

MESSUNG VON KÖRPERLICHER AKTIVITÄT IN EPIDEMIOLOGISCHEN STUDIEN
UND ASSOZIATIONEN MIT KREBSINZIDENZ



Dissertation
zur Erlangung des Doktorgrades
der Humanwissenschaften
(Dr. sc. hum.)

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Andrea Weber (geb. Hillreiner)
aus
Dachau

im Jahr
2021

MESSUNG VON KÖRPERLICHER AKTIVITÄT IN EPIDEMIOLOGISCHEN STUDIEN
UND ASSOZIATIONEN MIT KREBSINZIDENZ



Dissertation
zur Erlangung des Doktorgrades
der Humanwissenschaften
(Dr. sc. hum.)

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Andrea Weber (geb. Hillreiner)
aus
Dachau

im Jahr
2021

Dekan: Prof. Dr. Dirk Hellwig

Betreuer: Prof. Dr. Dr. Michael Leitzmann

Tag der mündlichen Prüfung: 13. Dezember 2021

Inhaltsverzeichnis

Tabellenverzeichnis	4
Abbildungsverzeichnis	5
Abkürzungsverzeichnis	7
1 Einleitung	9
2 Ziele	13
3 Messung der körperlichen Aktivität in der NAKO Gesundheitsstudie	15
3.1 Einleitung	15
3.2 Methoden.....	16
3.2.1 Studiendesign der NAKO Gesundheitsstudie.....	16
3.2.2 Qualitätssicherung in der NAKO Gesundheitsstudie	16
3.2.3 GPAQ Instrument	17
3.2.4 GPAQ Qualitätssicherung.....	18
3.2.5 cpar24 Instrument	19
3.2.6 cpar24 Qualitätssicherung.....	20
3.2.7 Akzelerometrie Instrument	22
3.2.8 Akzelerometrie Qualitätssicherung.....	23
3.3 Ergebnisse	24
3.3.1 GPAQ.....	24
3.3.2 cpar24.....	26
3.3.3 Akzelerometrie.....	28
3.4 Diskussion	32
3.5 Schlussfolgerung	34
4 Assoziationen zwischen kardiorespiratorischer Fitness und der Inzidenz von kolorektalem Karzinom in der UK Biobank	35
4.1 Einleitung	35
4.2 Methoden.....	36
4.2.1 Studiendesign und Stichprobe.....	36
4.2.2 Messung der kardiorespiratorischen Fitness	37
4.2.3 Nachbeobachtung und Verknüpfung mit Routinedaten.....	41
4.2.4 Kovariablen.....	42
4.2.5 Statistische Analysen	44

4.3	Ergebnisse	46
4.4	Diskussion	53
4.5	Schlussfolgerung	58
5	Assoziationen zwischen körperlicher Aktivität, Greifkraft, sedentärem Verhalten und der Inzidenz von malignem Melanom in der UK Biobank	60
5.1	Einleitung	60
5.2	Methoden.....	61
5.2.1	Studiendesign und Stichprobe.....	61
5.2.2	Erhebung der subjektiven körperlichen Aktivität	62
5.2.3	Messung der Greifkraft.....	63
5.2.4	Erhebung des subjektiven sedentären Verhaltens.....	64
5.2.5	Nachbeobachtung und Verknüpfung mit Routinedaten.....	65
5.2.6	Kovariablen.....	66
5.2.7	Statistische Analysen	68
5.3	Ergebnisse	70
5.3.1	Deskriptive Statistisk	70
5.3.2	Körperliche Aktivität	74
5.3.3	Greifkraft.....	74
5.3.4	Sedentäres Verhalten	79
5.4	Diskussion	79
5.5	Schlussfolgerung	82
6	Zusammenfassung	83
7	Anhang.....	85
7.1	Subjektive Instrumente zur Erfassung der körperlichen Aktivität	85
7.1.1	GPAQ Fragebogen in der NAKO	85
7.1.2	cpar24 Protokoll in der NAKO	87
7.1.3	IPAQ Fragebogen in der UK Biobank.....	89
7.1.4	Fragen zu sedentärem Verhalten in der UK Biobank	90
7.2	Zusätzliche Informationen.....	91
7.3	Zusätzliche Analysen	95
7.4	Auszüge R-Code.....	99
7.4.1	Auszüge R Code – Aufbereitung GPAQ	99
7.4.2	Auszüge R Code – Aufbereitung cpar24	100
7.4.3	Auszüge R Code – Deskriptive Plots Akzelerometrie.....	101

7.4.4	Auszüge R Code – Kardiorespiratorische Fitness und Risiko von kolorektalem Karzinom in der UK Biobank.....	104
7.4.5	Auszüge R Code – Körperliche Aktivität, Greifkraft, sedentäres Verhalten und Risiko von malignem Melanom in der UK Biobank	112
8	Literaturverzeichnis	118
9	Publikationsliste	129
10	Danksagung	131
11	Lebenslauf.....	132
12	Selbstständigkeitserklärung.....	134

Tabellenverzeichnis

Tabelle 1:	MET-Werte für die im GPAQ abgefragten Intensitäten.....	19
Tabelle 2:	Einteilung der cpar24 Aktivitäten in Intensitätskategorien nach MET-Wert und Position	22
Tabelle 3:	Charakteristika der Studienpopulation zur Basisuntersuchung nach direkter Standardisierung auf die Altersverteilung der Kohorte pro Quartile von PWC _{75%}	48
Tabelle 4:	ICD-10 Klassifizierung des malignen Melanoms und Einteilung in chronisch und sporadisch UV-exponierte Hautlokalisationen.....	66
Tabelle 5:	Charakteristika der Studienpopulation zur Basisuntersuchung nach direkter Standardisierung auf die Altersverteilung der Kohorte pro Quartile von körperlicher Aktivität.....	71
Tabelle 6:	Charakteristika der Studienpopulation zur Basisuntersuchung nach direkter Standardisierung auf die Altersverteilung der Kohorte pro Quartile von Greifkraft	72
Tabelle 7:	Charakteristika der Studienpopulation zur Basisuntersuchung nach direkter Standardisierung auf die Altersverteilung der Kohorte pro Quartile von sedentärem Verhalten	73
Tabelle S 1:	GPAQ Fragen und Antwortmöglichkeiten in der NAKO.....	85
Tabelle S 2:	IPAQ-SF Fragen und Antwortmöglichkeiten in der UK Biobank.....	89
Tabelle S 3:	Fragen und Antwortmöglichkeiten zum sedentären Verhalten in der UK Biobank.....	90
Tabelle S 4:	Gewichtungsschema für die Erstellung des UV-Sensitivitätsindex.....	93
Tabelle S 5:	Plausibilität des UV-Sensitivitätsindex.....	93
Tabelle S 6:	Liste von Berufen mit natürlicher oder künstlicher UV-Exposition.....	94
Tabelle S 7:	Complete Case Analyse: Assoziation zwischen körperlicher Aktivität, Greifkraft, sedentärem Verhalten und der Inzidenz von malignem Melanom	95

Abbildungsverzeichnis

Abbildung 1:	Minuten pro Tag in verschiedenen Kontexten (GPAQ) stratifiziert nach Alter und Geschlecht.....	25
Abbildung 2:	MET-Minuten pro Woche (GPAQ) stratifiziert nach Alter und Geschlecht.....	26
Abbildung 3:	MET-Minuten pro Tag (cpar24) stratifiziert nach Alter und Geschlecht	27
Abbildung 4:	MET-Minuten pro Tag in verschiedenen Kontexten (cpar24) stratifiziert nach Alter und Geschlecht	28
Abbildung 5:	Mindestanzahl an validen Akzelerometrie-Tragetagen	29
Abbildung 6:	Durchschnittliche Beschleunigung (Akzelerometrie) stratifiziert nach Alter und Geschlecht.....	30
Abbildung 7:	Orientierung der vertikalen Achse (Akzelerometrie) über den Tagesverlauf stratifiziert nach Alter und Geschlecht.....	31
Abbildung 8:	Flussdiagramm zum Ein- und Ausschluss der UK Biobank Teilnehmenden für die Analyse der Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom.....	37
Abbildung 9:	Ablauf der Ergometrie mit inkrementell ansteigender Belastung (schematische Darstellung).....	39
Abbildung 10:	Schätzung der PWC _{75%} während der inkrementellen Phase (schematische Darstellung).....	39
Abbildung 11:	Kardiorespiratorische Fitness stratifiziert nach Alter und Geschlecht	41
Abbildung 12:	Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom	50
Abbildung 13:	Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom bei Complete Case (A) und voller Imputation (B)	51
Abbildung 14:	Flussdiagramm zum Ein- und Ausschluss der UK Biobank Teilnehmenden für die Analyse der Assoziation zwischen körperlicher Aktivität, Greifkraft, sedentärem Verhalten und malignem Melanom	61
Abbildung 15:	MET-Minuten pro Woche (IPAQ) stratifiziert nach Alter und Geschlecht.....	63
Abbildung 16:	Greifkraft stratifiziert nach Alter und Geschlecht	64
Abbildung 17:	Sedentäres Verhalten stratifiziert nach Alter und Geschlecht	65

