments, the effect seems to follow similar trend between the different age groups. The results indicate that one or two portions of alcohol depending on the subject's weight do not reduce the recovery minutes during next night sleep. This finding is in line with earlier studies reporting dose-dependent disturbing effects of alcohol on sleep structure [15] and on nocturnal ANS activity [16].

Fig. 3 shows the partial dependence plot of sleep duration. Apparently, as sleep duration increases also the number of recovery minutes increases. However, the curve is slightly steeper with short than long sleep duration, and interestingly the recovery minutes suddenly increase at around 7 to 7.25 hours. This means that the proportion of recovery from the total sleep time (i.e. "efficiency of recovery") peaks with sleep durations around 7 hours. Approximately 7 hours of sleep has previously been associated with lowest mortality rates [17] and short and long sleep durations with poorer self-reported health [18]. In general, 7-9 hours of sleep is recommended for adults' good sleep hygiene [19]. One highly possible explanation for our finding is that those sleeping about 7 hours are the ones who have the most regular sleeping schedules resulting in the most efficient recovery during sleep.

Fig. 4 shows the partial dependence plot for physical exercise in different physical activity classes (sedentary (0-3), moderately active (4-6), athletes (7-10)). The results show that exercise minutes per day are negatively associated with the recovery minutes during next night sleep. However, the difference between the activity classes is clear; the persons with higher physical fitness have better recovery during the night's sleep. Thus, it seems that being in good aerobic fitness increases recovery minutes, but an acute physical exercise challenges the recovery process and for days with exercise the recovery minutes are actually decreased during the next night sleep. Hence, positive effect of physical exercise is seen with a delay and occurs through the increased aerobic fitness.

Fig. 5 shows the partial dependence plot on light physical activity. A few minutes of light physical activity seems to increase the recovery minutes during the next night sleep but higher minutes of even light physical activity are associated with lower recovery time during next night sleep.

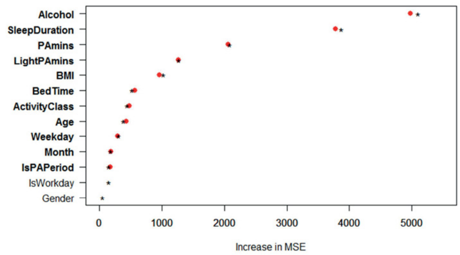


Figure 1. The importance of predictors for complete (black asterix) and compact model (red dots).

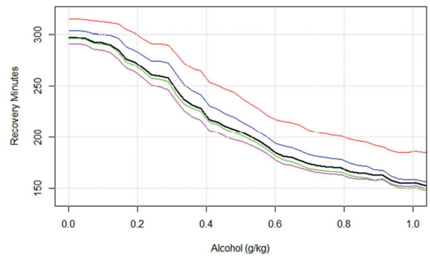


Figure 2. The partial dependence plot of alcohol for all (black line) and for age groups of 18-30 yrs (red), 31-40 yrs (blue), 41-50 yrs (green) and 51-65 yrs (magenta).

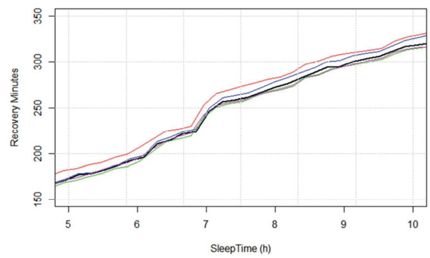


Figure 3. The partial dependence plot of sleep time for all (black line) and for age groups of 18-30 yrs (red), 31-40 yrs (blue), 41-50 yrs (green) and 51-65 yrs (magenta).

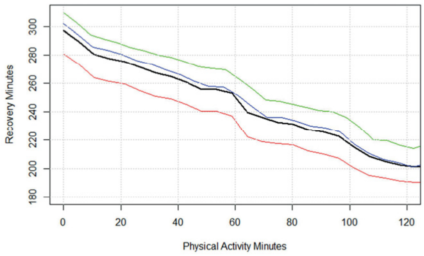


Figure 4. The partial dependence plot of physical exercise minutes for all (black line) and for sedentary (red), moderately active (blue) and athletic (green) subjects.

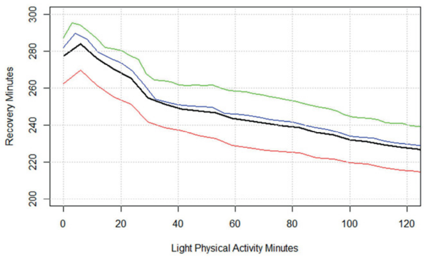


Figure 5. The partial dependence plot of light physical activity minutes for all (black line) and for sedentary (red), moderately active (blue) and athletic (green) subjects.

The results of the present study suggest that, even in the long run, being physically active is highly beneficial for health. In addition it increases the resources of the ANS but it is also important to take into consideration stress put upon the body by exercising and have days off from exercise.

The RF analysis showed interesting, non-linear effects of alcohol consumption, sleep time and physical activity on recovery minutes during sleep. As limitations of this study, the alcohol consumption was self-reported and timings of alcohol intake or physical activity were not considered in the analysis. The times of alcohol intake are not available in the data but the effect of timing of exercise could be examined separately by including only the persons who had exercised.

IV. Conclusion

In this paper we used RFs to examine which personal characteristics and daily lifestyle choices would be the most relevant for predicting recovery during next night's sleep calculated by assessing ANS balance from HRV. This was an exploratory study, with a limited list of predictors, but with a large and heterogeneous population of Finnish working-age adults containing unique self-reported data and physiological information obtained from HRV.

We found that alcohol was the strongest predictor for recovery and remarkably and dose-dependently decreased recovery minutes during the next night sleep. Alcohol was self-reported and the time of consumption was not known. Nevertheless, the study indicates that habitual drinking disturbs the nocturnal recovery process and is an extremely important lifestyle factor to take into consideration.

Better aerobic fitness was associated with increased number of recovery minutes, but acute physical activity during the day was, however, found to challenge the recovery process and even decrease recovery minutes during the next night sleep. Longer sleeping time enables more recovery minutes, but sleep duration around 7 hours was found to be the most efficient in terms of recovery, i.e. it contained the highest proportion of recovery minutes from the overall sleep duration.

Acknowledgment

This study was done under "Health Data Mining" and "the Finland Distinguished Professor Programme" projects, both financed by the Technology Development center of Finland (TEKES).

References

[1] H. Lindholm, "Physiological determinants and assessment of stress and recovery among media workers," 2013.
[2] T. Myllymäki *et al.*, "Effects of vigorous late-night exercise on sleep quality and cardiac autonomic activity," *J. Sleep Res.*, vol. 20, pp. 146-153, 2011.
[3] B. S. McEwen, "Physiology and neurobiology of stress and adaptation: central role of the brain," *Physiol. Rev.*, vol. 87, pp. 873-904, Jul, 2007.
[4] U. Acharya *et al.*, "Heart rate variability," in *Advances in Cardiac Signal Processing*, 1st ed., U. R. Acharya *et al.*, Eds. Springer Berlin Heidelberg, 2007, pp. 121-165.
[5] H. Bonnemeier *et al.*, "Circadian Profile of Cardiac Autonomic Nervous Modulation in Healthy Subjects:" *J. Cardiovasc. Electrophysiol.*, vol. 14, pp. 791-799, 2003.
[6] J. F. Brosschot *et al.*, "Daily worry is related to low heart rate variability during waking and the subsequent nocturnal sleep period," *International Journal of Psychophysiology*, vol. 63, pp. 39-47, 2007.
[7] K. Hottenrott *et al.*, "Heart rate variability and physical exercise. Current status," *Herz*, vol. 31, pp. 544-552, Sep, 2006.
[8] M. Chung *et al.*, "Sleep and autonomic nervous system changes—enhanced cardiac sympathetic modulations during sleep in permanent night shift nurses," *Scand. J. Work Environ. Health*, pp. 180-187, 2009.
[9] J. Kettunen and S. Saalasti, "Procedure for detection of stress by segmentation and analyzing a heart beat signal,", 17.11.2005, 2008.
[10] Firstbeat Technologies Ltd. VO2 estimation method based on heart rate measurement. 2005.
[11] C. E. Garber *et al.*, "American College of Sports Medicine position stand. 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.*, vol. 43, pp. 1334-1359, Jul, 2011.
[12] U. Grömping, "Variable importance assessment in regression: linear regression versus random forest," *The American Statistician,* vol. 63, 2009.
[13] R. Genuer *et al.*, "Variable selection using random forests," *Pattern Recog. Lett.*, vol. 31, pp. 2225-2236, 2010.
[14] B. Gregorutti *et al.*, "Correlation and variable importance in random forests," ArXiv Preprint arXiv:1310.5726, 2013.
[15] I. O. Ebrahim *et al.*, "Alcohol and sleep I: effects on normal sleep, " *Alcoholism: Clinical and Experimental Research,* vol. 37, pp. 539-549, 2013.
[16] Y. Sagawa *et al.*, "Alcohol Has a Dose-Related Effect on Parasympathetic Nerve Activity During Sleep," *Alcoholism: Clinical and Experimental Research,* vol. 35, pp. 2093-2100, 2011.
[17] D. F. Kripke *et al*, "Mortality associated with sleep duration and insomnia," *Arch. Gen. Psychiatry,* vol. 59, pp. 131-136, 2002.
[18] C. Frange *et al.*, "The impact of sleep duration on self-rated health," *Sleep Science,* vol. 7, pp. 107-113, 2014.
[19] M. Hirshkowitz *et al.* "National Sleep Foundation's sleep time duration recommendations: methodology and results summary," *Sleep Health,* vol. 1, pp. 40-43, 2015.

PUBLICATION
IV

Acute Effect of Alcohol Intake on Cardiovascular Autonomic Regulation During the First Hours of Sleep in a Large Real-World Sample of Finnish Employees: Observational Study

PIETILÄ, J., HELANDER, E., KORHONEN, I., MYLLYMÄKI, T., KUJALA, U.M. and LINDHOLM, H.

JMIR Mental Health, 5(1), pp. e23
DOI: 10.2196/mental.9519

Original Paper

Acute Effect of Alcohol Intake on Cardiovascular Autonomic Regulation During the First Hours of Sleep in a Large Real-World Sample of Finnish Employees: Observational Study

Julia Pietilä[1], MSc; Elina Helander[1], PhD; Ilkka Korhonen[1,2], PhD; Tero Myllymäki[2,3], MSc; Urho M Kujala[4], MD, PhD; Harri Lindholm[5,6], MD, PhD

[1]Faculty of Biomedical Sciences and Engineering, BioMediTech Institute, Tampere University of Technology, Tampere, Finland

[2]Firstbeat Technologies, Jyväskylä, Finland

[3]Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

[4]Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

[5]Finnish Institute of Occupational Health, Helsinki, Finland

[6]Nokia Technologies, Espoo, Finland

Corresponding Author:
Julia Pietilä, MSc
Faculty of Biomedical Sciences and Engineering
BioMediTech Institute
Tampere University of Technology
Korkeakoulunkatu 3
Tampere, 33720
Finland
Phone: 358 405475334
Fax: 358 331153015
Email: julia.pietila@tut.fi

Abstract

Background: Sleep is fundamental for good health, and poor sleep has been associated with negative health outcomes. Alcohol consumption is a universal health behavior associated with poor sleep. In controlled laboratory studies, alcohol intake has been shown to alter physiology and disturb sleep homeostasis and architecture. The association between acute alcohol intake and physiological changes has not yet been studied in noncontrolled real-world settings.

Objective: The aim of this study was to assess the effects of alcohol intake on the autonomic nervous system (ANS) during sleep in a large noncontrolled sample of Finnish employees.

Methods: From a larger cohort, this study included 4098 subjects (55.81%, 2287/4098 females; mean age 45.1 years) who had continuous beat-to-beat R-R interval recordings of good quality for at least 1 day with and for at least 1 day without alcohol intake. The participants underwent continuous beat-to-beat R-R interval recording during their normal everyday life and self-reported their alcohol intake as doses for each day. Heart rate (HR), HR variability (HRV), and HRV-derived indices of physiological state from the first 3 hours of sleep were used as outcomes. Within-subject analyses were conducted in a repeated measures manner by studying the differences in the outcomes between each participant's days with and without alcohol intake. For repeated measures two-way analysis of variance, the participants were divided into three groups: low (≤0.25 g/kg), moderate (>0.25-0.75 g/kg), and high (>0.75 g/kg) intake of pure alcohol. Moreover, linear models studied the differences in outcomes with respect to the amount of alcohol intake and the participant's background parameters (age; gender; body mass index, BMI; physical activity, PA; and baseline sleep HR).

Results: Alcohol intake was dose-dependently associated with increased sympathetic regulation, decreased parasympathetic regulation, and insufficient recovery. In addition to moderate and high alcohol doses, the intraindividual effects of alcohol intake on the ANS regulation were observed also with low alcohol intake (all *P*<.001). For example, HRV-derived physiological recovery state decreased on average by 9.3, 24.0, and 39.2 percentage units with low, moderate, and high alcohol intake, respectively. The effects of alcohol in suppressing recovery were similar for both genders and for physically active and sedentary subjects but stronger among young than older subjects and for participants with lower baseline sleep HR than with higher baseline sleep HR.

Conclusions: Alcohol intake disturbs cardiovascular relaxation during sleep in a dose-dependent manner in both genders. Regular PA or young age do not protect from these effects of alcohol. In health promotion, wearable HR monitoring and HRV-based analysis of recovery might be used to demonstrate the effects of alcohol on sleep on an individual level.

(JMIR Ment Health 2018;5(1):e23) doi:10.2196/mental.9519

KEYWORDS

heart rate; heart rate variability; sleep; alcohol drinking; autonomic nervous system; wearable electronic device

Introduction

Background

Sleep is a crucial period of physiological restoration, and it is the optimal state to assess the tonic component or the most relaxed state of the autonomic nervous system (ANS) in real-life conditions [1]. Poor sleep attenuates relaxation in the ANS [2], impairs regenerative physiological processes, causes metabolic disturbances, and has been associated with negative health outcomes [3]. Alcohol intake disturbs recovery, sleep homeostasis, and sleep architecture in several ways [4]. Alcohol affects negatively on stress-related cardiovascular adaptation in the ANS and hypothalamus-pituitary-adrenal axis [5]. Still, alcohol is used to relieve stress [6] or as sleep medicine [4]. Increased alcohol consumption is associated with long working hours, poor social support, and low job control [7].

Heart rate variability (HRV) is a widely used marker of cardiac autonomic regulation reflecting fluctuations in R-R intervals in short or extended time recordings [8] and is modulated by respiration, central vasoregulatory centers, peripheral baroreflex loops, and genetic factors [9]. HRV decreases with age, although differently in men and women [8]. In addition, suppressed HRV has been shown to predict occurrence of different diseases and conditions such as diabetic neuropathy or left ventricular dysfunction after acute myocardial infarction [10]. Traditionally, the HRV analysis is performed in time or frequency domains [10] but also novel analysis methods exist [11]. A widely used time domain measure of HRV is the root mean square of the successive differences (RMSSD) between adjacent R-R intervals, which mainly reflects the parasympathetic input of cardiac regulation [10]. In the frequency domain analysis, the high frequency (HF) band of HRV is considered to indicate parasympathetic regulation [10], whereas the low frequency (LF) band reflects both parasympathetic and sympathetic regulation [10,12]. The ratio between LF power and HF power (LF/HF ratio) has been suggested to reflect the balance between the two branches of the ANS, but this suggestion has not received a consensus [10,12]. Recently, a standardized reporting system in HRV-related behavioral studies was proposed [13].