Abbildung 18:	Assoziation zwischen körperlicher Aktivität und malignem Melanom an verschiedenen Lokalisationen (A) und stratifiziert nach UV-Sensitivität (B)	76
Abbildung 19:	Assoziation zwischen Greifkraft und malignem Melanom an verschiedenen Lokalisationen (A) und stratifiziert nach UV-Sensitivität (B)	77
Abbildung 20:	Assoziation zwischen sedentärem Verhalten und malignem Melanom an verschiedenen Lokalisationen (A) und stratifiziert nach UV-Sensitivität (B)	78
Abbildung S 1:	Screenshot cpar24: Auswahl von Aktivitäten, Spezifikation von Körperposition und Angabe von Start- und Endzeit.....	87
Abbildung S 2:	Screenshot cpar24: Spezifikation von Intensität.....	88
Abbildung S 3:	Screenshot cpar24: Übersicht aller eingetragenen Aktivitäten und Markierung von fehlenden Zeiträumen	88
Abbildung S 4:	Gerichteter Azyklischer Graph (DAG) zur Repräsentation von kausalen und verzerrenden Pfaden zwischen kardiorespiratorischer Fitness und kolorektalem Karzinom	91
Abbildung S 5:	Gerichteter Azyklischer Graph (DAG) zur Repräsentation von kausalen und verzerrenden Pfaden zwischen körperlicher Aktivität, Greifkraft, sedentärem Verhalten und malignem Melanom	92
Abbildung S 6:	Nichtlineare Zusammenhänge zwischen körperlicher Aktivität (C), sedentärem Verhalten (A, B, D, E) und malignem Melanom.....	98

Abkürzungsverzeichnis

a.n.g.	anderweitig nicht genannt
APIA	Application Programming Interface Accelerometry
cpar24	computer-based 24-hour physical activity recall (computer-basiertes 24-Stunden Erinnerungsprotokoll zur Erfassung der körperlichen Aktivität)
DAG	Directed Acyclic Graph (Gerichteter azyklischer Graph)
DNA	Deoxyribonucleic acid (Desoxyribonukleinsäure)
Df100k	Data Freeze 100.000 (Datenbasis zur Halbzeit der NAKO-Basiserhebung)
EKG	Elektrokardiogramm
ENMO	Euclidian Norm minus One (Euklidische Norm Minus Eins und negativen Werten auf 0 gesetzt)
GPAQ	Global Physical Activity Questionnaire (Globaler Fragebogen zur Erhebung der körperlichen Aktivität)
HR	Hazard Ratio
ICD	International Classification of Disease (Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme)
IPAQ-SF	International Physical Activity Questionnaire Short Form (Kurzform des Internationalen Fragebogens zur Erfassung der körperlichen Aktivität)
KA	Körperliche Aktivität
KI	Konfidenzintervall
MAD	Mean Amplitude Deviation (Mittlere Amplitudenabweichung)
MET	Metabolic equivalent of task (Metabolisches Äquivalent)
NA	not available (Fehlender Wert)
NAKO	NAKO Gesundheitsstudie
NSAID	non-steroidal anti-inflammatory drug (nichtsteroidales Antiphlogistikum)
PWC_{75%}	Physical Work Capacity at 75 % of maximum Heart Rate (Körperliche Leistungsfähigkeit bei 75 % der maximalen Herzfrequenz)
Q	Quartile
QS	Qualitätssicherung
QUAP	Questionnaire on Annual Physical Activity (Fragebogen zur jährlichen körperlichen Aktivität)
RR	Relatives Risiko

SOC	Standard Occupational Classification (Standardisierte Berufsklassifikation)
SOP	Standard Operating Procedure (Standardisierte Arbeitsanweisung)
SV	Sedentäres Verhalten
SZ	Studienzentrum
UV	Ultraviolettstrahlung
$\dot{V}O_{2max}$	Maximale Sauerstoffaufnahme
WHO	World Health Organisation (Weltgesundheitsorganisation)

1 Einleitung

Die epidemiologische Erforschung der Effekte körperlicher Aktivität auf die menschliche Gesundheit begann bereits in der Mitte des 20. Jahrhunderts, als verschiedene Berufsgruppen hinsichtlich ihres Risikos für koronare Herzerkrankungen verglichen wurden. Die Beobachtung, dass Personen in körperlich aktiven Berufen ein geringeres Mortalitätsrisiko hatten als Personen mit Berufen ohne körperliche Aktivität, lieferte einen ersten Hinweis auf die protektive Wirkung der körperlichen Aktivität (1). Seitdem gewann dieser Forschungszweig stetig an Interesse und es zeigten sich weitere inverse Assoziationen zwischen körperlicher Aktivität und beispielsweise Hypertonie, Schlaganfall, Diabetes mellitus Typ 2, bestimmten Krebsarten und Depression (2). Gerade in der letzten Dekade wurden viele neue wissenschaftliche Erkenntnisse generiert, sodass sowohl die Weltgesundheitsorganisation (WHO), als auch die amerikanische Regierung die bestehenden Empfehlungen zur körperlichen Aktivität aktualisierten (3, 4). Die vorangeschrittene epidemiologische Forschung erlaubte erstmals Aussagen zu spezifischen Subpopulationen (beispielsweise Personen, welche mit chronischen Erkrankungen leben), weiteren Krebsentitäten, sowie den Effekten des sedentären Verhaltens auf Krankheitsendpunkte (3-5). Fortschritte bei der gerätebasierten Messung der körperlichen Aktivität zeigten zudem, dass bereits geringste Mengen an körperlicher Aktivität protektive gesundheitliche Effekte erzielen können (3, 4, 6). Dennoch gibt es weiterhin Wissenslücken, beispielsweise in Bezug auf die Dosis-Wirkungsbeziehung, die Effekte leichter körperlicher Aktivität oder verschiedener Typen und Domänen körperlicher Aktivität und sedentären Verhaltens (7). Nach aktuellem Wissensstand werden bei Erwachsenen regelmäßig mindestens 150 - 300 Minuten moderate oder 75 - 150 Minuten intensive Ausdaueraktivitäten pro Woche, oder eine äquivalente Kombination, sowie zwei Mal wöchentlich muskelstärkende Aktivitäten, welche alle großen Muskelgruppen beanspruchen, empfohlen. Gleichzeitig soll die sedentäre Zeit verringert und bestenfalls mit körperlicher Aktivität beliebiger Intensität ersetzt werden (3, 4).