In addition to the cardiac autonomic regulation, HRV analysis may also provide useful information on sleep [14], the sleeping brain [15], and stress-related insufficient recovery [16], even though autonomic regulation during sleep is complex and varies during different sleep stages [14]. During slow wave sleep (so-called deep sleep), the parasympathetic regulation has been reported to be dominating and the sympathetic regulation to be attenuated, whereas the opposite is true for rapid eye movement sleep [14]. The sufficient amount of slow wave sleep has been associated with good physical and mental recovery [1]. Unconscious stress may be detected in physiological recordings made during cardiovascular stress recovery [17], and HRV may usefully reflect the adaptive resources of the ANS [18]. However, the limitations and pitfalls of HRV analysis as well as the physiological nature of HRV have to be taken into account in all interpretations.

Prior Work

The effect of acute alcohol intake on the ANS using heart rate (HR) and HRV parameters has been shown in the previous studies. In laboratory settings, high acute alcohol consumption (0.7 g/kg-1.0 g/kg) was associated with decreased HRV and increased HR in awake subjects [19]. The effect was also observed with lower doses (two drinks, not reported in g/kg units) [20]. In one laboratory study, both HRV and polysomnography were monitored after alcohol consumption [21]. The young healthy male subjects (n=10) were given no (0 g/kg, control), low (0.5 g/kg of ethanol), or high (1.0 g/kg) dose of alcohol. A dose-related effect of alcohol on HR and HRV during sleep was found, and the highest HR and lowest HRV were observed for high dose.

However, the effect of acute alcohol intake on the ANS during sleep has not been studied in noncontrolled free-living conditions or with large samples. Most published studies considering the effects of acute alcohol intake on HRV have involved only males, been rather small in number of participants, and included no comparison between genders or objective measurements of physical activity (PA) and recordings during sleep [9]. Thus, studies employing larger number of participants with both genders and considering the background parameters of the subjects such as age, body mass index (BMI) and PA, are needed.

Goal of This Study

The widespread use of wearable and connected consumer devices enables unobtrusive collection of massive amounts of data from large number of individuals during their daily life. These health-related datasets gathered under normal day-to-day circumstances outside of traditional clinical trials represent so called real-world data [22]. This real-world data collected in uncontrolled settings and outside of clinical trials may be exploited in research to complement the knowledge gained from the traditional clinical trials [23]. The multitude and variety of individuals and information included in real-world datasets allow studying aspects that cannot be studied to that extent in traditional clinical trials [23]. The real-world data has also the prospect to assess the generalizability of the findings from traditional clinical trials with specific populations and circumstances to broader populations and circumstances [22]. On the other hand, the associations found in real-world data can serve as hypotheses for further clinical trials [22]. To gain valid results from the real-world data, the data characteristics such as the sample bias, missing data, confounding, uncertainties and provenance of the data, must, however, be taken into account in the analysis [24].

Alcohol consumption is a universal health behavior associated with poor sleep [4], but to the authors' knowledge, there is not yet any study employing real-world data [9]. This study analyzes the effects of alcohol intake on the ANS during sleep in a large free-living population. An observational real-world dataset of continuous beat-to-beat R-R interval recordings and self-reported sleep times and alcohol consumption collected from over 40,000 subjects during their normal everyday life was employed for studying retrospectively the effect of acute alcohol intake on sleep. The intraindividual differences in the HRV during sleep associated with acute alcohol intake were studied from 4098 participants of various ages, BMI ranges, and PA levels in whom data with and without alcohol consumption during previous day was available. The purpose of the study was to assess the generalizability of the previous findings to broader population and to study associations between the characteristics of the subjects and the effects of acute alcohol intake on the HR and HRV parameters during sleep.

Methods

Data Collection

The original data sample contained 111,025 measurement days from 42,086 Finnish employees representing a wide range of blue- and white-collar workers in varying size companies. Employees had voluntarily participated in a preventive occupational health care program with the aim of improving their health habits and stress management. The program included a continuous beat-to-beat R-R interval recording for a few days during the participant's normal life. The R-R interval recordings were performed using Bodyguard (Firstbeat Technologies Ltd, Jyväskylä, Finland) wearable device that was attached on the chest with two electrodes. HRV indices, stress, recovery, and PA were computed with Firstbeat Analysis Server (Firstbeat Technologies Ltd) from the recorded R-R interval data, and together with other physiological measurements, they were used as health promotion tools at employees' workplaces. Employees were instructed not to participate in recordings if they had any disease stages or medications possibly affecting R-R intervals, for example, chronic heart rhythm disturbance, very high blood pressure (≥180/100 mm Hg), type 1 or 2 diabetes with autonomic neuropathy, severe neurological disease (eg, advanced multiple sclerosis or Parkinson disease), fever or other acute disease, or BMI >40 kg/m^2[25].

All the R-R interval recordings performed on the employees were analyzed and stored anonymously to a registry administered by Firstbeat Technologies Ltd. Each service provider conducting recordings for participants signed an agreement allowing Firstbeat Technologies Ltd to store the anonymized data and to use it for development and research purposes. The employers were responsible to inform their employees about the data usage. Following the agreements, a dataset was extracted from the registry for this study. The use of the dataset for research purposes was approved by the ethics committee of Tampere University Hospital (Reference No R13160).

Data Extraction

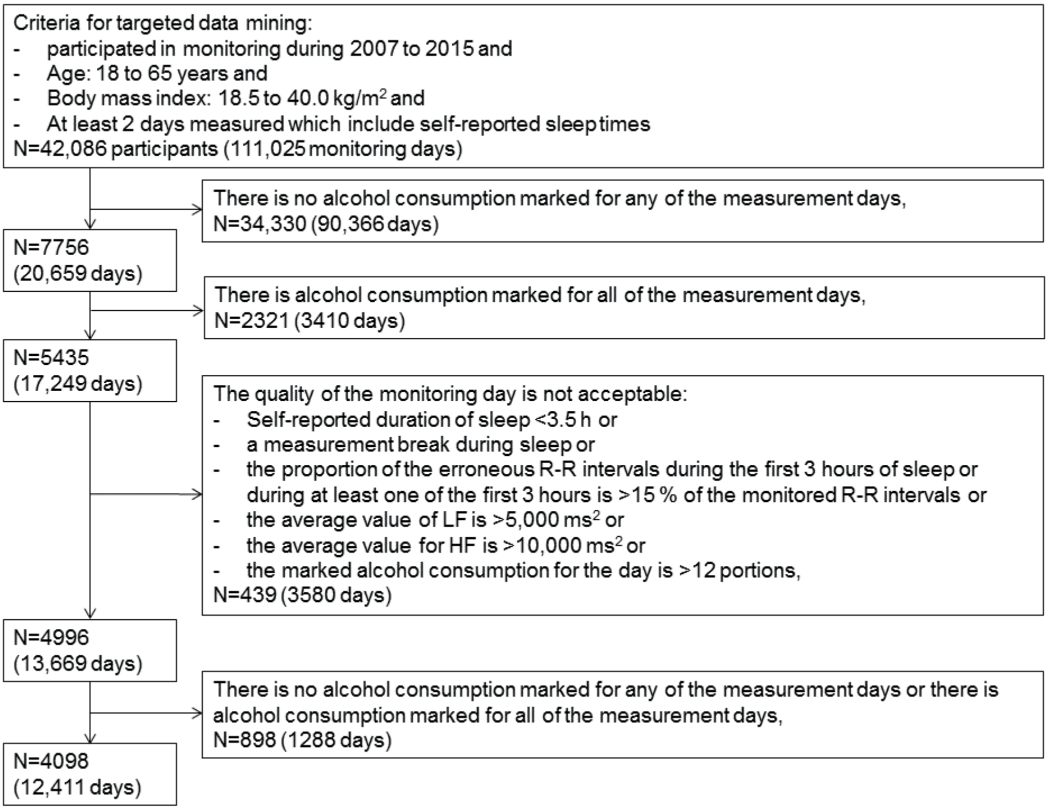
The dataset extracted from the registry to this study included the R-R interval recordings performed with the Bodyguard device (Firstbeat Ltd, Jyväskylä, Finland). The sampling frequency of the device is 1000 Hz for the R-R interval recording [26], and its mean absolute error for R-R intervals has been reported to be 4.45 ms [27]. An artifact correction was performed for the R-R intervals with Firstbeat Analysis Server [28], after which the mean absolute error of R-R intervals has been reported to be 2.27 ms [27]. For this study, the artifact-corrected beat-to-beat R-R intervals were analyzed for a 3-hour period starting 30 min after the self-reported onset of bedtime that is the most likely period for slow wave sleep.

From the artifact-corrected beat-to-beat R-R intervals, the average of HR in 10-min nonoverlapping windows and RMSSD with 5-min windows were calculated [10]. The frequency bands of HRV were assessed applying short-time Fourier transform on the artifact-corrected beat-to-beat R-R interval data. In addition to the traditional HRV measures, personalized HRV-derived indices of recovery were calculated with Firstbeat Analysis Server. The software detects the periods of recovery and thereafter estimates the magnitude of these recovery reactions based on a person's range of physiological reactions (eg, minimal and maximal HR) and time series variables related to parasympathetic and sympathetic modulation (eg, HR, HF power, LF power, and HRV-derived respiration rate) [29]. During recovery reactions, parasympathetic regulation predominates in the ANS [29]. The momentary absolute level of recovery reactions is estimated with parameters describing the magnitude of parasympathetic modulation, and it is high when HR is individually low and parasympathetic HRV is individually high [30].

For this study, the exclusion criteria were unknown or very high reported alcohol consumption (>12 portions of alcohol) during the recording day, unknown or very short self-reported sleeping time, and poor quality of HRV recordings (Figure 1). If a subject reported more than one sleep periods per day, only the longest sleep period was analyzed. Only subjects having a day with at least one portion of alcohol and a day with no alcohol intake were analyzed. The final analysis included 12,411 HRV recording days from 4098 individuals.

As background information, age, gender, and self-reported weight, height, and PA class modified from Ross and Jackson [31] were available. Participants were asked to note their alcohol intake as portions (1 portion=12 g of ethanol) for each measurement day preceding sleep. The exact timing of alcohol intake and smoking history were not available.

Figure 1. The selection of study population for the analyses.



Statistical Analyses

HR, RMSSD, LF/HF ratio, time considered as recovery (recovery percentage), and average of the momentary absolute levels of recovery reactions (recovery index) from a 3-hour period starting 30 min after the self-reported bedtime onset were considered as the outcome variables. Only the first 3 hours of sleep were analyzed, as most of the slow wave sleep typically occurs during the first hours of sleep [1]. During slow wave sleep, the parasympathetic regulation is dominating, and sufficient amount of slow wave sleep has been associated with both good physical and mental recovery [1].

All analyses were conducted in a within-subject repeated-measures manner by comparing the participants' outcome variables between days with and without alcohol intake. The within-subject design was used, as it allows studying the intraindividual effects of acute alcohol intake and controls for possible unknown confounders.

For the within-subject repeated-measures two-way analysis of variance (ANOVA), the participants' hourly averages of outcome variables were calculated for days with and without alcohol intake, and the participants were categorized into low (≤0.25 g/kg), moderate (>0.25-0.75 g/kg), and high (>0.75 g/kg) dose groups according to their alcohol intake during the day. Note that the groups also include the participant's reference with no alcohol, and the participants may have data in one, two, or all three dose groups. If a participant had more than 1 day with low, moderate, or high or with no alcohol intake, the outcome variables were averaged over those days. A repeated-measures two-way ANOVA was performed separately for each dose group to evaluate the difference and the shape of the hour-by-hour pattern in the outcome variables between the days with and without alcohol intake.

In the second analysis, the linear regression model was fitted for the difference in the average of the 3-hour HR and HRV parameters between the participant's days with and without alcohol intake. First, the 3-hour averages of the outcome variables were calculated for each measurement day. Thereafter, the difference in the participant's averages between the days with and without alcohol intake was calculated. If the participant had more than one measurement day without alcohol intake, the average of the measurement days' 3-hour outcome variable averages was employed. A dataset including the measurement day with the highest amount of reported alcohol intake from each participant was extracted, and a linear regression was fitted to the data using the difference in the outcome variables between the days with and without alcohol intake as a dependent variable. In addition to alcohol intake, all information available about the subjects was employed as independent variables in the regression models. The independent variables were continuous variables of alcohol dose (g/kg), age, PA class, BMI, and the 3-hour average of HR (bpm) during a night after a day without alcohol intake (baseline sleep HR) and gender as a categorical variable.

Age, BMI, and the baseline sleep HR were subtracted to baseline levels of 18 years, 18.5 kg/m^2, and 38 bpm, respectively. In addition, a linear regression with interactions between alcohol doses and other predictors was fitted.

All statistical analyses were conducted using R (The R Foundation for Statistical Computing) version 3.2.2. The level of significance in all analyses was set at <0.05. However, with data of this size, it is more important to focus on effect sizes than *P* values [32].

Results

Characteristics of the Study Population

From a larger cohort, this study included 4098 subjects who had continuous beat-to-beat R-R interval recordings of good quality with for at least 1 day with and for at least 1 day without alcohol intake. There was a significant proportion of female subjects in this study (Table 1). On average, the subjects were middle-aged, slightly overweight, and had regular PA 2 to 3 times per week, and the total weekly training amount being approximately 1 hour.

Neither PA class nor BMI differed significantly between the dose groups (*P*>.05, Table 2). High alcohol intake was more common among males (*P*<.001) and young subjects (*P*<.001). Average daily alcohol intake in the low, moderate, and high groups was 0.17, 0.45, and 1.1 g/kg, respectively, with the corresponding average number of reported alcohol portions being 1.1, 2.9, and 7.0 drinks.

Repeated-Measures Analysis of Variance Analyses

The means and 99% CIs for HR, the LF/HF ratio, RMSSD, the recovery percentage, and recovery index calculated from intraindividual HRV recordings during the first 3 hours of sleep in low, moderate, and high dose groups were calculated (Figures 2 and 3). Low HR and LF/HF ratio reflect increased parasympathetic regulation, and low RMSSD indicates increased sympathetic regulation in the ANS.

High alcohol intake had the greatest effect on the outcome variables. On average, HR was increased by 1.4 bpm with low, 4.0 bpm with moderate, and 8.7 bpm with high alcohol intake. The LF/HF ratio was increased by 0.1 with low, 0.3 with moderate, and 0.5 with high alcohol intake. RMSSD was decreased by 2.0 ms with low, 5.7 ms with moderate, and 12.9 ms with high alcohol intake. The recovery percentage was decreased by 9.3 percentage units with low, 24.0 percentage units with moderate, and 39.2 percentage units with high alcohol intake. The recovery index was decreased by 7.1 with low, 20.8 with moderate, and 40.2 with high alcohol intake.

For each dose group, the within-subject repeated-measures two-way ANOVA showed significant differences in all outcome variables (all *P*<.001) between the days with and without alcohol intake. In addition, the hourly HRV parameters differed significantly from each other (all *P*<.001). In high dose group comparisons, the interactions between the hour of sleep and alcohol intake were statistically significant for all outcome parameters (all *P*<.001), indicating that the hour-by-hour pattern in the HRV parameters during sleep was different for subjects between the days with high and no alcohol intake. For days with high alcohol intake, the average LF/HF ratio increased hour-by-hour during sleep, whereas the average LF/HF ratio increased from the first to the second hour of sleep but decreased from the second to the third hour of sleep for days with no alcohol intake. For days without alcohol intake, the recovery percentage and recovery index increased as the sleep progressed, but this did not occur after high alcohol intake. In moderate dose group comparisons, the interactions between hour and alcohol intake were statistically significant for the LF/HF ratio (*P*=.002) and recovery percentage (*P*=.01).

Table 1. Characteristics of the study population.

Demographic characteristic	All (N=4098), mean (SD; range)	Males (N=1811), mean (SD; range)	Females (N=2287), mean (SD; range)
Age (years)	45.1 (9.6; 19-65)	45.2 (9.4; 19-65)	44.9 (9.8; 19-65)
Physical activity class[a]	4.8 (1.8; 0-10)	4.9 (1.7; 0-10)	4.8 (1.8; 0-10)
Body mass index (kg/m^2)	26.0 (4.0; 18.5-39.9)	26.7 (3.5; 18.9-39.5)	25.4 (4.3; 18.5-39.9)

[a]Physical activity class range: 0 (physically inactive) to 10 (physically very active).

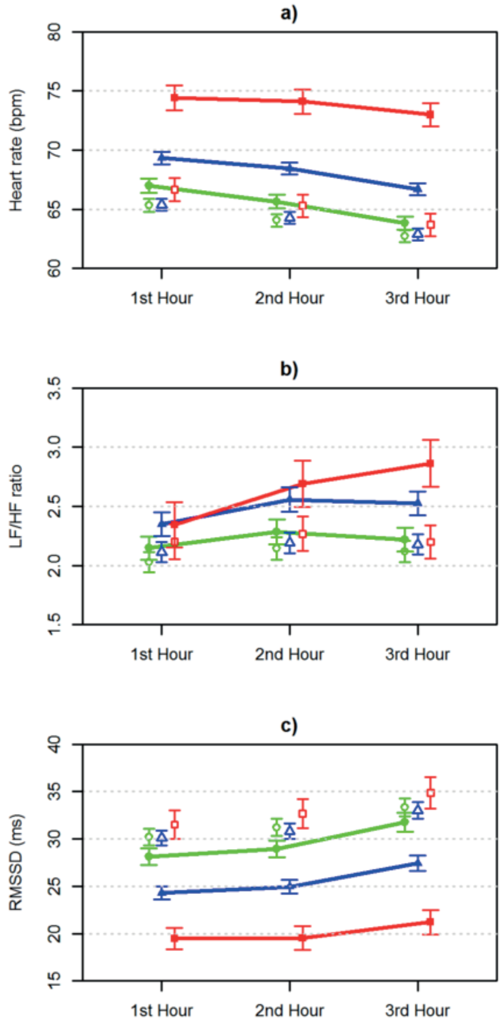
Table 2. Characteristics of low, moderate, and high dose groups during the heart rate variability (HRV) recordings.

Demographic characteristic	Low ≤0.25 g/kg (n=1752)	Moderate >0.25-0.75 g/kg (n=2194)	High >0.75 g/kg (n=716)	*P* value
Number of male subjects, n (%)	671 (38.29)	1010 (46.03)	380 (53.1)	<.001[a]
Age in years, mean (SD)	45.6 (9.0)	46.3 (9.3)	42.3 (10.7)	<.001[b]
Physical activity class, mean (SD)	4.9 (1.6)	4.9 (1.7)	4.6 (1.9)	.59[b]
Body-mass index in kg/m^2, mean (SD)	25.9 (4.2)	26.0 (3.8)	25.8 (3.7)	.10[b]
Weight in kg, mean (SD)	77.5 (16.2)	77.9 (14.3)	78.1 (14.1)	.31[b]

[a]Chi-square test.

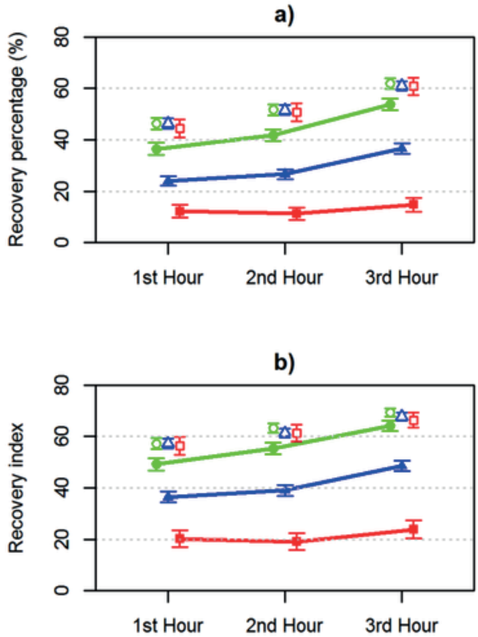
[b]One-way analysis of variance (ANOVA) adjusted for gender.

Figure 2. The effect of alcohol intake during the three first hour of sleep on a) heart rate, b) low frequency/high frequency (LF/HF) ratio, and c) root mean square of the successive differences (RMSSD). The marks green ●=low dose (≤0.25 g/kg), blue ▲=medium dose (>0.25-0.75 g/kg), and red ■=high dose (>0.75 g/kg) denote the averages, and corresponding white symbols denote the measures for the same persons without alcohol. Due to the size of the data and clarity of the figure, 99% CIs are shown, and the lines between hours are only shown for alcohol dose groups.



This shows that the hour-by-hour pattern was different between the days with moderate and no alcohol intake only for the LF/HF ratio and recovery percentage. In low dose comparisons, the hour-by-hour pattern in the LF/HF ratio (*P*=.51), RMSSD (*P*=.06), and recovery percentage (*P*=.08) during sleep was similar between the days with low and no alcohol intake. The HR (*P*<.001) and recovery index (*P*=.01) variables had a statistically significant interaction between the hour and alcohol intake, ie, their hour-by-hour pattern during sleep differed between the days with low and no alcohol intake. Visual inspection (Figures 2 and 3) showed that after low alcohol intake, the levels of outcome variables during the third hour approach their reference levels.

Figure 3. The effect of alcohol intake during the three first hour of sleep on a) recovery percentage, and b) recovery index. The marks green ●=low dose (≤0.25 g/kg), blue ▲=medium dose (>0.25-0.75 g/kg), and red ■=high dose (>0.75 g/kg) denote the averages, and corresponding white symbols denote the measures for the same persons without alcohol. Due to the size of the data and clarity of the figure, 99% CIs are shown, and the lines between hours are only shown for alcohol dose groups.



Linear Models

In linear model analysis, alcohol intake significantly affected the outcome variables (Tables 3 and 4). The results show that HR was increased with acute alcohol intake. For example, alcohol intake of 0.75 g/kg increased HR for subjects on average by 6.8 bpm compared with their nights without alcohol intake (Table 3). Alcohol intake increased the HR significantly more among young than older subjects: alcohol intake of 0.75 g/kg increased HR on average by 1.8 bpm more for a 30-year old person than for a 60-year old person (Table 4). In addition, alcohol intake increased HR significantly more among subjects with lower than higher baseline sleep HR: alcohol intake of 0.75 g/kg and an increase of 10 bpm in baseline HR decreased the difference in HR by 3.4 bpm, on average (Table 4). The increase in HR with alcohol intake was similar for subjects despite their gender, PA level, or BMI (Table 4).

Alcohol intake increased the LF/HF ratio, and this effect was slightly stronger among males and subjects with higher PA level (Table 4). However, the coefficients of determination for the LF/HF ration linear regression models were low (Tables 3 and 4), indicating that the input variables employed in the models did not explain the variation in the LF/HF ratio well. RMSSD was decreased by alcohol intake at all ages, but the effect was stronger in younger than older subjects (Table 4). On average, RMSSD was decreased with high alcohol intake (0.75 g/kg) by 10.9 ms for a 30-year old subject but only by 4.7 ms for a 60-year old subject (Table 4). In addition, alcohol intake decreased RMSSD more for subjects with lower baseline HR (Table 4).

Recovery percentage decreased significantly by increased alcohol intake (Tables 3 and 4). An 80-kg person drinking five portions of alcohol (0.75 g/kg) has on average 45 min less recovery (25.25 percentage units) during the first 3 hours of sleep than without alcohol (Table 3). The recovery percentage was decreased significantly more by alcohol intake for subjects with lower baseline HR than with higher baseline HR (Table 4). The decrease in recovery percentage with alcohol was similar regarding the other background parameters of the subjects (Table 4). In addition, the recovery index was attenuated with alcohol intake (Tables 3 and 4). Alcohol intake attenuated the recovery index slightly more in subjects with higher BMI than with lower BMI (Table 4). The other background parameters did not have a significant interaction with alcohol intake (Table 4).

When the effects for alcohol and background characteristics were controlled, the difference in recovery percentage was strongly correlated with the difference in HR (Pearson partial correlation coefficient and the coefficient of determination: $r=-.70$, $R^2=.486$, $P<.001$) and in RMSSD ($r=.51$, $R^2=.262$, $P<.001$) but only moderately correlated with the change in LF/HF ratio ($r=-.27$, $R^2=.071$, $P<.001$). Similarly, the difference in the recovery index was strongly correlated with the difference in HR ($r=-.63$, $R^2=.388$, $P<.001$) and in RMSSD ($r=.49$, $R^2=.236$, $P<.001$) but only moderately correlated with the change in LF/HF ratio ($r=-.27$, $R^2=.074$, $P<.001$). The partial correlation between the difference in the recovery percentage and the recovery index was moderate ($r=.48$, $R^2=.229$, $P<.001$).

Table 3. The linear regression models without interaction components for the average of heart rate (HR), low frequency/high frequency (LF/HF) ratio, root mean square of the successive differences (RMSSD), recovery percentage, and recovery index during the first 3 hours of sleep. BMI: body mass index.

Outcome	HR	LF/HF ratio	RMSSD	Recovery percentage	Recovery index
Intercept, estimate (SE)	10.87 (0.66)[a]	0.715 (0.115)[a]	−15.32 (0.98)[a]	−56.09 (3.25)[a]	−28.27 (3.70)[a]
Alcohol (g/kg), estimate (SE)	8.49 (0.29)[a]	0.425 (0.051)[a]	−12.24 (0.44)[a]	−33.67 (1.45)[a]	−36.63 (1.65)[a]
Physical activity class, estimate (SE)	−0.48 (0.06)[a]	−0.019 (0.011)	0.37 (0.09)[a]	1.62 (0.31)[a]	1.81 (0.35)[a]
Age (0=18 years), estimate (SE)	−0.03 (0.01)[b]	−0.001 (0.002)[a]	0.14 (0.02)[a]	−0.06 (0.05)	−0.08 (0.06)
BMI (0=18.5 kg/m^2), estimate (SE)	0.22 (0.03)[a]	0.002 (0.005)	−0.26 (0.04)[a]	−0.81 (0.14)[a]	−0.89 (0.15)[a]
Gender (0=female, 1=male), estimate (SE)	−1.70 (0.21)[a]	0.014 (0.037)	2.09 (0.32)[a]	7.22 (1.06)[a]	5.24 (1.21)[a]
Baseline sleep HR (0=38 bpm), estimate (SE)	−0.33 (0.01)[a]	−0.021 (0.002)[a]	0.41 (0.02)[a]	1.67 (0.06)[a]	0.86 (0.07)[a]
Adjusted coefficient of determination for the model	0.267	0.039	0.245	0.230	0.135

[a] $P<.001$.
[b] $P<.01$.

Table 4. The linear regression models with interaction components for the average of heart rate (HR), low frequency/high frequency (LF/HF) ratio, root mean square of the successive differences (RMSSD), recovery percentage, and recovery index during the first 3 hours of sleep.

Outcome	HR	LF/HF ratio	RMSSD	Recovery percentage	Recovery index
Intercept, estimate (SE)	7.61 (1.04)[a]	0.724 (0.182)[a]	−10.55 (1.55)[a]	−49.34 (5.13)[a]	−26.73 (5.86)[a]
Alcohol (g/kg), estimate (SE)	14.47 (1.56)[a]	0.386 (0.272)	−20.68 (2.32)[a]	−46.05 (7.69)[a]	−38.48 (8.78)[a]
Physical activity class, estimate (SE)	−0.33 (0.10)[b]	−0.051 (0.018)[b]	0.51 (0.15)[a]	2.29 (0.51)[a]	2.57 (0.59)[a]
Age (0=18 years), estimate (SE)	0.03 (0.02)	0.004 (0.003)	−0.008 (0.03)	0.07 (0.09)	−0.05 (0.10)
BMI (0=18.5 kg/m^2), estimate (SE)	0.19 (0.04)[a]	0.005 (0.008)	−0.18 (0.07)[b]	−0.60 (0.22)[b]	−0.42 (0.25)
Gender (0=female, 1=male), estimate (SE)	−1.38 (0.35)[a]	−0.010 (0.062)	2.04 (0.53)[a]	6.47 (1.75)[a]	2.95 (2.00)
Baseline sleep HR (0=38 bpm), estimate (SE)	−0.28 (0.02)[a]	−0.019 (0.004)[a]	0.31 (0.03)[a]	1.25 (0.10)[a]	0.71 (0.11)[a]
Alcohol x physical activity class, estimate (SE)	−0.27 (0.17)	0.065 (0.029)[c]	−0.32 (0.25)	−1.47 (0.83)	−1.59 (0.95)
Alcohol x age	−0.12 (0.03)[a]	−0.009 (0.005)	0.26 (0.04)[a]	−0.04 (0.14)	0.27 (0.16)
Alcohol x BMI, estimate (SE)	0.08 (0.08)	−0.009 (0.014)	−0.15 (0.12)	−0.39 (0.39)	−1.05 (0.45)[c]
Alcohol x gender, estimate (SE)	−0.56 (0.60)	0.251 (0.104)[c]	−0.13 (0.89)	1.35 (2.95)	4.87 (3.37)
Alcohol x baseline sleep HR, estimate (SE)	−0.08 (0.03)[c]	−0.005 (0.005)	0.18 (0.05)[a]	0.83 (0.15)[a]	0.30 (0.17)
Adjusted coefficient of determination for the model	0.271	0.042	0.255	0.236	0.137

[a] $P<.001$.
[b] $P<.01$.
[c] $P<.05$.