Die Ableitung und Aktualisierung dieser Empfehlungen, sowie die Überprüfung deren Einhaltung setzen eine möglichst valide und reliable Messung der körperlichen Aktivität und des sedentären Verhaltens voraus (4). Auch eine korrekte Definition und Abgrenzung der unterschiedlichen Konstrukte ist essenziell, um Studien vergleichen zu können. Die körperliche Aktivität ist definiert als „durch die Skelettmuskulatur erzeugte körperliche Bewegung, welche in Energieverbrauch resultiert“ (8). Sie lässt sich durch die Komponenten Häufigkeit, Dauer,

Intensität und Typ (z.B. Ausdauer- oder muskelstärkende Aktivitäten) beschreiben und kann in verschiedenen Domänen (z.B. Arbeit, Fortbewegung, Haushalt und Freizeit) stattfinden (4, 9). Das sedentäre Verhalten wird als Verhalten im Wachzustand, welches einen niedrigen Energieverbrauch von $\leq 1,5$ metabolischen Äquivalenten (MET) bei sitzender, zurückgelehnter oder liegender Körperposition aufweist, definiert (10). Die meisten bisherigen Studien zu den Assoziationen zwischen körperlicher Aktivität, sedentärem Verhalten und Krankheitsendpunkten basieren auf subjektiv gemessenen Expositionen, da diese Methodik leicht anzuwenden und kostengünstig ist (11, 12). Viele Fragebögen, Protokolle oder Tagebücher mit unterschiedlichem Detailgrad und Abfragezeitraum wurden bereits für spezifische Populationen und Fragestellungen entwickelt und eingesetzt (13). Um Vergleichbarkeit zwischen verschiedenen Populationen und Studien zu gewährleisten wurden zudem internationale Fragebögen entwickelt (z.B. Global Physical Activity Questionnaire (GPAQ), International Physical Activity Questionnaire (IPAQ) (14, 15)). Allen subjektiven Methoden ist jedoch gemein, dass sie von Erinnerungsvermögen, Einschätzung und Aufrichtigkeit der Teilnehmenden abhängig sind, was sie anfällig für Informationsbias macht (11, 16). Deshalb wurden in den letzten Jahrzehnten auch vermehrt gerätebasierte Messmethoden wie Pedometer (Schrittzähler) oder Akzelerometer (Beschleunigungssensoren) eingesetzt (17). Während diese in den 80ern noch eine Seltenheit waren und aufgrund von Kosten und Speicherkapazitäten nur im kleineren Umfang eingesetzt werden konnten, haben technologische Fortschritte mittlerweile den Einsatz in großen epidemiologischen Studien erlaubt (17). Allerdings liefern Akzelerometer keine Kontextinformationen und können bislang nur begrenzt bestimmte Bewegungsarten unterscheiden oder gar erfassen (13). Eine Kombination aus subjektiven und gerätebasierten Methoden wird deshalb für eine möglichst umfassende und valide Messung der körperlichen Aktivität empfohlen (7). Die körperliche Fitness ist definiert als „eine Reihe von Attributen, welche Personen besitzen oder erlangen und welche nötig sind, um körperlich aktiv zu sein“ (8). Sie besteht unter anderem aus den Komponenten kardiorespiratorische Fitness und Muskelkraft (8), welche in epidemiologischen Studien meist objektiv (beispielsweise mittels Ergometer-Fitnesstest oder Messung der Handgreifkraft) bestimmt werden. Körperliche Aktivität und körperliche Fitness sind positiv miteinander assoziiert (4), jedoch stellen sie unterschiedliche Konstrukte dar (18). So wird die körperliche Fitness nicht allein durch Umwelt und Lebensstil determiniert, sondern auch zu einem großen Anteil durch die Genetik (19, 20). Während sowohl körperliche Aktivität als auch körperliche Fitness inverse Assoziationen mit Krankheitsendpunkten zeigten (21), sind die

Interaktionen und unabhängigen Effekte der beiden Konstrukte noch nicht abschließend geklärt (4).

Auch die biologischen Mechanismen, über welche körperliche Aktivität und Fitness protektiv auf die Kanzerogenese wirken sind nicht endgültig erforscht, jedoch werden verschiedene Hypothesen postuliert. Neben direkten Effekten auf die Tumorzellen, wie beispielsweise Einleitung des programmierten Zelltodes oder Hemmung der Vermehrung von Tumorzellen wird eine generelle Steigerung der Immunfunktion und eine Verbesserung der Insulinsensitivität diskutiert (22-25). Auch eine Verminderung von oxidativem Stress und chronischen Entzündungsvorgängen, sowie eine Verbesserung der Desoxyribonukleinsäure-(DNA)-Reparaturmechanismen könnten den antikanzerogenen Effekt von körperlicher Aktivität und Fitness erklären (22, 26-30). Mittlerweile gilt der protektive Einfluss der körperlichen Aktivität auf die Inzidenz von Brust-, Kolon-, Blasen-, Endometrium-, Ösophagus Adeno-, Magen- und Nierenkarzinom als erwiesen (4). Die Risikoreduktionen lagen bei 10 bis 20 % (4). Kardiorespiratorische Fitness zeigte in einigen Studien eine inverse Assoziation mit der Inzidenz und Mortalität von Gesamtkrebs (31-33), sowie einigen Krebsentitäten (z.B. Lungen- oder Kolorektalkarzinom) (34-38), wobei die meisten Studien nur in Männern durchgeführt wurden. Auch zwischen Greifkraft und Kolorektal-, Lungen- und Brustkarzinom zeigten sich erste inverse Assoziationen (39). Im Gegensatz hierzu wird diskutiert, ob sedentäres Verhalten über eine Steigerung der systemischen Entzündung und Insulinresistenz, sowie über Östrogen-vermittelte DNA-Schäden das Krebsrisiko erhöht (40). In den letzten Jahren wurde ein positiver Zusammenhang zwischen sedentärem Verhalten und dem Risiko für Kolon-, Lungen- und Brustkarzinom identifiziert (40). Trotz dieser vielfältigen Erkenntnisse zu den Effekten von körperlicher Aktivität, Fitness und sedentärem Verhalten auf das Krebsrisiko gibt es weiterhin großen Forschungsbedarf (4). Unter anderem wird empfohlen, Assoziationen zwischen körperlicher Aktivität und bisher wenig erforschten Krebsentitäten zu untersuchen. Da es sich hierbei oft um seltenere Krebstypen handelt, sind groß angelegte prospektive Kohortenstudien, gepoolte Analysen oder Metaanalysen vonnöten. Auch werden Informationen zu den Effekten spezifischer Typen von körperlicher Aktivität, beispielsweise muskelstärkenden Aktivitäten, sowie objektiv gemessener körperlicher Aktivität benötigt (4).

Obwohl die positiven Effekte von körperlicher Aktivität und Fitness weithin bekannt sind, lag die Prävalenz von unzureichender körperlicher Aktivität weltweit bei 27,5 % (41). In Deutschland erreichten 55 % der Erwachsenen nicht die empfohlene Menge an Ausdaueraktivitäten in der Freizeit (42). Dennoch schätzte eine kürzlich veröffentlichte Studie,

dass bereits jetzt jährlich mindestens 3,9 Millionen vorzeitige Tode durch einen körperlich aktiven Lebensstil verhindert werden (43). Durch Förderung von Sport und Bewegung könnte diese Zahl noch weiter gesteigert werden. Insbesondere Krebs wird zu einem großen Anteil durch Umwelt- oder Lebensstilfaktoren, wie beispielsweise körperliche Aktivität, verursacht (44, 45). Gleichzeitig steigen Inzidenz und Mortalität von Krebs weltweit an. Für das Jahr 2018 wurden weltweit 18,1 Millionen neue Krebsfälle und 9,6 Millionen Krebstode geschätzt (44). Die WHO beschloss im Jahr 2018 einen globalen Aktionsplan mit dem Ziel, die körperliche Inaktivität bis 2030 um 15 % zu reduzieren (46). Auch sollen die aktuellen WHO-Empfehlungen zur körperlichen Aktivität nach spätestens 10 Jahren überarbeitet werden, was anschaulich verdeutlicht, dass die Forschung zu den positiven Effekten der körperlichen Aktivität keinesfalls abgeschlossen ist (3).

2 Ziele

Die vorliegende Arbeit soll einen Beitrag zu dieser Forschung liefern und vorhandene Wissenslücken füllen.