Discussion

Principal Findings

Impact of alcohol on autonomic nervous system control during sleep has been earlier demonstrated in controlled conditions with relatively small samples. This study demonstrated that this effect is also clearly seen in noncontrolled conditions with wearable HR monitoring and HRV analysis. In the large heterogeneous, noncontrolled, and free-living study population, alcohol intake caused a dose-dependent effect in cardiac autonomic regulation during the first 3 hours of self-reported sleep time. Intraindividually, HR remained elevated, parasympathetic recovery was delayed, and sympathetic dominance was prolonged after alcohol intake compared with recordings with no alcohol. The effects in cardiac autonomic regulation were observed already with low doses of alcohol.

These findings accord with previous studies reporting dose-related effects of alcohol on parasympathetic indices of HRV during sleep in laboratory conditions [21]. Increased HR partly explains the attenuated HRV indices during sleep following alcohol intake [33], and prolonged elevation in the LF/HF ratio supports the role of sympathetic regulation in alcohol-related delayed HRV recovery [34]. Even a moderate amount of alcohol was shown to attenuate recovery in the ANS in this study. This accords with the results of a previous study where two drinks caused significantly decreased RMSSD and increased HR and LF/HF ratio, and one drink altered RMSSD but not HR or LF/HF ratio [20]. In this study, HR and the LF/HF ratio were affected also in the low dose (≤0.25 g/kg) group where about 90% of the measurement days involved one drink and the rest two drinks.

The strength of this study was the large study population representing a sample of Finnish employees with the whole span of working age, different BMI categories and PA levels, and both genders. With the large free-living sample, this study provided real-world evidence and enabled further studying the effects of personal background parameters on the effects of alcohol intake on the ANS. The main limitations of the study were not knowing the exact alcohol doses and the exact times of alcohol consumption and sleep. The higher alcohol intakes may have been underestimated. In addition, the alcohol drinking habits of the participants were not known.

Most previous studies considering the effects of alcohol on the ANS have used male subjects only, and differences between the sexes have not been examined [9]. With a significant proportion of female participants, this study showed alcohol mainly affecting the ANS similar among men and women, although the LF/HF ratio showed sympathetic dominance being slightly stronger in men than in women after alcohol intake. The large age range of the participants allowed studying the interaction effect of age and alcohol intake on the ANS. The effect of alcohol intake on the change in HR and RMSSD was stronger in young subjects than in older subjects, but the effect of alcohol on the LF/HF ratio, the recovery percentage, or the recovery index was not age-dependent.

Our findings on the modifiable disease risk factors are in agreement with previous data on that physical inactivity and high BMI reduce HRV [35] and show that consumption of alcohol reduces HRV in all the PA and BMI categories. In fact, this study showed that regular PA does not attenuate the effects of alcohol intake on the ANS. The changes in HR and RMSSD because of alcohol intake were similar for physically active and sedentary participants in this study. The physically active participants actually displayed in LF/HF ratio even stronger intraindividual sympathetic dominance because of alcohol intake than the sedentary subjects did. Consequently, being physically active does not seem to protect from the negative effects of alcohol intake on the ANS during sleep. This aspect is important to consider given that the alcohol consumption is common also among physically active individuals, and there may even be a dose-response relationship between alcohol consumption and level of PA [36]. Even though exercise is beneficial for general health among alcohol users [37], alcohol has been reported to negatively affect HRV recovery after exercise [38]. Thus, clinically important is to note that the risk of exercise-related cardiac events might be raised by prolonged sympathetic tone during recovery. PA on the current day of alcohol intake, a factor that might affect HRV parameters [9], was not estimated, although it would be possible with our material. However, the effect of alcohol on the amount of recovery during sleep has been reported to strongly overwhelm the effect of other daily activities, including PA [39].

Poor sleep associates with negative health behaviors, ill health, and decreased work ability [40,41]. This study might offer some new tools for health promotion in occupational and primary health care to show practically, on individual and personal level, based on wearable HRV monitoring, the negative effects of alcohol on sleep. Demonstration of the insufficient recovery after using alcohol may be very important for individuals who consume alcohol repeatedly day after day and may suffer from accumulated consequences of insufficient recovery. The personalized indices of recovery, recovery percentage, and recovery index were found to accord with the RMSSD and HR. Importantly, the recovery percentage was found to be independent of age, and the recovery index had only a slight interaction effect between alcohol intake and BMI. These personalized recovery parameters can be used as a tool of health promotion in occupational health care to better manage interindividual differences in HRV and to visualize the associations between alcohol consumption and sleep.

Conclusions

The study demonstrates, with big uncontrolled data from unobtrusive wearable monitoring, that alcohol intake results in suppression of parasympathetic regulation of the ANS in a dose-response manner. Being physically active and young appears to provide no protection from alcohol-induced suppression of parasympathetic regulation, a finding that needs to be considered given the literature evidence that increased PA associates with higher alcohol usage among nonalcoholics. Personalized HRV measures such as recovery percentage may be more practical in occupational health settings to demonstrate the effect of alcohol on sleep than, eg, RMSSD, which is strongly age-dependent.

Acknowledgments

This study was supported by the Finnish Funding Agency for Technology and Innovation (TEKES).

Conflicts of Interest

HL is currently employed in the Digital Health Lab of Nokia Technologies, and TM and I are currently employed by Firstbeat Technologies Ltd, Jyväskylä, Finland.

References

1. von Bonin D, Grote V, Buri C, Cysarz D, Heusser P, Moser M, et al. Adaption of cardio-respiratory balance during day-rest compared to deep sleep--an indicator for quality of life? Psychiatry Res 2014 Nov 30;219(3):638-644. [doi: 10.1016/j.psychres.2014.06.004] [Medline: 25011731]
2. Burton AR, Rahman K, Kadota Y, Lloyd A, Vollmer-Conna U. Reduced heart rate variability predicts poor sleep quality in a case-control study of chronic fatigue syndrome. Exp Brain Res 2010 Jul;204(1):71-78. [doi: 10.1007/s00221-010-2296-1] [Medline: 20502886]
3. Porkka-Heiskanen T, Zitting K, Wigren H. Sleep, its regulation and possible mechanisms of sleep disturbances. Acta Physiol (Oxf) 2013 Aug;208(4):311-328. [doi: 10.1111/apha.12134] [Medline: 23746394]
4. Thakkar MM, Sharma R, Sahota P. Alcohol disrupts sleep homeostasis. Alcohol 2015 Jun;49(4):299-310 [FREE Full text] [doi: 10.1016/j.alcohol.2014.07.019] [Medline: 25499829]
5. Thayer JF, Hall M, Sollers JJ, Fischer JE. Alcohol use, urinary cortisol, and heart rate variability in apparently healthy men: Evidence for impaired inhibitory control of the HPA axis in heavy drinkers. Int J Psychophysiol 2006 Mar;59(3):244-250. [doi: 10.1016/j.ijpsycho.2005.10.013] [Medline: 16325293]
6. Rod NH, Grønbaek M, Schnohr P, Prescott E, Kristensen TS. Perceived stress as a risk factor for changes in health behaviour and cardiac risk profile: a longitudinal study. J Intern Med 2009 Nov;266(5):467-475 [FREE Full text] [doi: 10.1111/j.1365-2796.2009.02124.x] [Medline: 19570055]
7. Virtanen M, Jokela M, Nyberg ST, Madsen IE, Lallukka T, Ahola K, et al. Long working hours and alcohol use: systematic review and meta-analysis of published studies and unpublished individual participant data. Br Med J 2015 Jan 13;350:g7772 [FREE Full text] [Medline: 25587065]
8. Bonnemeier H, Richardt G, Potratz J, Wiegand UK, Brandes A, Kluge N, et al. Circadian profile of cardiac autonomic nervous modulation in healthy subjects: differing effects of aging and gender on heart rate variability. J Cardiovasc Electrophysiol 2003 Aug;14(8):791-799. [Medline: 12890036]
9. Romanowicz M, Schmidt JE, Bostwick JM, Mrazek DA, Karpyak VM. Changes in heart rate variability associated with acute alcohol consumption: current knowledge and implications for practice and research. Alcohol Clin Exp Res 2011 Jun;35(6):1092-1105. [doi: 10.1111/j.1530-0277.2011.01442.x] [Medline: 21332532]
10. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation 1996 Mar 01;93(5):1043-1065 [FREE Full text] [Medline: 8598068]
11. Sassi R, Cerutti S, Lombardi F, Malik M, Huikuri HV, Peng C, et al. Advances in heart rate variability signal analysis: joint position statement by the e-Cardiology ESC Working Group and the European Heart Rhythm Association co-endorsed by the Asia Pacific Heart Rhythm Society. Europace 2015 Sep;17(9):1341-1353. [doi: 10.1093/europace/euv015] [Medline: 26177817]
12. Billman GE. The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. Front Physiol 2013;4:26 [FREE Full text] [doi: 10.3389/fphys.2013.00026] [Medline: 23431279]
13. Quintana DS, Alvares GA, Heathers JA. Guidelines for Reporting Articles on Psychiatry and Heart rate variability (GRAPH): recommendations to advance research communication. Transl Psychiatry 2016 May 10;6:e803 [FREE Full text] [doi: 10.1038/tp.2016.73] [Medline: 27163204]
14. Tobaldini E, Nobili L, Strada S, Casali KR, Braghiroli A, Montano N. Heart rate variability in normal and pathological sleep. Front Physiol 2013 Oct 16;4:294 [FREE Full text] [doi: 10.3389/fphys.2013.00294] [Medline: 24137133]
15. Chouchou F, Desseilles M. Heart rate variability: a tool to explore the sleeping brain? Front Neurosci 2014;8:402 [FREE Full text] [doi: 10.3389/fnins.2014.00402] [Medline: 25565936]
16. Rieger A, Stoll R, Kreuzfeld S, Behrens K, Weippert M. Heart rate and heart rate variability as indirect markers of surgeons' intraoperative stress. Int Arch Occup Environ Health 2014 Feb;87(2):165-174. [doi: 10.1007/s00420-013-0847-z] [Medline: 23370764]
17. Brosschot JF, Geurts SA, Kruizinga I, Radstaak M, Verkuil B, Quirin M, et al. Does unconscious stress play a role in prolonged cardiovascular stress recovery? Stress Health 2014 Aug;30(3):179-187. [doi: 10.1002/smi.2590] [Medline: 25100269]
18. Föhr T, Tolvanen A, Myllymäki T, Järvelä-Reijonen E, Rantala S, Korpela R, et al. Subjective stress, objective heart rate variability-based stress, and recovery on workdays among overweight and psychologically distressed individuals: a cross-sectional study. J Occup Med Toxicol 2015;10:39 [FREE Full text] [doi: 10.1186/s12995-015-0081-6] [Medline: 26504485]
19. Koskinen P, Virolainen J, Kupari M. Acute alcohol intake decreases short-term heart rate variability in healthy subjects. Clin Sci (Lond) 1994 Aug;87(2):225-230. [Medline: 7924168]
20. Spaak J, Tomlinson G, McGowan CL, Soleas GJ, Morris BL, Picton P, et al. Dose-related effects of red wine and alcohol on heart rate variability. Am J Physiol Heart Circ Physiol 2010 Jun;298(6):H2226-H2231 [FREE Full text] [doi: 10.1152/ajpheart.00700.2009] [Medline: 20418480]

21. Sagawa Y, Kondo H, Matsubuchi N, Takemura T, Kanayama H, Kaneko Y, et al. Alcohol has a dose-related effect on parasympathetic nerve activity during sleep. Alcohol Clin Exp Res 2011 Nov;35(11):2093-2100. [doi: 10.1111/j.1530-0277.2011.01558.x] [Medline: 21848959]
22. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - what is it and what can it tell us? N Engl J Med 2016 Dec 08;375(23):2293-2297. [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]
23. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. Br J Cancer 2014 Feb 04;110(3):551-555 [FREE Full text] [doi: 10.1038/bjc.2013.725] [Medline: 24495873]
24. Garrison Jr LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. Value Health 2007;10(5):326-335 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00186.x] [Medline: 17888097]
25. Mutikainen S, Helander E, Pietilä J, Korhonen I, Kujala UM. Objectively measured physical activity in Finnish employees: a cross-sectional study. Br Med J Open 2014 Dec 10;4(12):e005927 [FREE Full text] [doi: 10.1136/bmjopen-2014-005927] [Medline: 25500160]
26. Firstbeat Technologies Ltd. Bodyguard 2 technical details URL: https://support.firstbeat.com/en/support/solutions/articles/9000114468-bodyguard-2-technical-details [WebCite Cache ID 6vJNo6JbP]
27. Parak J, Korhonen I. Firstbeat Technologies Ltd. 2013. Accuracy of Firstbeat Bodyguard 2 beat-to-beat heart rate monitor URL: https://assets.firstbeat.com/firstbeat/uploads/2015/11/white_paper_bodyguard2_final.pdf [WebCite Cache ID 6vJJnrXG6]
28. Saalasti S. Neural Networks for Heart Rate Time Series Analysis. Jyväskylä, Finland: University of Jyväskylä; 2003.
29. Kettunen J, Saalasti S. Patent Images. 2008. Procedure for detection of stress by segmentation and analyzing a heart beat signal URL: https://patentimages.storage.googleapis.com/c0/da/86/0013898102e400/US20050256414A1.pdf [WebCite Cache ID 6xsx0lFJX]
30. Firstbeat Technologies Ltd. 2014. Stress and Recovery Analysis Method Based on 24-hour Heart Rate Variability URL: https://assets.firstbeat.com/firstbeat/uploads/2015/10/Stress-and-recovery_white-paper_20145.pdf [WebCite Cache ID 6vJCwTnzu]
31. Ross RM, Jackson AS. Exercise Concepts, Calculations, and Computer Applications. Carmel, Indiana, USA: Benchmark Press; 1990.
32. Lin M, Lucas Jr HC, Shmueli G. Research commentary—too big to fail: large samples and the p-value problem -Value Problem. Inf Syst Res 2013 Dec;24(4):906-917. [doi: 10.1287/isre.2013.0480]
33. Ryan JM, Howes LG. Relations between alcohol consumption, heart rate, and heart rate variability in men. Heart 2002 Dec;88(6):641-642 [FREE Full text] [Medline: 12433903]
34. Boschloo L, Vogelzangs N, Licht CM, Vreeburg SA, Smit JH, van den Brin W, et al. Heavy alcohol use, rather than alcohol dependence, is associated with dysregulation of the hypothalamic-pituitary-adrenal axis and the autonomic nervous system. Drug Alcohol Depend 2011 Jul 01;116(1-3):170-176. [doi: 10.1016/j.drugalcdep.2010.12.006] [Medline: 21330066]
35. Rennie K, Hemingway H, Kumari M, Brunner E, Malik M, Marmot M. Effects of moderate and vigorous physical activity on heart rate variability in a British study of civil servants. Am J Epidemiol 2003 Jul 15;158(2):135-143. [Medline: 12851226]
36. Piazza-Gardner AK, Barry AE. Examining physical activity levels and alcohol consumption: are people who drink more active? Am J Health Promot 2012;26(3):e95-104. [doi: 10.4278/ajhp.100929-LIT-328] [Medline: 22208422]
37. Leasure JL, Neighbors C, Henderson CE, Young CM. Exercise and alcohol consumption: what we know, what we need to know, and why it is important. Front Psychiatry 2015;6:156 [FREE Full text] [doi: 10.3389/fpsyt.2015.00156] [Medline: 26578988]
38. Wiklund U, Karlsson M, Öström M, Messner T. Influence of energy drinks and alcohol on post-exercise heart rate recovery and heart rate variability. Clin Physiol Funct Imaging 2009 Jan;29(1):74-80. [doi: 10.1111/j.1475-097X.2008.00837.x] [Medline: 19016812]
39. Pietila J, Helander E, Myllymaki T, Korhonen I, Jimison H, Pavel M. Exploratory analysis of associations between individual lifestyles and heart rate variability -based recovery during sleep. Conf Proc IEEE Eng Med Biol Soc 2015;2015:2339-2342. [doi: 10.1109/EMBC.2015.7318862] [Medline: 26736762]
40. Clark AJ, Salo P, Lange T, Jennum P, Virtanen M, Pentti J, et al. Onset of impaired sleep as a predictor of change in health-related behaviours; analysing observational data as a series of non-randomized pseudo-trials. Int J Epidemiol 2015 Jun;44(3):1027-1037. [doi: 10.1093/ije/dyv063] [Medline: 25969504]
41. Paunio T, Korhonen T, Hublin C, Partinen M, Koskenvuo K, Koskenvuo M, et al. Poor sleep predicts symptoms of depression and disability retirement due to depression. J Affect Disord 2015 Feb 01;172:381-389 [FREE Full text] [doi: 10.1016/j.jad.2014.10.002] [Medline: 25451441]