Wie bereits beschrieben, empfiehlt die WHO für zukünftige Forschung die subjektive und objektive Messung der körperlichen Aktivität in groß angelegten prospektiven Beobachtungsstudien, um die Effekte sowie Dosis-Wirkungsbeziehungen verschiedener Typen und Domänen von körperlicher Aktivität und sedentärem Verhalten auf die menschliche Gesundheit zu untersuchen (7). Anhand der prospektiven Kohortenstudie NAKO werden verschiedene subjektive und gerätebasierte Instrumente vorgestellt, sowie die Methodik und Ergebnisse der Qualitätssicherungen dargelegt. Die Aufbereitung und Ableitung neuer Variablen soll den parallelen Einsatz subjektiver und gerätebasierter Daten zur körperlichen Aktivität in Assoziationsstudien gewährleisten. Insbesondere werden behandelt:

- Der subjektive Fragebogen GPAQ zur körperlichen Aktivität einer typischen Woche
- Das subjektive computerbasierte 24-Stunden Protokoll zur Zeitverwendung cpar24
- Die Messung der körperlichen Aktivität mittels triaxialer 7-Tage Akzelerometrie

Für viele Krebsarten liegt noch nicht ausreichend Evidenz vor, um einen Einfluss von körperlicher Aktivität als gesichert anzunehmen. Die Assoziationen zwischen körperlicher Fitness und Krebsinzidenz wurden bislang größtenteils nur in Männern untersucht. Auch die Interaktionen zwischen körperlicher Aktivität und körperlicher Fitness in Hinblick auf die Kanzerogenese sind nicht geklärt. Daher untersucht die vorliegende Arbeit folgende Fragestellungen mit Daten der prospektiven Kohortenstudie UK Biobank:

- Assoziation zwischen kardiorespiratorischer Fitness und Inzidenz von kolorektalem Karzinom, Kolonkarzinom, Rektumkarzinom sowie Kolonkarzinom an proximalen und distalen Lokalisationen

Diese Ergebnisse wurden bereits im European Journal of Epidemiology veröffentlicht (siehe S. 129).

- Assoziation zwischen körperlicher Aktivität, Greifkraft, sedentärem Verhalten und der Inzidenz von malignem Melanom, sowie malignem Melanom an Körperlokalisationen mit chronischer oder sporadischer Exposition durch Ultraviolettstrahlung (UV).

Diese Arbeit wurde im British Journal of Cancer veröffentlicht (siehe S. 129).

Aufgrund der bislang zu kurzen Follow-up Zeit der UK Biobank und der hieraus resultierenden geringen Teststärke konnten keine Assoziationen mit Akzelerometrie-Daten als Exposition oder Krebsmortalität als Endpunkt untersucht werden.

3 Messung der körperlichen Aktivität in der NAKO Gesundheitsstudie

3.1 Einleitung

Epidemiologische Studien werden unter anderem durchgeführt, um Risiko- und Schutzfaktoren für Erkrankungen, sowie die Häufigkeit und Verbreitung dieser zu erforschen (47). Voraussetzung für korrekte Schlussfolgerungen ist die zuverlässige und valide Erhebung von Daten. Die Messung der körperlichen Aktivität nimmt einen wichtigen Stellenwert ein, da sie als Exposition, Endpunkt als auch Kovariable für viele Fragestellungen benötigt wird, um beispielsweise unverzerrte Assoziationen zwischen Lebensstilfaktoren und Krankheitsendpunkten zu identifizieren. Auch um die Level an körperlicher Aktivität in verschiedenen Populationen zu dokumentieren und über die Zeit zu vergleichen oder Interventionsprogramme zu evaluieren werden praktikable Messmethoden benötigt (13). Wie bereits erwähnt ist die körperliche Aktivität jedoch ein sehr komplexes Konstrukt, was die Messung im Alltagskontext von Studienteilnehmenden entsprechend herausfordernd gestaltet (48, 49). Während subjektive Methoden wie Fragebögen zwar kostengünstig, schnell und einfach anzuwenden sind, treten hierbei häufig Misklassifikationen durch Informationsbias (beispielsweise mangelndes Erinnerungsvermögen oder Antwortverzerrung in Richtung erwünschtes Verhalten) auf (49). Messungen mit Beschleunigungssensoren sind zwar weniger anfällig für Informationsbias, allerdings ist die Anwendung teurer und aufwendiger und manche Aktivitäten wie isometrische Muskelkontraktionen, das Tragen von Lasten oder bergauf gehen können nicht gut erfasst werden (13). Außerdem gibt es bislang keinen Konsens bezüglich Aufbereitung und Auswertung der Daten und die Aufzeichnung der Beschleunigung liefert keine Kontextinformationen (50, 51). Vergleiche zwischen subjektiven und objektiven Methoden zeigten, dass die Instrumente unterschiedliche Aspekte erfassten (48). Daher wird empfohlen, eine Kombination aus Selbstangaben und gerätebasierten Messmethoden in epidemiologischen Studien anzuwenden, um so auch Kontextinformationen zu körperlicher Aktivität und sedentärem Verhalten zu erhalten (7, 48). In der NAKO entschied man sich für ein multimodales Konzept bestehend aus 3 subjektiven Instrumenten mit unterschiedlichem Abfragezeitraum und Detailgrad (QUAP, GPAQ, cpar24) sowie der Messung mittels triaxialem Akzelerometer (52).

3.2 Methoden

3.2.1 Studiendesign der NAKO Gesundheitsstudie

Die NAKO Gesundheitsstudie ist die größte bevölkerungsbezogene Kohortenstudie Deutschlands (53). Sie soll eine Datenbasis zur Identifizierung von Risiko- und Schutzfaktoren sowie Prädiktoren und Biomarkern für häufige chronische Erkrankungen wie Diabetes, Krebserkrankungen oder kardiovaskulären Erkrankungen liefern (53, 54). Zwischen 2014 und 2019 wurden über 200.000 Männer und Frauen (jeweils 50 %) im Alter von 20 bis 69 Jahren alters- und geschlechtsstratifiziert aus Zufallsstichproben der 18 deutschlandweiten Studienregionen rekrutiert und untersucht (53). Alle Teilnehmenden wurden vorab ausführlich informiert und aufgeklärt (z.B. über die Freiwilligkeit, Ablauf, Datenspeicherung, Datenschutz und Möglichkeit des Widerrufs) und gaben ihre schriftliche Einwilligung zur Teilnahme an allen oder bestimmten Untersuchungsmodulen (53). Alle Dokumente wurden vorab durch die lokalen Ethikkommissionen geprüft und das Datenschutzkonzept und der Ethikkodex wurden zudem studienbegleitend kontrolliert (53, 55). Für alle Teilnehmenden war ein circa 4 stündiges Untersuchungsprogramm (Level 1) mit standardisierten Interviews, Touchscreen-Befragungen, körperlichen Untersuchungen (z.B. Anthropometrie, Greifkraftmessung, Blutdruckmessung, Akzelerometrie, Spirometrie) und Bioprobengewinnung vorgesehen. Bei den subjektiven Erhebungsinstrumenten wurden beispielsweise ausführliche Informationen zu Soziodemographie, Lebensstilfaktoren, bereits bestehenden Erkrankungen oder der Familienhistorie von Erkrankungen erhoben. Bei einer zufällig ausgewählten Teilstichprobe von 20 % wurde das erweiterte Level 2 Programm mit zusätzlichen Untersuchungen (beispielsweise Ruhe-Elektrokardiogramm (EKG), Langzeit-EKG mit Erfassung der Atmungsfunktion während des Schlafes, Ultraschallmessung des Abdominalfettes oder Echokardiographie) angeboten (53). In einigen Studienzentren wurde außerdem eine submaximale Fahrradergometrie oder eine Ganzkörper Magnetresonanztomographie-Untersuchung durchgeführt (53). Die longitudinale Beobachtung wird durch regelmäßige postalische Befragungen zu inzidenten Erkrankungen, sowie durch eine erneute Untersuchung aller Teilnehmenden zwischen 2018 und 2023 gewährleistet (53). Langfristig soll zusätzlich eine Verknüpfung mit Sekundär- und Registerdaten (z.B. Krebsregister, Krankenversicherung, Sozialversicherung) realisiert werden (53).

3.2.2 Qualitätssicherung in der NAKO Gesundheitsstudie

In der NAKO Gesundheitsstudie gelten sehr hohe Qualitätsstandards, welche sich an den Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis, aufgestellt

von der Deutschen Gesellschaft für Epidemiologie, orientieren (56). Vor Beginn der Basiserhebung wurden Machbarkeitsstudien durchgeführt, um das gesamte Untersuchungsprogramm zu testen (53, 54). Studienbegleitend wurde die Qualität durch interne Organe, sowie durch das Robert Koch Institut als unabhängige externe Einrichtung überprüft (53). Für alle Untersuchungsmodule wurden standardisierte Arbeitsanweisungen (SOPs) von den Modulverantwortlichen erstellt und deren Einhaltung wurde durch regelmäßige Schulungen, Supervisionen, Site-Visits und Zertifizierungen gewährleistet (53). Die Untersucher erhielten bei der Dateneingabe durch Plausibilitätsgrenzen und Warnmeldungen bei unvollständig abgeschlossenen Modulen ein direktes Feedback, um fehlende oder unplausible Werte direkt zu vermeiden. Zudem wurde die Datenqualität für jedes Untersuchungsmodul von den verantwortlichen Wissenschaftlern überprüft. Um hiermit nicht bis zum Abschluss der Basiserhebung warten zu müssen wurde bereits zur Halbzeit der Erstuntersuchung (Erhebungszeitraum 14.3.2014 - 17.3.2017) ein sogenannter „Data freeze“ durchgeführt und ein Datensatz mit 101.839 Teilnehmenden (df100k) zur Verfügung gestellt (53). Die Ergebnisse der Qualitätssicherung (QS) wurden in Berichten festgehalten, welche unter anderem die Vollständigkeit der Untersuchungen, Verteilung der Werte, unplausible Werte, Definition abgeleiteter Werte oder Maßnahmenvorschläge zur Qualitätssicherung beinhalteten (57). Die Arbeitsgruppe Datenkontrolle überprüfte diese Berichte und forderte gegebenenfalls Ergänzungen oder Revisionen an. Erst nach Abschluss dieser Qualitätskontrolle wurden die entsprechenden Variablen in die Studiendatenbank überführt und konnten folglich für Publikationsvorhaben beantragt werden. Im Folgenden wird die Qualitätssicherung der Module GPAQ (gesamte Erstuntersuchung) und cpar24 (df100k) sowie der Akzelerometrie Rohdaten (df100k) vorgestellt.

3.2.3 GPAQ Instrument

Der GPAQ (Global Physical Activity Questionnaire) wurde von der WHO zur Erfassung und Überwachung der körperlichen Aktivität entwickelt (58). Ziel war es ein Instrument zu schaffen, welches für verschiedene Länder und Kulturen adaptiert werden kann aber dennoch vergleichbare Ergebnisse liefert (59). In der Originalversion war der GPAQ als persönliches Interview konzipiert und Beispielbilder für die jeweils abgefragten Aktivitäten sollten passend zur Population gezeigt werden (60). Der GPAQ wurde bislang in über 100 Ländern sowohl als Fragebogen zum Selbstausfüllen, als auch als Interview durchgeführt (58). Beide Versionen zeigten hierbei im Vergleich zur Akzelerometrie gute Reliabilität (61, 62) und Validität (61, 63). Generell waren die Selbstausfüllerfragebögen vergleichbar mit den Interviewbasierten

Erhebungen und auch der Einsatz der Beispielbilder hatte keinen Einfluss auf die Validität der Ergebnisse (60, 64, 65). Der GPAQ erfragt die körperliche Aktivität einer typischen Woche anhand von 15 Fragen zu Intensität (moderat, intensiv), Häufigkeit (Tage pro Woche) und Dauer (Minuten pro Tag) in den Domänen Arbeit, Fortbewegung und Freizeit. Eine weitere Frage erhebt Informationen zum sedentären Verhalten (Minuten pro Tag) (58, 66). In der NAKO Gesundheitsstudie war der GPAQ als Selbstausfüllerfragebogen im Touchscreen für alle Teilnehmenden vorgesehen (Tabelle S 1). Falls die Touchscreen Befragung aufgrund technischer Probleme nicht funktionierte, wurde der GPAQ mit der Notfall-Papierversion abgefragt, um den Informationsverlust möglichst gering zu halten.

3.2.4 GPAQ Qualitätssicherung

Im Rahmen der Qualitätssicherung wurden die mit dem GPAQ erhobenen Daten auf Vollständigkeit geprüft und fehlende Werte sowie Werteverteilungen pro Studienzentrum, Altersgruppe und Geschlecht dargestellt. Die Analysevorschriften der WHO wurden verwendet, um die Daten auf Plausibilität zu prüfen sowie neue Variablen abzuleiten (66). Sobald für Teilnehmende in einer der Domänen (moderate bzw. intensive körperliche Aktivität in der Arbeit, bei der Fortbewegung oder in der Freizeit) invalide Angaben vorlagen, wurden diese von allen Analysen ausgeschlossen. Als invalide Antwort galten Zeitangaben von weniger als 10 Minuten pro Tag beziehungsweise mehr als 960 Minuten pro Tag (66). Begründet wurde dies durch die dem GPAQ zugrundeliegenden Empfehlungen für körperliche Aktivität, welche eine Mindestdauer von 10 Minuten vorgeben (67), sowie die Annahme, dass mehr als 16 Stunden körperliche Aktivität in einer der Domänen unrealistisch erscheinen. Zudem wurden Teilnehmende ausgeschlossen, wenn in allen Domänen fehlende Werte vorlagen (66).

Eine Vielzahl an neuen Variablen wurde basierend auf den GPAQ Analysevorschriften generiert. Die Zeit pro Woche in den einzelnen Subdomänen wurde durch Multiplikation von Häufigkeit (Anzahl Tage pro Woche) und Dauer (Minuten pro Tag) berechnet. MET-Minuten pro Woche wurden durch Multiplikation der Zeit pro Woche mit dem entsprechenden MET-Wert (Tabelle 1) berechnet. Die WHO begründet die Auswahl dieser MET-Werte durch den Kalorienverbrauch, welcher im Vergleich zu ruhigem Sitzen (1 MET entspricht $1 \text{ kcal} \cdot \text{kg}^{-1} \cdot \text{Stunde}^{-1}$) bei moderater Aktivität um das vierfache und bei intensiver Aktivität um das achtfache erhöht ist (66).

Tabelle 1: MET-Werte für die im GPAQ abgefragten Intensitäten

Domäne	Subdomäne	MET-Wert
Arbeit	Moderate körperliche Aktivität in der Arbeit	4,0
	Intensive körperliche Aktivität in der Arbeit	8,0
Fortbewegung	Moderate körperliche Aktivität bei der Fortbewegung (Fahrradfahren und Gehen)	4,0
Freizeit	Moderate körperliche Aktivität in der Freizeit	4,0
	Intensive körperliche Aktivität in der Freizeit	8,0

Zusätzlich wurden kategoriale Variablen erstellt, welche angeben ob ein Teilnehmender die WHO Empfehlungen zur körperlichen Aktivität erreicht (≥ 600 MET-Minuten pro Woche) (67) oder ob körperliche Aktivität in den einzelnen Domänen berichtet wurde. Außerdem wurden Variablen erstellt, welche den Anteil arbeitsbezogener, fortbewegungsbezogener oder freizeitbezogener körperlicher Aktivität an der gesamten körperlichen Aktivität angeben. Für das sedentäre Verhalten (Minuten pro Tag) wurden alle Werte von 0 bis 1440 als plausibel angesehen (66). Auszüge aus dem R Code zur Qualitätssicherung und Ableitung neuer GPAQ Variablen befinden sich in Abschnitt 7.4.1.

3.2.5 cpar24 Instrument

Der cpar24 (computer-based 24-hour physical activity recall) ist ein webbasiertes Instrument zur Erhebung aller Aktivitäten des vergangenen Tages (68). Das Erinnerungsprotokoll wurde 2011 am Institut für Epidemiologie und Präventivmedizin der Universität Regensburg durch Simone Kohler entwickelt und von der externen Softwarefirma „Webstollen“ programmiert. Ziel war die Gewinnung detaillierter und kontextbezogener Informationen zu körperlicher Aktivität und sedentärem Verhalten, welche in der NAKO zudem zur Plausibilisierung von Extremwerten in den Akzelerometrie-Daten oder von Schlafenszeiten verwendet werden können.