Abbreviations

ANS: Autonomic nervous system
ANOVA: Analysis of variance

BMI: Body mass index
HF: high frequency
LF: low frequency
HR: heart rate
HRV: heart rate variability
PA: physical activity
RMSSD: root mean square of the successive differences

Edited by G Eysenbach; submitted 29.11.17; peer-reviewed by J Spaak, S Bhatlekar, T Riestra; comments to author 16.12.17; revised version received 01.02.18; accepted 27.02.18; published 16.03.18

Please cite as:
Pietilä J, Helander E, Korhonen I, Myllymäki T, Kujala UM, Lindholm H
Acute Effect of Alcohol Intake on Cardiovascular Autonomic Regulation During the First Hours of Sleep in a Large Real-World Sample of Finnish Employees: Observational Study
JMIR Ment Health 2018;5(1):e23
URL: http://mental.jmir.org/2018/1/e23/
doi:10.2196/mental.9519
PMID:29549064

PUBLICATION
V

Physical Activity, Body Mass Index and Heart Rate Variability-Based Stress and Recovery in 16 275 Finnish Employees: a Cross-Sectional Study

FÖHR, T., PIETILÄ, J., HELANDER, E., MYLLYMÄKI, T., LINDHOLM, H., RUSKO, H. and KUJALA, U.M.

BMC Public Health, 16:701
DOI: 10.1186/s12889-016-3391-4

Föhr *et al. BMC Public Health* (2016) 16:701
DOI 10.1186/s12889-016-3391-4

BMC Public Health

RESEARCH ARTICLE Open Access

Physical activity, body mass index and heart rate variability-based stress and recovery in 16 275 Finnish employees: a cross-sectional study

Tiina Föhr[1*], Julia Pietilä[2], Elina Helander[2], Tero Myllymäki[3], Harri Lindholm[4], Heikki Rusko[5] and Urho M. Kujala[1]

Abstract

Background: Physical inactivity, overweight, and work-related stress are major concerns today. Psychological stress causes physiological responses such as reduced heart rate variability (HRV), owing to attenuated parasympathetic and/or increased sympathetic activity in cardiac autonomic control. This study's purpose was to investigate the relationships between physical activity (PA), body mass index (BMI), and HRV-based stress and recovery on workdays, among Finnish employees.

Methods: The participants in this cross-sectional study were 16 275 individuals (6863 men and 9412 women; age 18–65 years; BMI 18.5–40.0 kg/m^2). Assessments of stress, recovery and PA were based on HRV data from beat-to-beat R-R interval recording (mainly over 3 days). The validated HRV-derived variables took into account the dynamics and individuality of HRV. Stress percentage (the proportion of stress reactions, workday and working hours), and stress balance (ratio between recovery and stress reactions, sleep) describe the amount of physiological stress and recovery, respectively. Variables describing the intensity (i.e. magnitude of recognized reactions) of physiological stress and recovery were stress index (workday) and recovery index (sleep), respectively. Moderate to vigorous PA was measured and participants divided into the following groups, based on calculated weekly PA: inactive (0 min), low (0 < 150 min), medium (150–300 min), and high (>300 min). BMI was calculated from self-reported weight and height. Linear models were employed in the main analyses.

Results: High PA was associated with lower stress percentages (during workdays and working hours) and stress balance. Higher BMI was associated with higher stress index, and lower stress balance and recovery index. These results were similar for men and women ($P < 0.001$ for all).

Conclusion: Independent of age and sex, high PA was associated with a lower amount of stress on workdays. Additionally, lower BMI was associated with better recovery during sleep, expressed by a greater amount and magnitude of recovery reactions, which suggests that PA in the long term resulting in improved fitness has a positive effect on recovery, even though high PA may disturb recovery during the following night. Obviously, several factors outside of the study could also affect HRV-based stress.

Keywords: Body mass index, Heart rate variability, Physical activity, Physiological stress, Stress, Stress assessment

* Correspondence: tiina.fohr@jyu.fi
[1]Department of Health Sciences, University of Jyväskylä, P.O. Box 35FIN-40014 Jyväskylä, Finland
Full list of author information is available at the end of the article

Background

Physical activity (PA) is known to have positive effects on health [1, 2]. Routine PA reduces stress and enhances psychological wellbeing, which is particularly important for the prevention and management of cardiovascular disease, among other chronic diseases [3]. Regular PA is known to reduce the risk of many adverse health outcomes. Some PA is better than none; however, for most health outcomes, additional benefits are achieved if the amount of PA increases through higher intensity, greater frequency, and/or longer duration. According to the 2008 Physical Activity Guidelines for Americans, most health benefits occur with at least 150 total minutes of moderate intensity or at least 75 min of vigorous intensity aerobic PA per week. However, additional benefits occur with more PA [4, 5]. In addition to the beneficial effects of PA on physical health, these guidelines are also relevant for mental health [6]. Although leisure-time PA has increased among Finnish adults [7], physical inactivity is a major problem and risk for health, in all countries. Furthermore, physical inactivity is associated with being overweight [8] and the current rate of overweight adults worldwide has been described as an epidemic or even a pandemic. This situation is a major public health risk because being overweight is associated with diseases including coronary heart disease, stroke, diabetes and cancer [9].

Together with physical inactivity and overweight, stress at work is a major public health risk. It may even lead to cardiovascular disease [10] without complete recovery [11]. Stress has been shown to reduce participation in leisure-time PA [12, 13]. Furthermore, workplace stress may predict a future increased risk of insufficient PA [14]. Normal weight is associated with good self-reported subjective health [15], including low stress levels [16, 17]. Evidence suggests that psychosocial stress is associated with the development of adiposity [18]. However, according to previous studies, the association between subjective stress and body composition is inconsistent, with evidence both supporting [16, 19] and refuting [20, 21] the idea that stress is associated with adiposity. A recent systematic review reported that the associations of psychosocial factors at work with weight-related outcomes were weak and somewhat inconsistent [22].

Psychological stress causes sympathetic responses in the autonomic nervous system (ANS), such as reduced heart rate variability (HRV) [23]. HRV refers to the variation in intervals between heartbeats and reflects cardiac autonomic modulation. Physiological stress can be defined as an increased body activation level, when sympathetic activity dominates the ANS and parasympathetic activation is low. Stress is associated with reduced HRV, owing to attenuated parasympathetic and/or increased sympathetic activity in cardiac autonomic control. Recovery refers to a reduced body activation level, when parasympathetic activation dominates the ANS over sympathetic activity [24–26]. HRV analysis can be used as a complementary tool to assess general health [27]. HRV analysis during sleep has the potential to explore the sleeping brain, with possible implications for mental health [28]. Previous HRV-studies have mainly used traditional time-domain and frequency-domain measures of HRV, such as root mean square of successive R-R intervals (RMSSD) and the ratio of low frequency power to high frequency power (LF/HF ratio). The traditional measures of HRV represent the average level of the autonomic activity over a period of the time. Cardiac autonomic activity is very dynamic and varies during the day depending on stress, recovery and PA. Therefore, the usability of the traditional measures of HRV is limited in real-life conditions. Additionally, these measures are very individual which further limits their usability in stress assessment and clinical work. However, it is also possible to provide applied heart rate (HR) and HRV-derived stress and recovery variables that take into account the dynamic changes in autonomic activity and individuality of HRV including information that is difficult to obtain from traditional measures of HRV.

The majority of previous studies on the association of PA with stress have used subjective assessment methods or traditional measures of HRV in the assessment of stress. The previous studies support the association of PA with increased HRV [29, 30]. However, accurate and objective methods are needed to reliably assess PA, as well as to assess HRV-based stress and recovery in real-life. By utilizing a method that acknowledges the dynamics and individuality in HRV in real-life, the aim of this study was to investigate the extent to which PA and BMI are associated with HRV-based indicators of stress and recovery on workdays. The study was conducted among 16 275 Finnish employees who had participated in beat-to-beat R-R interval recording as a part of lifestyle counseling between 2007 and 2015. More specifically, accounting for age and sex, we investigated the prevalence of stress and recovery according to the participants' objectively measured PA level and self-reported body mass index (BMI). Uniqueness of the present study is in the individual and dynamic method used in the assessment of physiological stress and recovery.

Methods

Study design and participants

This cross-sectional study investigated the amount and intensity of objective HRV-based stress and recovery on workdays in a real-life sample of 16 275 Finnish employees (6863 men and 9412 women; age 18–65 years; BMI 18.5–40.0 kg/m^2). The participants nonselectively represent a cross-section of typical Finnish employees including both manual and non-manual labour employees.

The majority of the participants were apparently healthy without chronic diseases. The exclusion criteria for participation in the R-R interval recordings included severe cardiac disease, very high blood pressure (≥180/100 mmHg), type 1 or 2 diabetes with autonomic neuropathy, severe neurological disease, fever or other acute disease, and BMI >40 kg/m^2. These exclusion criteria represented by the analysis software manufacturer are presented in detail previously [31]. The characteristics of the participants are presented in Table 1.

Data collection

The novel methodology used to determine the participants' stress, recovery and level of weekly PA, was based on HRV data from beat-to-beat R-R interval recordings. These recordings were voluntarily performed on employees as a part of the preventive occupational health care programs provided by their employers between 2007 and 2015. The clinical purpose of these measurements is presented comprehensively in a previous paper by Mutikainen et al. [31]. The data recordings used in the previous study were gathered between 2007 and 2013, with a study population size of 9554. These data were used in the present study, supplemented with recordings from 2014 to 2015. The study had a further inclusion criterion of a minimum of 4.5 h beat-to-beat R-R interval recording during sleep after a workday. Another inclusion criterion was the availability of R-R interval data, including at least one workday (≥4 h of work) and one day off, with a measurement period of 16–30 h/day (from wake-up to wake-up). Participants who had consumed alcohol on the monitoring days were excluded. Information about workdays, working hours, days off and sleep periods was obtained from diaries that the participants were requested to keep during the measurement period. The analyzed data consisted of successfully recorded (measurement error <15 % and <30-min recording break) days. The flow of the participants included in the analysis is presented in Fig. 1.

HRV-based assessment of PA, stress and recovery

Ambulatory beat-to-beat R-R interval data were used to determine the amount and intensity of PA, stress and recovery. Using the Firstbeat Bodyguard device (Firstbeat Technologies Ltd., Jyväskylä, Finland), real-life R-R interval data were recorded, usually over 3 days (typically two workdays and one day off) and analyzed using Firstbeat Analysis Server software (version 6.3, Firstbeat Technologies Ltd.), which included a powerful artifact detection and correction feature for irregular ectopic beats and signal noise. The software calculates HRV indices second-by-second using the short-time Fourier transform method, and calculates HR- and HRV-derived variables of respiration rate, oxygen consumption, on-off kinetics (increasing or decreasing HR), and parameters describing excess post-exercise oxygen consumption using neural networks. Thereafter, the software divides the measurement data into coherent data segments and categorizes these segments into different physiological states, such as PA of different intensities, stress and recovery [32–34], by taking into account individual characteristics (e.g. individual levels and scales of HR and HRV, and the individual relationships between HRV and autonomic control) [35]. The categorization of the data is described in Additional file 1: Table S1. More information about this analysis method is available in a paper by Firstbeat Technologies Ltd. [36].

Detection of stress and recovery variables

After data categorization, the HRV-based variables describing the amount and intensity of stress and recovery on workdays were detected. Stress percentages (i.e. proportions of stress reactions, during the day and during working hours) and stress balance (ratios between recovery and stress reactions during sleep) describe the amount of stress and recovery, respectively. The variables describing the intensity (i.e. magnitude of recognized reactions) of stress and recovery were stress index (during the day) and recovery index (during sleep), respectively. These variables and their calculations are presented in Additional file 1: Table S1. The correlation coefficient between two consecutive workdays varied from 0.74 to 0.88 for the traditional HRV variables, from 0.64 to 0.93 for HRV-derived variables of stress, and from 0.42 to 0.49 for HRV-derived variables of recovery during sleep.

Calculation of weekly PA

Background information about age, sex, self-reported height and weight, and self-reported PA class [37] modified from Ross and Jackson [38], was collected in

Table 1 Characteristics of the participants

Variable	All ($n = 16275$)			Men ($n = 6863$)			Women ($n = 9412$)		
	Mean ± SD	Min	Max	Mean ± SD	Min	Max	Mean ± SD	Min	Max
Age	44.8 ± 9.9	18.0	65.0	44.5 ± 9.9	18.0	65.0	45.0 ± 9.9	18.0	65.0
Body mass index (kg/m^2)	26.0 ± 4.1	18.5	40.0	26.6 ± 3.5	18.6	40.0	25.5 ± 4.4	18.5	40.0
Self- reported activity class 0–10	4.8 ± 1.8	0.0	10.0	4.9 ± 1.9	0.0	10.0	4.8 ± 1.8	0.0	10.0
Physical activity (mins/week)	186 ± 227	0	2629	246 ± 258	0	2629	142 ± 189	0	1865

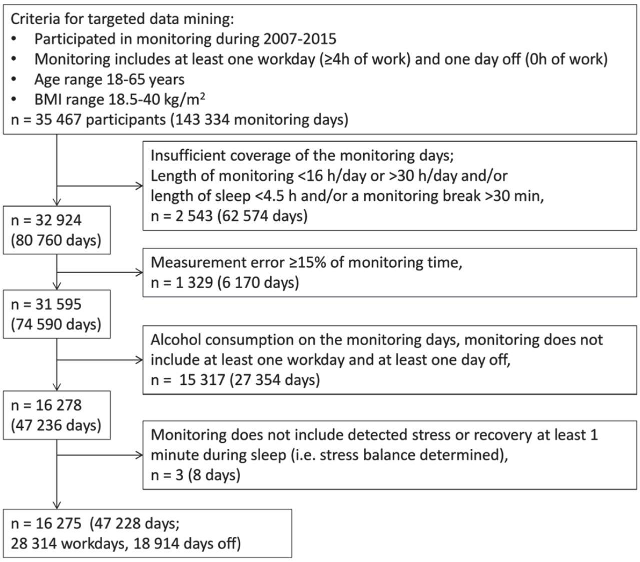


Fig. 1 Flow of the participants and measurement days included in the analysis (BMI, body mass index)

conjunction with R-R interval recordings using questionnaires. Background information was used to estimate maximal HR [39] and maximal VO_2 [40] which were then used in the estimation of VO_2. The maximal HR used for further calculations was corrected accordingly if a period with HR higher than the estimated maximal was found from the recording. From the second-by-second VO_2 estimations, each participant's mean VO_2 for each minute of the measurement day was calculated. The minute-by-minute VO_2 estimations were then converted to multiples of the resting metabolic rate (MET) by dividing the VO_2 values by 3.5. The total number of 1-min segments within the following thresholds: moderate PA 3 to <6 METs and vigorous PA ≥6 METs, during each measurement day (including days off), were calculated. Continuous bouts of PA lasting for ≥10 min were included in the estimation of weekly PA. These continuous bouts of PA were calculated separately for workdays and days off, and, if the measurement period included two or more workdays or days off, an average was calculated. The activity minutes score for each day (moderate PA minutes + vigorous PA minutes × 2) was calculated. Thereafter, the amount of PA was extrapolated using the following formula: PA minutes per week = (5 × mean workday activity score) + (2 × mean day-off activity score). These calculations have been previously described in more detail [31]. Based on the weekly PA minutes, the participants were divided into the following PA groups: inactive (0 min), low (0 < 150 min), medium (150–300 min) and high (>300 min).