Im Rahmen einer NAKO Pilotstudie wurde der cpar24 bei 67 Teilnehmenden gegen die Akzelerometrie (Actigraph GT3x+) validiert (68). Hierbei unterschätzte der cpar24 leichte körperliche Aktivität um 123 Minuten und überschätzte moderate bis intensive körperliche Aktivität um 89 Minuten. Das sedentäre Verhalten war bei beiden Erhebungsmethoden vergleichbar (Unterschied von 7 Minuten). Insgesamt zeigte sich eine mäßige Übereinstimmung zwischen cpar24 und Akzelerometrie bei sedentärem Verhalten ($r = 0,54$), leichter körperlicher Aktivität ($r = 0,46$) und moderater bis intensiver körperlicher Aktivität ($r = 0,50$) (68). Eine wiederholte Anwendung des cpar24 nach circa 3 Stunden zeigte moderate

bis starke Reliabilität für sedentäres Verhalten ($r = 0,75$), leichte körperliche Aktivität ($r = 0,65$) und moderate bis intensive körperliche Aktivität ($r = 0,92$). Die in den verschiedenen Intensitäten verbrachten Zeiten unterschieden sich nicht statistisch signifikant zwischen den beiden cpar24 Befragungen (68).

In der NAKO Basisuntersuchung wurde der cpar24 mindestens allen Teilnehmenden angeboten, die auch ein Akzelerometer oder eine Somnowatch (Gerät zur Erfassung von Schlafcharakteristiken mit zusätzlichem Langzeit-EKG) trugen. In vier Studienzentren (Essen, Leipzig, Neubrandenburg und Regensburg) wurde der cpar24 allen Teilnehmenden angeboten. Bei Einwilligung erhielten die Teilnehmenden Zugangsdaten und meldeten sich hiermit zwei Tage nach dem Studienzentrumsbesuch von zu Hause aus auf der Internetseite www.cpar24.de an. Hier erhielten sie eine Anleitung in schriftlicher Form und als Video. Wie bei einem interaktiven Kalender konnten die Teilnehmenden aus 262 Aktivitäten wählen um den Ablauf des vorangegangenen Tages (von Mitternacht bis Mitternacht) zu rekonstruieren (siehe Abbildung S 1) (68). Diesen Aktivitäten waren MET-Werte aus dem Kompendium von Ainsworth (Version 2011) (69) zugeordnet. Zur besseren Übersichtlichkeit waren die Aktivitäten in 13 Oberkategorien unterteilt (Schlafen, Tätigkeiten im Liegen / Körperpflege, Umkleiden / Ernährung / Hausarbeit / Gartenarbeit, Heimwerken / Einkaufen, Termine / Beruf, Arbeit / Familienleben, Soziale Kontakte / Freizeitaktivitäten, Hobby / Sport / Gehen, Fortbewegung, Reisen / Außer-Haus Aktivitäten / Sonstige Tätigkeiten) (68). Falls die gewünschte Aktivität nicht in der Liste aufgeführt wurde oder nicht gefunden werden konnte, war eine Freitexteingabe über das Hilfe-Menü möglich. Für 12 Aktivitäten wurde zusätzlich die Körperposition (im Sitzen verbrachter Anteil) abgefragt (Abbildung S 1). Für 22 weitere Aktivitäten konnte die Intensität (leicht, mittel und anstrengend) spezifiziert werden (siehe Abbildung S 2). Zu jedem Aktivitätseintrag wurde die Start- und Endzeit in 5-Minuten-Schritten abgefragt. Das Eintragen von zwei parallelen oder sich überschneidenden Aktivitäten war, wie in der Literatur empfohlen, möglich (Abbildung S 3) (70). Am Ende der Befragung erhielten die Teilnehmenden Feedback zu ihrer erbrachten körperlichen Aktivität. Zudem hatten die Teilnehmenden die Möglichkeit die Benutzerfreundlichkeit des cpar24 zu bewerten, einen Kommentar zu hinterlassen und die Zeit, welche sie zum Ausfüllen benötigt hatten, anzugeben (68).

3.2.6 cpar24 Qualitätssicherung

Da es sich bei dem cpar24 um ein neues, eigens für die NAKO entwickeltes Instrument ohne etablierte Aufbereitungsvorschrift handelt, mussten zunächst Problemquellen und

Korrekturbedarf identifiziert werden, um die Datenintegrität vor der Herausgabe zu Analyse Zwecken zu garantieren. Einige Teile der Qualitätssicherung (z.B. die Aufbereitung der sich überschneidenden Aktivitäten sowie die semiautomatisierte Zuordnung von Freitexteingaben) wurden im Rahmen einer Masterarbeit von Maximilian Ott (Studiengang Informatik der Ostbayerische Technische Hochschule Regensburg) unter Anleitung und externer Betreuung von Andrea Weber entwickelt.

Vollständigkeit (relativ zum Tragestatus von Akzelerometer und Somnowatch), Aktualität bei multiplen Zugängen und SOP konformes Ausfülldatum (innerhalb des Akzelerometer-Tragezeitraums) wurden überprüft. Die in der Datenbank hinterlegten MET-Werte wurden, wenn nötig, aktualisiert und an die angegebene Intensität (MET-Wert jeweils für leicht, mittel und anstrengend) und Körperposition (prozentualer MET-Wert für Sitzen und Stehen) angepasst. Da im cpar24 nur die zwei anstrengendsten gleichzeitig durchgeführten Aktivitäten dokumentiert werden sollten, wurden alle weiteren parallel eingetragenen Aktivitäten ignoriert. Damit die Gesamtdauer trotz gleichzeitiger Aktivitäten 24 Stunden pro Tag nicht überstieg, wurden überlappende Zeiträume zu je 50% auf die beiden parallelen Aktivitäten aufgeteilt. Den in der Freitexteingabe hinzugefügten Aktivitäten mussten passende Aktivitäten aus der cpar24 Liste zugeordnet werden, um einerseits Datenschutz zu gewährleisten und andererseits passende MET-Werte für die Analysen zu erhalten. Das von Maximilian Ott entwickelte Verfahren zur semiautomatisierten Zuordnung der Freitexteingaben bediente sich verschiedener String-matching Methoden aus dem R Paket stringdist (z.B. edit-basiert, q-gram basiert oder heuristisch) (71, 72). Je nach Übereinstimmungsgrad zwischen den Strings wurde die Aktivität automatisch aus dem Kompendium zugeordnet oder musste manuell aus einer Reihe vorgeschlagener Aktivitäten ausgewählt werden. Alle zugeordneten Aktivitäten wurden zudem manuell auf Übereinstimmung bezüglich Kontext und Intensität überprüft. Basierend auf den MET-Werten wurden alle Aktivitäten in Intensitätskategorien eingeteilt (Tabelle 2). Sedentäres Verhalten wurde wie empfohlen zusätzlich durch die Körperposition bestimmt, wodurch beispielsweise „Gießen (dabei gehen und stehen)“ als leichte körperliche Aktivität klassifiziert wurde (73).

Da im cpar24 weiche Plausibilitätskontrollen für die Dauer einiger Aktivitäten hinterlegt waren, wurden die Zeitangaben als valide akzeptiert. Um Vergleichbarkeit zwischen allen Teilnehmenden zu gewährleisten, wurden alle Einträge mit unvollständigen Angaben zum Tagesablauf (< 1440 Minuten) markiert. Als weitere, von der Arbeitsgruppe Datenkontrolle vorgeschlagene Maßnahme, wurden alle cpar24 Datensätze mit weniger als vier Aktivitätseinträgen als unplausibel angesehen.

Tabelle 2: Einteilung der cpar24 Aktivitäten in Intensitätskategorien nach MET-Wert und Position

Intensitätskategorie	Definition
Intensiv	$\geq 6,0$ MET
Moderat	$\geq 3,0$ MET und $< 6,0$ MET
Leicht	$> 1,5$ MET und $< 3,0$ MET
Sedentär	$\leq 1,5$ MET und wach und sitzend, liegend
Schlafen	Aktivitäten Schlafen; Dösen, Schlummern

Die beschriebenen Datenaufbereitungen (siehe auch Ausschnitte R Code in Abschnitt 7.4.2) führten zur Ableitung neuer Variablen (z.B. Aktivitätsdauer nach Korrektur für Überschneidung, MET-Minuten). Abschließend wurden in einem Teildatensatz Benutzerbewertung, angegebene Ausfülldauer und Kommentare der Teilnehmenden ausgewertet.

3.2.7 Akzelerometrie Instrument

In der NAKO wurden zudem triaxiale Akzelerometer eingesetzt, welche die Beschleunigung ($\text{mg} \approx 9,81 \cdot 10^{-3} \cdot \text{m} \cdot \text{s}^{-2}$) in allen drei Raumachsen (vertikal, medio-lateral, und antero-posterior) über 7 Tage aufzeichneten. Diese Informationen können zur Ableitung des Ausmaßes der körperlichen Aktivität herangezogen werden. In der frühen Phase der NAKO wurden die Akzelerometer wGT3x+ der Firma ActiGraph (Pensacola, FL, USA) eingesetzt und, nachdem diese nicht mehr produziert wurden, die Akzelerometer wGT3x-BT (52). Die Akzelerometriemessung war bei mindestens 50 % aller Level 1 Teilnehmenden (80.000 Teilnehmende in der Basisuntersuchung) vorgesehen. Im Studienzentrums wurden die Akzelerometer vollständig geladen und standardisiert nach der NAKO SOP initialisiert. Hierfür wurde das Programm ActiLife, sowie das am Deutschen Institut für Ernährung entwickelte Programm APIA (application programming interface accelerometry) verwendet (74). Die Messfrequenz wurde auf 100 Hz eingestellt und die ActiGraph spezifischen Filteroptionen zur Datenreduktion (Schlafmodus, Energiesparmodus) wurden deaktiviert, um die Anzahl an herstellerabhängigen Vorverarbeitungsschritten zu minimieren, und damit die Transparenz und Reproduzierbarkeit unserer Aufbereitung zu maximieren. Die Teilnehmenden erhielten ausführliche Informationen zur Untersuchung und wurden instruiert das Gerät durchgehend, also auch während des Schlafens, für eine Woche bis zum 8. Tag nach Besuch im Studienzentrums zu tragen. Der Sensor sollte nur in der Sauna, beim Tauchen und bei

Wasserkontakt, welcher länger als 30 Minuten andauert abgenommen werden. Das Gerät wurde durch die Untersuchenden oberhalb der rechten Hüfte auf der mittleren Axillarlinie mit einem Gurt befestigt und konnte oberhalb und unterhalb der Kleidung getragen werden. Die Teilnehmenden sollten während der Trageweche möglichst alle Aktivitäten wie gewohnt durchführen. Für das Zurücksenden der Geräte erhielten die Teilnehmenden einen vorfrankierten Umschlag. Nach Eingang der Akzelerometer im Studienzentrum wurden die Daten heruntergeladen und die Geräte für den nächsten Einsatz vorbereitet. Für die Rückmeldung an die Teilnehmenden im Ergebnisbrief wurden mittels ActiLife Variablen abgeleitet (Zeit in moderater und intensiver körperlicher Aktivität sowie Anzahl an Schritten).

3.2.8 Akzelerometrie Qualitätssicherung

Die Qualitätssicherung der Ergebnisbrief-Variablen, sowie die Überprüfung von Vollständigkeit und Optimierungsbedarf wurde durch die Modulverantwortliche Dr. Sylvia Gastell bei 41.216 Akzelerometrie-Datensätzen durchgeführt und ist daher nicht Bestandteil dieser Arbeit (52). Im Folgenden wird die Qualitätssicherung der Akzelerometrie-Rohdaten beschrieben, welche am Institut für Epidemiologie und Präventivmedizin der Universität Regensburg unter der Leitung von Prof. Dr. Dr. Leitzmann und in Zusammenarbeit mit Dr. Hansjörg Baurecht und Dr. Vincent van Hees, einem unabhängigen Berater und Experten zur Aufbereitung von Akzelerometrie-Rohdaten, durchgeführt wurde.

Die Akzelerometrie-Rohdaten lagen als ActiGraph-spezifisches Datenformat .gt3x vor und hatten eine Datengröße von ca. 10 TB. Da die Konvertierung von .gt3x zu .csv, sowie die Aufbereitung der rohen Akzelerometrie-Daten sehr zeit- und rechenintensiv war, wurde dies auf dem Hochleistungsrechner Athene der Universität Regensburg durchgeführt (ca. 24 Stunden pro 15.360 Dateien bei paralleler Verarbeitung). Für die Aufbereitung wurde das von Vincent van Hees entwickelte Open Source R Paket GGIR Version 2.3-0 verwendet, wobei die entsprechenden Parameter an das NAKO Studienprotokoll angepasst wurden (75).

Mittels Autokalibrierung wurden Korrekturfaktoren abgeleitet, um die Ruhemessungen aller 3 Achsen auf 1 g (Gravitationsbeschleunigung) einzustellen, und so etwaige Kalibrierungsfehler bei der Herstellung der Geräte zu beheben (76). Nichttragezeiten pro 15 Minuten Intervall wurden anhand von Standardabweichung ($< 13,0$ mg) und Wertebereich ($< 50,0$ mg) des umschließenden 60 Minuten Intervalls identifiziert (77). Pro 5 Sekunden Epoche wurden die Beschleunigungsmessgrößen ENMO (Euklidische Norm Minus 1 und negativen Werten gleich 0 gesetzt (77)) und MAD (Epochen-Mittelwert von euklidischer Norm pro Zeitschritt abzüglich

des Mittelwerts, genannt Mittlere Amplitudenabweichung (78)) generiert. Die vertikale Achse wurde als Achse mit höchster verzögerter 24-Stunden Autokorrelation identifiziert. Die Orientierung der vertikalen Achse, relativ zur horizontalen Ebene wurde mit in GGIR implementierten Standardformeln berechnet. Als Nichttragezeit klassifizierte oder signalbegrenzte („signal clipping“) Zeitabschnitte wurden imputiert, indem sie durch den entsprechenden Durchschnittswert dieses Zeitpunktes aus den anderen Messtagen ersetzt wurden (75). Die moderate bis intensive körperliche Aktivität wurde geschätzt, indem die Zeit oberhalb eines aus der Literatur entnommenen hüftspezifischen ENMO-Schwellenwertes von 70 mg aggregiert wurde (79). Die Zeit in moderater bis intensiver körperlicher Aktivität wurde sowohl ohne, als auch mit Berücksichtigung der 10 Minuten Mindestdauer (wobei mindestens 80 % der 5-Sekunden-Intervalle die Intensitätskriterien erfüllen müssen) geschätzt. Variablen zur Beschreibung der Beschleunigung beziehungsweise des Ausmaßes der körperlichen Aktivität wurden pro Teilnehmendem, pro Tag und pro 15-Minuten Intervall jedes Tages aggregiert (75).

In Anlehnung an frühere Akzelerometrie-Untersuchungen wurde die minimal benötigte Anzahl an validen Tragetagen bestimmt, indem in einem vollständigen Teildatensatz der Mittelwert zufällig gezogener Kombinationen der ersten 1-7 Tage gegen den Mittelwert der vollständigen 7 Tage mittels Intraklassen-Korrelationskoeffizient verglichen wurde (80, 81). Beschreibende Statistiken und Abbildungen zur Überprüfung der Datenplausibilität wurden mit der Statistiksoftware R erstellt (siehe Auszüge im Abschnitt 7.4.3).

3.3 Ergebnisse

3.3.1 GPAQ

Im GPAQ Qualitätssicherungsdatensatz der Basisuntersuchung befanden sich 205.264 Teilnehmende. Bei 196.327 (96 % aller Teilnehmenden) war der GPAQ als vollständig durchgeführt dokumentiert worden. Bei den einzelnen Fragen zeigten sich Fehlraten von 3,9 % bis 4,4 %.