Assessment of body composition

BMI was calculated from the self-reported weight and height (kg/m^2). The participants were then divided into the following groups: normal weight (18.5 to <25 kg/m^2), overweight (25 to <30 kg/m^2) and obese (30–40 kg/m^2).

Analysis

Data processing and statistical analysis were performed using R 3.2.2 version (R Foundation for Statistical Computing). P-values were two-sided and a p-value of <0.05 was considered statistically significant. Because of the size of the data, 99 % confidence intervals (CIs) were determined (Fig. 2) instead of conventional 95 % CIs.

The main outcome variables of the study were stress percentage and stress index, calculated for the whole day, stress percentage calculated for working hours, and stress balance and recovery index calculated for sleep. These variables were derived from the beat-to-beat R-R interval recordings on workdays. For a more detailed description of the variables see Additional file 1: Table S1. In addition, HR and traditional HRV parameters, including RMSSD and the LF/HF ratio, were calculated from the beat-to-beat R-R interval recordings on workdays. These variables were calculated separately for waking hours and sleep, and RMSSD was calculated using a 5-min window. If the measurement period of a subject included two or more workdays, an average was calculated and the mean values of the outcome variables were used in the analysis.

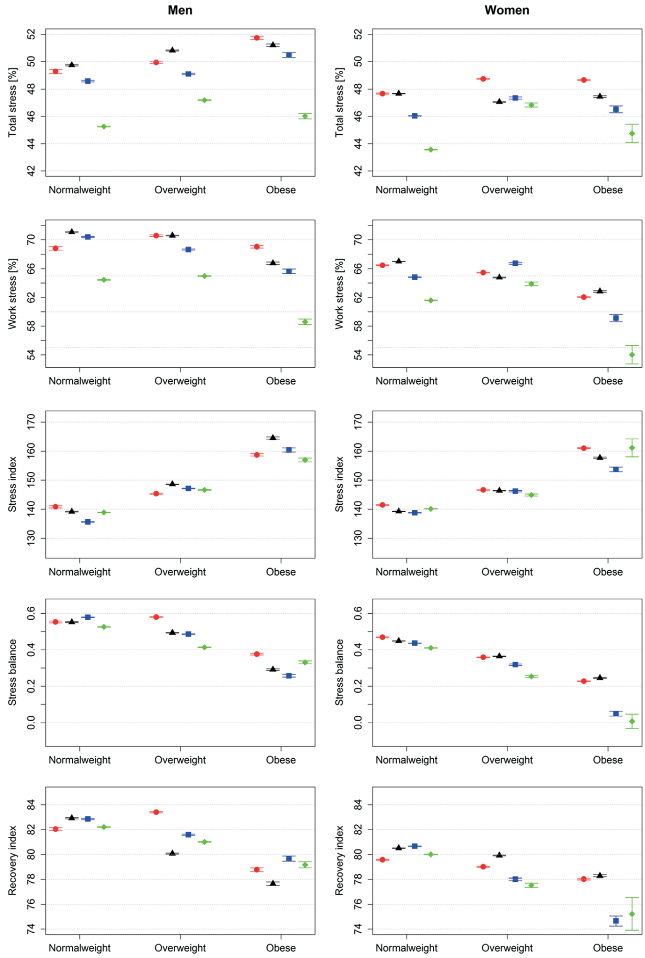


Fig. 2 Stress and recovery by physical activity and body mass index groups with age-controlled mean values and 99 % CIs for output variables. Physical activity groups: inactive (0 min/week), *red* ●; low (0 < 150 min/week), *black* ▲; medium (150–300 min/week), *blue* ■; high (>300 min/week), *green* ♦

For the descriptive statistics, the means and standard deviations of the outcome variables were calculated separately for men and women, and stratified based on PA, BMI and age. Differences in the outcome variables between PA, BMI and age groups were tested using the Kruskal-Wallis test. The results are shown in Additional file 2: Tables S2-S4. To show the effects of BMI and PA group on the HRV-based stress and recovery variables, the age-controlled mean values and 99 % CIs for the HRV-based stress and recovery variables, by BMI and PA group, are presented in Fig. 2.

Linear models were employed to study the effects of PA group, BMI and age on HRV-based stress and recovery variables. In the models, age and BMI were incorporated as continuous predictor variables and objectively measured PA group was incorporated as a categorical predictor variable. The models were generated separately for men and women. The reference value for age was set to 18 years and for BMI to 18.5 kg/m^2. A simple linear least squares regression model (procedure *lm* in R) was applied to predict the stress percentage during the day. As confirmed by visual inspection, the assumption of linear regression considering the normal distribution of the residuals was not fulfilled for stress percentage during working hours, stress index and recovery index. Thus, a Box-Cox transformation was applied on these dependent variables [41]. The Box-Cox coefficient was determined by maximizing the log-likelihood function and was rounded to two decimal places before transformation. Tobit regression model (procedure *vglm* using iteratively reweighted least squares in R) was applied for modeling stress balance with a fixed lower and upper limit of −1 and 1, respectively. The interactions of the predictors were not included in the final regression models because the coefficient of determination for the interaction models was only a few percentage points greater than for the simple models.

Results

The total number of workdays included in the analysis was 28 314, with measurements obtained from 16 275 participants (men 6863; women 9412). The participants' characteristics are shown in Table 1. The participants' mean age was 44.8 years (44.5 years for men; 45.0 years for women) and the mean BMI was 26.0 kg/m^2 (26.6 kg/m^2 for men; 25.5 kg/m^2 for women). The participants' mean self-reported activity class was 4.8 (4.9 for men; 4.8 for women) indicating that, on average, the participants were involved in PA 2–3 times per week and their total weekly PA was about 2 h. The mean weekly minutes of objective monitoring-based PA was 186 (246 for men; 142 for women). The mean weekly minutes of PA in the group of low PA was 78 for men and 74 for women, in the group of medium PA 222 for men and 215 for women, and in the group of high PA 545 for men and 496 for women. The number of participants in the PA, BMI, and age groups is presented in Additional file 3: Table S5.

Differences in outcome variables between PA groups (Additional file 2: Table S2) were statistically significant except for LF/HF ratio during waking hours and sleep, stress balance in men, and HR and stress balance in women. For both men and women, the high PA group had the highest RMSSD (during waking hours and during sleep) and recovery index, and the lowest stress percentage (during the day and during working hours) and stress index.

Differences in outcome variables between BMI groups were statistically significant for both men and women (Additional file 2: Table S2). Normal-weight individuals had the highest RMSSD (both during waking hours and during sleep), stress balance and recovery index, the lowest stress percentage during the day and the lowest stress index. Stress percentage during working hours was lowest in obese individuals.

In both sexes, differences in outcome variables were statistically significant between age groups, except for HR during sleep in women (Additional file 2: Table S3). The youngest age group (18–30 years) had the highest RMSSD (both during waking hours and during sleep), stress balance and recovery index, and the lowest stress index. Stress percentages during the day were lowest in the youngest age group in women, and in the oldest age group (51–65 years) in men. Stress percentages during working hours were lowest in the oldest age group in both sexes.

Figure 2 shows the effect of PA and BMI group on the stress and recovery variables with the effect of age controlled. The high PA group had the lowest mean stress percentage during the day and during working hours in all three BMI groups, after adjustment for age (Fig. 2). Mean stress index values increased as the BMI group changed from normal weight to overweight and overweight to obese, regardless of the PA group. In addition, regardless of the PA group, obese individuals had the lowest stress balance and recovery index.

The linear model results are shown in Table 2. Medium ($P < 0.05$) and high ($P < 0.001$) PA groups, lower BMI ($P < 0.001$), and older age ($P < 0.001$) were associated with lower stress percentages during the day. Medium ($P < 0.05$) and high ($P < 0.001$) PA level, higher BMI ($P < 0.001$), and older age ($P < 0.001$) were associated with lower stress percentages during working hours. Stress percentage results during the day and during working hours were similar for men and women. Higher BMI ($P < 0.001$) and older age ($P < 0.001$) were associated with higher stress index, both in men and in women. In addition, medium ($P < 0.01$) and high ($P < 0.01$) PA were associated with lower stress index in women.

Medium ($P < 0.01$) and high ($P < 0.001$) PA, and higher BMI ($P < 0.001$) were associated with lower stress balance, both in men and in women. Moreover, older age was associated with lower stress balance in men ($P < 0.001$). Higher BMI and older age were associated with lower recovery index, in men and in women ($P < 0.001$). BMI explained the highest proportion of variance in stress balance (2.2 % for men and 3.1 % for women) compared with PA and age.

Discussion

The purpose of this study was to investigate the amount and intensity of objective HRV-based stress and recovery on workdays. The sample group comprised 16 275 Finnish employees, who had participated in beat-to-beat R-R interval recording as a part of lifestyle counseling in the course of their everyday lives between 2007 and 2015. More specifically, the relationships between PA, BMI, and HRV-based stress and recovery were investigated. For both sexes, a high level of PA and lower BMI

Table 2 Results of the linear models

	Men					Women				
	Parameter Estimate	95 % CI Lower	Upper	*P* value	Variance explained (%)	Parameter Estimate	95 % CI Lower	Upper	*P* value	Variance explained (%)
Stress (%), 24 h [a]					2.356[h]					1.476 [h]
Intercept	52.3153	50.7730	53.8575	<0.001		50.0658	48.9790	51.1527	<0.001	
Age (18 years = 0)	−0.1263	−0.1600	−0.0926	<0.001	0.780[i]	−0.0869	−0.1151	−0.0587	<0.001	0.387 [i]
BMI (18.5 kg/m^2 = 0)	0.1541	0.0580	0.2502	0.002	0.144 [i]	0.0909	0.0264	0.1553	0.006	0.081 [i]
Physical activity level (inactive = 0)					1.586 [i]					1.050 [i]
Low physical activity class	0.3235	−0.6903	1.3372	0.53		−0.8892	−1.5471	−0.2313	0.008	
Medium physical activity class	−1.1405	−2.2144	−0.0667	0.04		−1.9539	−2.7608	−1.1470	<0.001	
High physical activity class	−3.9695	−5.0062	−2.9328	<0.001		−4.3772	−5.2620	−3.4923	<0.001	
Stress (%), working hours [b]					4.152 [h]					1.721 [h]
Intercept	807.3480	777.0476	837.6483	<0.001		677.8091	656.0738	699.5443	<0.001	
Age (18 years = 0)	−4.8729	−5.5355	−4.2103	<0.001	2.942 [i]	−2.66482	−3.22838	−2.10127	<0.001	0.905 [i]
BMI (18.5 kg/m^2 = 0)	−4.6640	−6.5515	−2.7766	<0.001	0.341 [i]	−4.54067	−5.82984	−3.25149	<0.001	0.504 [i]
Physical activity level (inactive = 0)					1.590 [i]					0.584 [i]
Low physical activity class	−5.5094	−25.4261	14.4073	0.59		−3.40975	−16.5664	9.746938	0.61	
Medium physical activity class	−24.7847	−45.8819	−3.6874	0.021		−17.9339	−34.0695	−1.79829	0.030	
High physical activity class	−85.2896	−105.6576	−64.9215	<0.001		−61.7878	−79.4833	−44.0923	<0.001	
Stress index, 24 h [c]					27.113 [h]					27.448 [h]
Intercept	1.2406	1.2403	1.2409	<0.001		1.2414	1.2412	1.2417	<0.001	
Age (18 years = 0)	0.0001	0.0001	0.0001	<0.001	21.992 [i]	0.0001	0.0001	0.0001	<0.001	19.655 [i]
BMI (18.5 kg/m^2 = 0)	0.0002	0.0001	0.0002	<0.001	3.819 [i]	0.0001	0.0001	0.0001	<0.001	3.031 [i]
Physical activity level (inactive = 0)					0.115 [i]					0.145 [i]
Low physical activity class	0.0002	0.0000	0.0004	0.024		−0.0001	−0.0003	0.0000	0.031	
Medium physical activity class	0.0000	−0.0002	0.0002	0.91		−0.0002	−0.0004	−0.0001	0.002	
High physical activity class	0.0001	−0.0001	0.0003	0.47		−0.0003	−0.0004	−0.0001	0.002	
Stress balance, sleep [e]					3.669 [h]					3.244 [h]
Intercept1	0.9975	0.9287	1.0663	<0.001		0.6166	0.5672	0.6661	<0.001	
Intercept2	−0.5292	−0.5503	−0.5082	<0.001		−0.5537	−0.5705	−0.5368	<0.001	
Age (18 years = 0)	−0.0066	−0.0081	−0.0052	<0.001	1.153 [i]	−0.0007	−0.0019	0.0006	0.31	0.006 [i]
BMI (18.5 kg/m^2 = 0)	−0.0273	−0.0315	−0.0231	<0.001	2.179 [i]	−0.0253	−0.0282	−0.0224	<0.001	3.080 [i]
Physical activity level (inactive = 0)					0.445 [i]					0.233 [i]
Low physical activity class	−0.0715	−0.1162	−0.0269	0.002		−0.0126	−0.0423	0.0172	0.41	

Table 2 Results of the linear models *(Continued)*

Medium physical activity class	−0.0799	−0.1272	−0.0326	<0.001		−0.0572	−0.0937	−0.0207	0.002	
High physical activity class	−0.1327	−0.1784	−0.0870	<0.001		−0.0822	−0.1222	−0.0421	<0.001	
Recovery index, sleep [f]					9.685 [h]					3.297 [h]
Intercept	592375.4298	573358.1497	611392.7100	<0.001		457913.8301	444648.7930	471178.8673	<0.001	
Age (18 years = 0)	−5092.7330	−5508.5947	−4676.8712	<0.001	7.753 [i]	−2100.7061	−2444.6430	−1756.7692	<0.001	1.501 [i]
BMI (18.5 kg/m^2 = 0)	−4825.0722	−6009.6783	−3640.4662	<0.001	0.921 [i]	−4038.5841	−4825.3668	−3251.8014	<0.001	1.065 [i]
Physical activity level (inactive = 0)					0.154 [i]					0.026 [i]
Low physical activity class	−19214.5686	−31714.7934	−6714.3437	0.003		1051.0555	−6978.4656	9080.5766	0.80	
Medium physical activity class	−6153.1124	−19394.2693	7088.0445	0.36		−4452.2347	−14299.7786	5395.3093	0.38	
High physical activity class	−9653.5430	−22437.0465	3129.9606	0.14		−5732.9247	−16532.4796	5066.6302	0.30	

BMI body mass index
[a] Linear regression
[b] Box-Cox linear regression using transformation coefficient of 1.62
[c] Box-Cox linear regression using transformation coefficient of −0.79
[e] Tobit regression
[f] Box-Cox linear regression using transformation coefficient of 3.18
[h] The proportion of variance explained by the whole model
[i] The proportion of variance explained by the predictor variable. Calculated as the difference between the proportion of variance explained by the whole model and the proportion of variance explained by a model including all the predictor variables, except for the predictor in question

were associated with lower amounts of stress on workdays. Additionally, the results showed that both high PA and higher BMI were associated with a lower amount of recovery during sleep. Additional PA (above the generally recommended aerobic PA level of over 150 min of moderate PA per week), was associated with the additional health benefits of a low amount of HRV-based stress on workdays and during working hours. Lower BMI was associated with better recovery during sleep, expressed by a greater amount and magnitude of recovery reactions (i.e. quality of recovery). This suggests that PA in the long term resulting in improved physical fitness has a positive effect on recovery, even though high PA may disturb recovery during the following night. The results of the present study showing an association of BMI and objectively measured PA with HRV-based stress during the workday are in line with previous studies.

The finding of the present study on the association of high PA with low HRV-based stress on workdays is in line with previous studies. Both moderate and vigorous PA are found to be associated with higher HRV [29]. Additionally, PA has been found to have positive effects on subjective stress. For instance, Birdee et al. [42] found that, among a large group of employees, physically active employees reported less difficulty coping with stress, more happiness and a higher rate of competency than inactive employees. Our previous study used the same measurement method to assess stress as in the present study, and found higher PA and physical fitness were associated with lower stress among men [43]. However, to our knowledge, this study is unique in its focus on the additional health benefits from PA exceeding the recommended level, in the context of stress. The results showed that PA level affects stress percentage more than BMI, especially in women, and the decrease in the amount of stress following a change from inactivity to high PA appears to be impossible to achieve by weight loss alone. When stress percentage was calculated without the time spent on PA, the association between higher PA with lower stress percentage remained (data not shown).

The present findings of an association between lower BMI and lower amount of stress on workdays, and an association of higher BMI with lower amount of stress during working hours, are also in line with previous studies. Furthermore, an additional analysis (data not shown) showed that having a higher BMI was associated with a higher amount of PA during working hours. Previously, HRV profiles were found to be relatively poor among obese individuals [44] and improved after weight loss [45]. Previous studies also suggest that individuals with lower socioeconomic status are more likely to be obese and more likely to be in physically active employment than their counterparts with higher socioeconomic status [46, 47]. So, time spent in PA leads to less time for other physiological body states, such as stress, during working hours. Another possible explanation for these findings is that among obese individuals, the physiological state of the body is detected as PA instead of stress, as HR increases and HRV decreases easily. Therefore, caution is required when interpreting the results. Previously, the association of BMI with stress has been studied using mainly subjective methods. For instance, Nyberg et al. [16] found both obesity and being underweight to be associated with high levels of work-related stress [16], independent of sex. Additionally, employees of normal weight report the lowest prevalence of emotional exhaustion and chronic psychological complaints compared with underweight, overweight and obese individuals [17]. In general, the evidence is weak and inconsistent for associations of psychosocial factors at work with weight-related outcomes [22]. However, based on previous [43–45] and present results, the association of obesity with HRV and HRV-based stress seems to be consistent.

The group of high active, consisting largely of young and normal-weight individuals, had the best quality of recovery during sleep when age and weight were not taken into account. The linear models revealed that BMI and age explained greater proportion of variance in the quality of recovery than the level of PA. Further, the linear models showed that the non-significant association of high PA with lower quality of recovery during sleep was negative. Additionally, high PA was significantly associated with lower amount of recovery during sleep. This finding may be explained by the estimation of PA level occurring on the same days that stress and recovery during sleep were determined. We did not take into account the timing of PA in the analysis of the present study. The findings of Myllymäki et al. [48] suggest that vigorous late-night exercise may have effects on cardiac autonomic control of heart during the first sleeping hours. They found higher nocturnal HR after the exercise day compared to the control day but no differences between the days in nocturnal HRV. Additionally, previous literature suggests that PA during working hours and leisure-time may show different effects on cardiac autonomic regulation. High PA during working hours has been found to be associated with poor cardiovascular health, including reduced HRV [49]. Recovery of HRV is also dependent on training background, and type, intensity and duration of exercise [50]. The lower the physical fitness and the higher the intensity of exercise, the slower the recovery of HRV after exercise [51]. Hynynen et al. [52] reported that even an exercise that was perceived as light and easy may have prolonged effects on nocturnal HRV during the following night. Our previous study with a smaller study population used a subjective method to assess PA; using laboratory conditions to assess physical fitness, our previous results showed that physical fitness was associated with better recovery during sleep, even though PA was not [43]. In line with this Pietilä et al. [53] found good physical fitness to be associated with good recovery, even though PA was found to disturb the recovery of the following night. It appears that PA on the same day may disturb nighttime recovery, but in the long term, PA and good fitness enhance recovery during sleep. This is supported by our additional analysis (data not shown), which showed higher recovery in a day without PA compared with a day with PA, among high-PA individuals. Further, the present finding of an association of lower BMI with a higher amount and better quality of recovery supports the idea that good fitness enhances recovery during sleep. However, the effect of the timing of PA on HRV-based recovery during following night should be studied further.

The present findings suggest that, although older individuals are not stressed as often, their stress reactions are stronger and recovery is weaker than their younger counterparts. Weaker recovery among older individuals was expected, as it is known that aging reduces HRV [24]. Compared with PA and BMI, age was most strongly associated with the amount of stress during working hours, and intensity of stress and recovery reactions. These results suggest that recovery of older individuals is weakened. However, the findings of Soares-Miranda et al. [30] showing both cross-sectional and longitudinal association of PA with more favorable HRV among older adults emphasizes the importance of PA among older individuals. These findings should be considered for instance in policymaking when planning to lengthen working careers.

The results were similar between men and women. The variances explained by the linear models were mostly slightly higher for men. The men in this study had a slightly higher amount of stress than women during the whole workday and during working hours. Men also had a higher intensity of stress reactions and a lower amount and quality of recovery during sleep compared with women. This finding is in line with previous evidence that men have stronger physiological responses to psychological stress than women, including greater cardiovascular activation [18].

This study has strengths and weaknesses. While the measurement method may have had a significant impact on the measured PA levels, a strength of this study is that the weekly PA amount was calculated based on the objective measurement of PA periods lasting over 10 min. The validated ambulatory beat-to-beat R-R interval-based method [31, 54] used to assess the amount and intensity of PA has been shown to provide more accurate estimates of the intensity of PA than HR information [54, 55]. Even though the participants were informed to continue with normal daily living under the wellness assessment, individuals may have a tendency to be more active than usual during this type of short-time assessment. At least 5 consecutive days of pedometer monitoring has been suggested to achieve reliable and valid 1-year PA estimates [56]. However, another study suggests that three days would be sufficient to achieve valid results [57]. The existing literature indicates a need for valid, accurate and reliable measures of PA for assessing current and changing PA levels and the relationships between PA and health outcomes [58].

We used a novel HRV-based method to assess the amount and intensity of stress and recovery. HRV has been suggested as a feasible stress assessment method [59–61], and the stability of 24-h recording is high [24]. In our study, the sustainability of the HRV and HRV-based measures of stress and recovery between 2 consecutive days was quantified, and all the correlations were found to be statistically significant. The method used in the present study has been validated against neuroendocrine responses to stress, and the indicators of stress and recovery during sleep have been found to be associated with free salivary cortisol response after awakening [62]. Additionally, the method has been utilized in previous studies [43, 63–65] and the findings of these studies further support the validity and reliability of this HRV-based method. For instance, previous studies have found an association of higher HRV-based stress and lower recovery with higher perceived stress [63, 64]. Although traditional HRV measures are required to assess quality and clinical correlates of the recordings, these new ways of presenting findings improve the usability of HRV recordings in health promotion. Traditional HRV measures are not included in the main study analyses. However, the descriptive statistics show the similarity between the traditional HRV measures and the HRV-based stress and recovery variables. HRV-based methods that take individual characteristics and dynamic changes in cardiac autonomic activity into account and provide easily understandable variables of stress and recovery can be informative and suitable measures for field and clinical conditions. Individual written feedback together with verbal feedback and discussion of the HRV recording results would be optimal (an example of the feedback the participants received is shown in Additional file 4: Figure S1). It should be noted here that the method we used did not distinguish eustress from distress. However, division into these two types of stress may be impractical because of similar physiological responses to both stress forms.

The major strength of the present study is the very large sample, which included both non-manual and manual labor employees. The study sample was not a random sample from the Finnish population, but a real-life sample of Finnish employees who voluntary performed beat-to-beat R-R interval recording as a part of the preventive occupational health care programs provided by their employers. Even though the participants of the present study may represent a group of employees who are more interested about their health than the average person, their BMI profiles were similar to ordinary working-aged Finnish people [66], except for the individuals with BMI over 40 were excluded from the present study. Thus, the findings of this study are generalizable to ordinary Finnish employees. Nonetheless, it is a weakness of the present study that as it was a real-life/data-mining type study, we did not have detailed individual information about the participants, including the information about the profession or socioeconomic status of the participants. Additionally, the fact that the information about weight and height (needed for BMI calculation) was based on self-reports may have yielded an underestimation of BMI in the study sample [67]. Most of the participants were apparently healthy; however, the inclusion of individuals with chronic diseases and/or on medications may have had an effect on HRV. However, our large sample size should have compensated for these inclusions, leading to statistically significant results. The use of real-life data was a strength of the study; however, daytime stress is affected by many confounding factors [24], very few of which were controlled for in our analysis. Participants who had consumed alcohol on the monitoring days were excluded from the analyses of the present study. Unfortunately, we did not have information for example about participants' smoking or caffeine consumption. Clearly, outside factors appear to have affected HRV-based stress because the explanation ratios of the linear models were rather small. In summary, the large real-life study sample of the present study can be considered as either a strength or a weakness depending on the perspective. For instance, in future it would be interesting to study the association of socioeconomic status with HRV-based stress and recovery by taking into account the effect of the level of PA.

This study used novel, validated [43, 62–64] HRV-based technology to assess stress, recovery and PA in real-life. The results suggest that high PA and lower BMI are associated with a lower amount of stress on workdays independently of age and sex. Additionally, the results suggest that having a lower BMI is associated with lower intensity of stress reactions on workdays, and a higher amount and better quality of recovery during sleep. This, together with existing evidence, suggests that long-term PA, resulting in improved physical fitness, has a positive effect on recovery, even though high PA was associated with a lower amount of recovery on the following night. In summary, the present results support the beneficial effects of PA on health. However, owing to the cross-sectional study design, it is not possible to draw conclusions about the direction of the associations. Previous literature suggests that the association may be reciprocal; that is, inactivity may cause stress or stress may be a factor that leads to inactivity. Overall, most of the literature finds that the experience of stress impairs efforts to be physically active [68], even though PA is beneficial in stress management [65]. More research on the causal relations between PA and HRV-based stress and recovery is needed. Randomized controlled trials investigating the effect of increasing different types of PA and timing of PA are warranted.

Conclusions

The results provide important information about the associations of objectively measured PA and body weight with objectively measured physiological stress in Finnish employees. This information could be used in future policymaking and focused upon by employers. Although the beneficial effects of PA on health are well documented, these results may be beneficial by, for example, increasing employer willingness to invest greater resources in increasing the PA of employees.

Additional files

Additional file 1: Table S1. Detection of heart rate variability-based stress and recovery. (DOCX 21 kb)

Additional file 2: Table S2. Characteristics (mean ± SD) of the measures derived from beat-to-beat R-R interval recording, by physical activity groups. Table S3. Characteristics (mean ± SD) of the measures derived from beat-to-beat R-R interval recording, by body mass index groups. Table S4. Characteristics (mean ± SD) of the measures derived from beat-to-beat R-R interval recording, by age groups. (DOCX 31 kb)

Additional file 3: Table S5. The number of participants by age, physical activity and body mass index groups. (DOCX 21 kb)

Additional file 4: Figure S1. An example (1 of the 3 days) of the feedback from the wellbeing assessment of the participants. (PDF 262 kb)

Abbreviations

ANS, autonomic nervous system; BMI, body mass index; HF, high frequency; HR, heart rate; HRV, heart rate variability; LF, low frequency; MET, multiple of the resting metabolic rate; PA, physical activity; RMSSD, root mean square of successive R-R intervals; VO_2, oxygen uptake; VO_{2max}, maximal oxygen uptake

Acknowledgements

This study was supported by the Finnish Funding Agency for Technology and Innovation (Tekes) under Health Data Mining project, grant 40116/14. T. Föhr's work was supported by a personal grant from the Juho Vainio Foundation.

Availability of data and materials

Data is owned by Firstbeat Technologies Ltd. Researchers interested in using the data are advised to contact the corresponding author.

Authors' contributions

All authors participated in planning the study design and statistical analyses, they reviewed and edited the manuscript and approved the final manuscript. TF and JP drafted the manuscript. JP and EH carried out the analyses. TF is the guarantor of this work and accepts full responsibility for the content of the article.

Competing interests

T. Myllymäki is an employee of and H. Rusko is a stockholder in Firstbeat Technologies Ltd. They did not contribute to writing the study conclusions. No financial or other conflicts of interest are declared by the other authors.

Ethics approval and consent to participate

The data obtained from the R-R interval recordings were analyzed and anonymously stored in a database administered by the software manufacturer (Firstbeat Technologies Ltd). Firstbeat Technologies Ltd and each service provider (e.g., occupational health care unit) who conducted the recordings for employees (participants) signed an agreement providing Firstbeat Technologies Ltd the right to store the data in an anonymized form and to use it for development and research purposes with a statement that employers must inform their employees about its use. According to the agreement, Firstbeat Technologies Ltd extracted an anonymous data file from the registry for the present research purposes. The study protocol was approved by the ethics committee of Tampere University Hospital (Reference No R13160).

Author details

[1]Department of Health Sciences, University of Jyväskylä, P.O. Box 35FIN-40014 Jyväskylä, Finland. [2]Department of Signal Processing, Tampere University of Technology, P.O. Box 527FIN-33101 Tampere, Finland. [3]Department of Psychology, University of Jyväskylä, P.O. Box 35FIN-40014 Jyväskylä, Finland. [4]Finnish Institute of Occupational Health, P.O. Box 40FIN-00251 Helsinki, Finland. [5]Department of Biology of Physical Activity, University of Jyväskylä, P.O. Box 35FIN-40014 Jyväskylä, Finland.

Received: 19 April 2016 Accepted: 28 July 2016
Published online: 02 August 2016

References

1. Haskell WL, Blair SN, Hill JO. Physical activity: health outcomes and importance for public health policy. Prev Med. 2009;49(4):280–2.
2. Kujala UM. Evidence on the effects of exercise therapy in the treatment of chronic disease. Br J Sports Med. 2009;43(8):550–5.
3. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ. 2006;174(6):801–9.
4. Physical Activity Guidelines Advisory Committee. Report of the Physical Activity Guidelines Advisory Committee. 2008. http://health.gov/paguidelines/. Accessed 4 Dec 2015.
5. Fletcher GF, Ades PA, Kligfield P, Arena R, Balady GJ, Bittner VA, et al. Exercise standards for testing and training: a scientific statement from the American Heart Association. Circulation. 2013;128(8):873–934.
6. Jonsdottir IH, Rödjer L, Hadzibajramovic E, Börjesson M, Ahlborg G. A prospective study of leisure-time physical activity and mental health in Swedish health care workers and social insurance officers. Prev Med. 2010;51(5):373–7.
7. Helldán A, Helakorpi S. Health Behaviour and Health among the Finnish Adult Population, Spring 2014. 2015. http://urn.fi/URN:ISBN:978-952-302-447-2. Accessed 4 Dec 2015.
8. Vartanian LR, Shaprow JG. Effects of weight stigma on exercise motivation and behavior: a preliminary investigation among college-aged females. J Health Psychol. 2008;13(1):131–8.
9. Shook RP, Blair SN, Duperly J, Hand GA, Matsudo SM, Slavin JL. What is causing the worldwide rise in body weight? US Endocrinol. 2014;10:44–52.
10. Kivimäki M, Nyberg ST, Batty GD, Fransson EI, Heikkilä K, Alfredsson L, et al. Job strain as a risk factor for coronary heart disease: a collaborative meta-analysis of individual participant data. The Lancet. 2012;380(9852):1491–7.
11. Kivimäki M, Virtanen M, Elovainio M, Kouvonen A, Väänänen A, Vahtera J. Work stress in the etiology of coronary heart disease—a meta-analysis. Scand J Work Environ Health. 2006;32(6):431–42.
12. Kouvonen A, Kivimäki M, Cox SJ, Cox T, Vahtera J. Relationship between work stress and body mass index among 45,810 female and male employees. Psychosom Med. 2005;67(4):577–83.
13. Sonnentag S, Jelden S. Job stressors and the pursuit of sport activities: a day-level perspective. J Occup Health Psychol. 2009;14(2):165–81.
14. Kouvonen A, Vahtera J, Oksanen T, Pentti J, Väänänen AK, Heponiemi T, et al. Chronic workplace stress and insufficient physical activity: a cohort study. Occup Environ Med. 2013;70(1):3–8.
15. Asztalos M, Huybrechts I, Temme E, Van Oyen H, Vandevijvere S. Association of physical activity, waist circumference and body mass index with subjective health among Belgian adults. Eur J Public Health. 2013;24(2):205–9.
16. Nyberg ST, Heikkilä K, Fransson EI, Alfredsson L, De Bacquer D, Bjorner JB, et al. Job strain in relation to body mass index: pooled analysis of 160 000 adults from 13 cohort studies. J Intern Med. 2012;272:65–73.
17. Proper KI, Koppes LLJ, van Zwieten MHJ, Bemelmans WJE. The prevalence of chronic psychological complaints and emotional exhaustion among overweight and obese workers. Int Arch Occup Environ Health. 2012;85(5):537–45.
18. Wardle J, Chida Y, Gibson EL, Whitaker KL, Steptoe A. Stress and Adiposity: A Meta-Analysis of Longitudinal Studies. Obesity. 2011;19(4):771–8.
19. Block JP, He Y, Zaslavsky AM, Ding L, Ayanian JZ. Psychosocial stress and change in weight among US adults. Am J Epidemiol. 2009;170(2):181–92.
20. Overgaard D, Gyntelberg F, Heitmann B. Psychological workload and body weight: is there an association? A review of the literature. Occup Med. 2004;54(1):35–41.
21. Armon G, Shirom A, Berliner S, Shapira I, Melamed S. A prospective study of the association between obesity and burnout among apparently healthy men and women. J Occup Health Psychol. 2008;13(1):43–57.
22. Solovieva S, Lallukka T, Virtanen M, Viikari-Juntura E. Psychosocial factors at work, long work hours, and obesity: a systematic review. Scand J Work Environ Health. 2013;39(3):241–58.
23. Montano N, Porta A, Cogliati C, Costantino G, Tobaldini E, Casali KR, et al. Heart rate variability explored in the frequency domain: a tool to investigate the link between heart and behavior. Neurosci Biobehav Rev. 2009;33(2):71–80.
24. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology: Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Eur Heart J. 1996;93(5):1043–65.
25. Malliani A, Pagani M, Lombardi F, Cerutti S. Cardiovascular neural regulation explored in the frequency domain. Circulation. 1991;84(2):482–92.
26. Acharya UR, Joseph KP, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. Med Biol Eng Comput. 2006;44(12):1031–51.
27. Jarczok MN, Kleber ME, Koenig J, Loerbroks A, Herr RM, Hoffmann K, et al. Investigating the associations of self-rated health: heart rate variability is more strongly associated than inflammatory and other frequently used biomarkers in a cross sectional occupational sample. PloS one. 2015;10(2):e0117196. doi:10.1371/journal.pone.0117196.
28. Chouchou F, Desseilles M. Heart rate variability: a tool to explore the sleeping brain? Front Neurosci. 2014;8:402. doi:10.3389/fnins.2014.00402.
29. Rennie KL, Hemingway H, Kumari M, Brunner E, Malik M, Marmot M. 2003. Effects of moderate and vigorous physical activity on heart rate variability in a British study of civil servants. Am J Epidemiol. 2003;158(2):135–43.
30. Soares-Miranda L, Sattelmair J, Chaves P, Duncan G, Siscovick DS, Stein PK, et al. Physical activity and heart rate variability in older adults: the Cardiovascular Health Study. Circulation. 2014;129(21):2100–10.
31. Mutikainen S, Helander E, Pietila J, Korhonen I, Kujala UM. Objectively measured physical activity in Finnish employees: a cross-sectional study. BMJ Open. 2014. doi:10.1136/bmjopen-2014-005927.
32. Saalasti S. Neural networks for heart rate time series analysis [dissertation]. Jyväskylä: University of Jyväskylä; 2003. https://jyx.jyu.fi/dspace/bitstream/handle/123456789/13267/951391707X.pdf?sequence=1. Accessed 9 Dec 2015.
33. Kettunen J, Saalasti S, inventors; Firstbeat Technologies Ltd, assignee. Procedure for deriving reliable information on respiratory activity from heart period measurement. United States patent US 7,460,901 B2. 2 Dec 2008. http://worldwide.espacenet.com/publicationDetails/originalDocument?CC=US&NR=7460901B2&KC=B2&FT=D&ND=4&date=20081202&DB=EPODOC&locale=fi_FI. Accessed 9 Dec 2015.
34. Kettunen J, Saalasti S, inventors; Firstbeat Technologies Ltd, assignee. Procedure for detection of stress by segmentation and analyzing heart beat signal. United States patent US 7,330,752 B2.12 Feb 2008. http://worldwide.espacenet.com/publicationDetails/originalDocument?FT=D&date=20080212&DB=EPODOC&locale=fi_FI&CC=US&NR=7330752B2&KC=B2&ND=4. Accessed 9 Dec 2015.

35. Martinmäki K, Rusko H, Kooistra L, Kettunen J, Saalasti S. Intraindividual validation of heart rate variability indexes to measure vagal effects on hearts. Am J Physiol Heart Circ Physiol. 2006;290(2):640–7.
36. Firstbeat Technologies Ltd. Stress and recovery analysis method based on 24-hour heart rate variability. 2014. http://www.firstbeat.com/app/uploads/2015/10/Stress-and-recovery_white-paper_20145.pdf. Accessed 8 Dec 2015.
37. Firstbeat Technologies Ltd. EPOC based training effect assessment. 2012. http://www.firstbeat.com/app/uploads/2015/10/white_paper_training_effect.pdf. Accessed 8 Dec 2015.
38. Ross RM, Jackson AS. Exercise concepts, calculations, and computer applications. Carmel: Benchmark Press; 1990.
39. Jones NL. Clinical exercise testing. 3rd ed. Philadelphia: W.B. Saunders; 1988.
40. Jackson AS, Blair SN, Mahar MT, Wier LT, Ross RM, Stuteville JE. Prediction of functional aerobic capacity without exercise testing. Med Sci Sports Exerc. 1990;22(6):863–70.
41. Osborne JW. Improving your data transformations: Applying the Box-Cox transformation. Practical Assessment, Research & Evaluation. 2010;15(12):1–9.
42. Birdee GS, Byrne DW, McGown PW, Rothman RL, Rolando LA, Holmes MC, et al. Relationship between physical inactivity and health characteristics among participants in an employee-wellness program. J Occup Environ Med. 2013;55(5):514–9.
43. Teisala T, Mutikainen S, Tolvanen A, Rottensteiner M, Leskinen T, Kaprio J, et al. Associations of physical activity, fitness, and body composition with heart rate variability-based indicators of stress and recovery on workdays: a cross-sectional study. J Occup Med Toxicol. 2014;34:26–40. doi:10.1186/1745-6673-9-16.
44. Zahorska-Markiewicz B, Kuagowska E, Kucio C, Klin M. Heart rate variability in obesity. Int J Obes Relat Metab Disord. 1993;17(1):21–3.
45. Rissanen P, Franssila-Kallunki A, Rissanen A. Cardiac parasympathetic activity is increased by weight loss in healthy obese women. Obes Res. 2001;9(10):637–43.
46. Lallukka T, Sarlio-Lähteenkorva S, Kaila-Kangas L, Pitkäniemi J, Luukkonen R, Leino-Arjas P. Working conditions and weight gain: a 28-year follow-up study of industrial employees. Eur J Epidemiol. 2008;23(4):303–10.
47. McLaren L. Socioeconomic status and obesity. Epidemiol Rev. 2007;29:29–48.
48. Myllymäki T, Kyröläinen H, Savolainen K, Hokka L, Jakonen R, Juuti T, et al. Effects of vigorous late-night exercise on sleep quality and cardiac autonomic activity. J Sleep Res. 2011;20:146–53.
49. Kang D, Kim Y, Kim J, Hwang Y, Cho B, Hong T, et al. Effects of high occupational physical activity, aging, and exercise on heart rate variability among male workers. Ann Occup Environ Med. 2015;27(1):1–11.
50. Kaikkonen P, Rusko H, Martinmäki K. Post-exercise heart rate variability of endurance athletes after different high-intensity exercise interventions. Scand J Med Sci Sports. 2008;18(4):511–9.
51. Seiler S, Haugen O, Kuffel E. Autonomic recovery after exercise in trained athletes: intensity and duration effects. Med Sci Sports Exerc. 2007;39(8):1366–73.
52. Hynynen E, Vesterinen V, Nummela A, Rusko H. Effects of moderate and heavy endurance exercise on nocturnal HRV. Int J Sports Med. 2010;31:428–32.
53. Pietilä J, Helander E, Myllymäki T, Korhonen I, Jimison H, Pavel M. Exploratory analysis of associations between individual lifestyles and heart rate variability -based recovery during sleep. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2015. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7318862. Accessed 10 Jun 2016.
54. Smolander J, Ajoviita M, Juuti T, Nummela A, Rusko H. Estimating oxygen consumption from heart rate and heart rate variability without individual calibration. Clin Physiol Funct Imaging. 2011;31(4):266–71.
55. Firstbeat Technologies Ltd. VO2 estimation method based on heart rate measurement. 2005. http://www.firstbeat.com/userData/firstbeat/download/white_paper_vo2_estimation.pdf. Accessed 10 Jun 2016.
56. Kang M, Bassett DR, Barreira TV, Tudor-Locke C, Ainsworth B, Reis JP, et al. How many days are enough? A study of 365 days of pedometer monitoring. Res Q Exerc Sport. 2009;80(3):445–53.
57. Tudor-Locke C, Burkett L, Reis J, Ainsworth B, Macera C, Wilson D. How many days of pedometer monitoring predict weekly physical activity in adults? Prev Med. 2005;40(3):293–8.
58. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act. 2008. doi:10.1186/1479-5868-5-56.
59. Van Amelsvoort L, Schouten E, Maan A, Swenne C, Kok F. Occupational determinants of heart rate variability. Int Arch Occup Environ Health. 2000;73(4):255–62.
60. Sammito S, Thielmann B, Seibt R, Klussmann A, Weippert M, Böckelmann I. Guideline for the application of heart rate and heart rate variability in occupational medicine and occupational science. ASU Int. 2015. doi:10.17147/ASUI.2015-06-09-03.
61. Beauchaine TP, Thayer JF. Heart rate variability as a transdiagnostic biomarker of psychopathology. Int J Psychophysiol. 2015;98(2):338–50. doi:10.1016/j.ijpsycho.2015.08.004.
62. Rusko H, Rönkä T, Uusitalo A, Kinnunen U, Mauno S, Feldt T et al. Stress and relaxation during sleep and awake time, and their associations with free salivary cortisol after awakening. Nordic Ergonomics Society congress 2006. http://www.firstbeat.com/app/uploads/2015/10/rusko_et_al_nes_2006_congress.pdf. Accessed 9 Dec 2015.
63. Uusitalo A, Mets T, Martinmäki K, Mauno S, Kinnunen U, Rusko H. Heart rate variability related to effort at work. Appl Ergon. 2011;42(6):830–8.
64. Föhr T, Tolvanen A, Myllymaki T, Jarvela-Reijonen E, Rantala S, Korpela R, et al. Subjective stress, objective heart rate variability-based stress, and recovery on workdays among overweight and psychologically distressed individuals: a cross-sectional study. J Occup Med Toxicol. 2015;10:39. doi:10.1186/s12995-015-0081-6.
65. Föhr T, Tolvanen A, Myllymäki T, Järvela-Reijonen E, Peuhkuri K, Rantala S et al. Physical activity, heart rate variability-based stress and recovery, and subjective stress during a 9-month study period. Scand J Med Sci Sports. Advance online publication. doi: 10.1111/sms12683
66. Männistö S, Laatikainen T, Vartiainen E. Suomalaisten lihavuus ennen ja nyt. 2012. http://urn.fi/URN:ISBN:978-952-245-792-9. Accessed 29 Jun 2016.
67. Gorber SC, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. Obesity Reviews. 2007;8(4):307–26.
68. Stults-Kolehmainen MA, Sinha R. The effects of stress on physical activity and exercise. Sports Med. 2014;44(1):81–121.