Der Anteil von Teilnehmenden, welche körperliche Aktivität in der Arbeit angaben, betrug 21 % bei intensiver körperlicher Aktivität beziehungsweise 43 % bei moderater körperlicher Aktivität und war in der ältesten Gruppe am höchsten. Männer gaben häufiger intensive körperliche Aktivität in der Arbeit an als Frauen (25 % versus 18 %). Insgesamt gaben 72 % der Teilnehmenden körperliche Aktivität bei der Fortbewegung an. In der Domäne Freizeit gaben 51 % der Teilnehmenden intensive körperliche Aktivität und 74 % der Teilnehmenden moderate körperliche Aktivität an. Die Zeit pro Tag in sedentärem Verhalten nahm mit

zunehmendem Alter stetig ab (von 474 Minuten (SD = 256 Minuten) pro Tag bei den unter 30-Jährigen bis 325 Minuten (SD = 200 Minuten) pro Tag bei den über 69-Jährigen). Frauen berichteten mit 389 Minuten (SD = 234 Minuten) pro Tag weniger sedentäres Verhalten als Männer mit 435 Minuten (SD = 257 Minuten) pro Tag.

Bei der Aufbereitung nach dem GPAQ Analyse Guide zeigte sich, dass 20.789 Teilnehmende (10 %) invalide Angaben bei den Fragen zur körperlichen Aktivität getätigt hatten und somit von den Analysen ausgeschlossen werden mussten. Die Frage zu sedentärem Verhalten hatten 9075 Teilnehmende (4 %) invalide beantwortet und mussten demnach von den Analysen ausgeschlossen werden. Bei der Stichprobe der 184.457 Teilnehmenden mit validen Einträgen zur körperlichen Aktivität und validen Angaben zu Alter und Geschlecht waren die Teilnehmenden im Durchschnitt 49 Minuten (SD = 80) pro Tag in der Freizeit, 56 Minuten (SD = 100) pro Tag bei der Fortbewegung und 108 Minuten (SD = 188) pro Tag während der Arbeit körperlich aktiv. Die in den Domänen verbrachte Zeit stieg größtenteils mit zunehmendem Alter an (Abbildung 1). Pro Tag verbrachten die Teilnehmenden im Durchschnitt 212 Minuten (SD = 251 Minuten) mit moderater bis intensiver körperlicher Aktivität. Im Durchschnitt akkumulierten die Teilnehmenden 7427 MET-Minuten pro Woche (SD = 9383 MET-Minuten, Abbildung 2) und erfüllten so zu einem Großteil (88 %) die WHO Empfehlungen zu körperlicher Aktivität (≥ 600 MET-Minuten pro Woche).

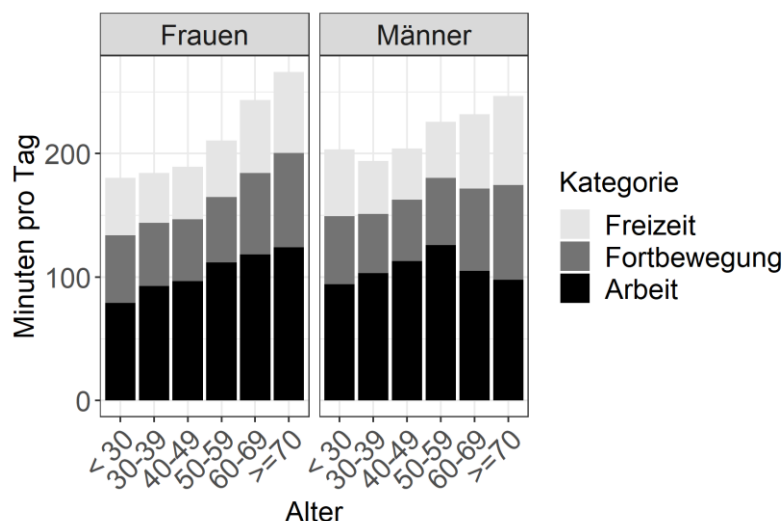


Abbildung 1: Minuten pro Tag in verschiedenen Kontexten (GPAQ) stratifiziert nach Alter und Geschlecht

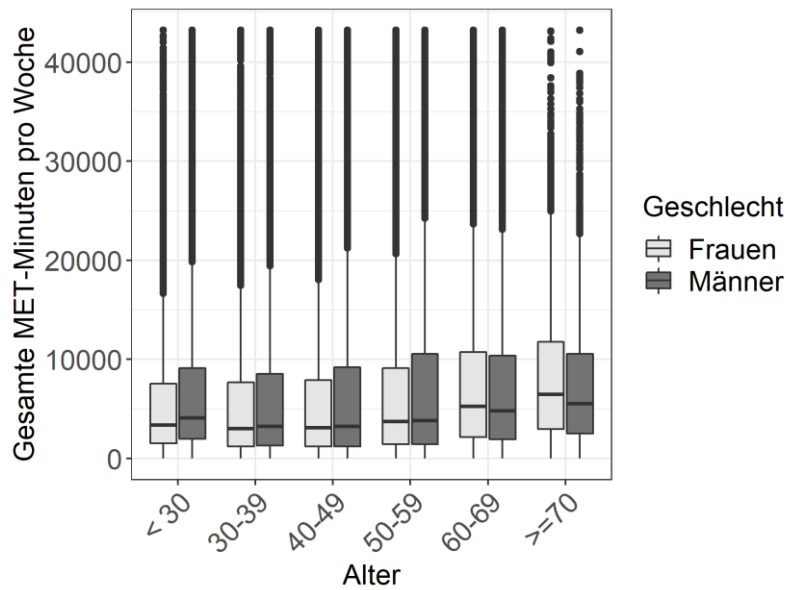


Abbildung 2: MET-Minuten pro Woche (GPAQ) stratifiziert nach Alter und Geschlecht

Zur besseren Darstellbarkeit wurden die Daten an der 99. Perzentile trunziert.

3.3.2 cpar24

Der zur Qualitätsprüfung ausgegebene Datensatz enthielt 41.421 cpar24 Zugänge. Gemessen an der Zielpopulation der Akzelerometer- oder Somnowatchtragenden ($n = 55.081$), erhielten 56 % einen cpar24 Zugang. Zusätzlich erhielten 9755 Teilnehmende ohne Akzelerometer oder Somnowatch einen cpar24 Zugang. Nach Ausschluss von nicht eindeutig zuordenbaren Zugängen ($n = 270$), nicht ausgefüllten cpar24 Zugängen ($n = 12.243$), und veralteten cpar24 Zugängen ($n = 183$) verblieben 28.745 cpar24 Zugänge. Über 90 % der Zugänge wurden innerhalb des Akzelerometer-Tragezeitraums initialisiert, die meisten wie vorgeschrieben am 2. Tag nach dem Studienzentrumsbesuch. Die Teilnehmenden beschrieben ihren Tag im Durchschnitt mit 23 Aktivitäten. Nach Korrektur der sich überlappenden Einträge (genutzt von 4362 Teilnehmenden) dauerte eine Aktivität im Durchschnitt 60,7 Minuten ($SD = 98,8$ Minuten). Über 90 % der Teilnehmenden, welche den cpar24 begonnen hatten, vervollständigten die 1440 Minuten und nur 28 dieser Teilnehmenden berichteten weniger als 4 Aktivitätseinträge. Im Qualitätssicherungs-Datensatz lagen 1371 Freitexteingaben (1220 einzigartige) von 300 verschiedenen Teilnehmenden vor. Frauen und ältere Teilnehmende nutzten die Möglichkeit der Freitexteingabe häufiger als Männer und jüngere Teilnehmende. Von den 1220 Freitexteingaben konnten 516 automatisch zugeordnet werden (davon 468 korrekt). Zusätzlich konnten 169 Aktivitäten aus den Vorschlägen übernommen werden. Die

restlichen 583 Freitexteinträge wurden manuell zugeordnet. Die meisten Freitexteinträge behandelten Aktivitäten, die auch in der Aktivitätsliste vorhanden gewesen wären.

Pro Tag verbrachten die Teilnehmenden im Durchschnitt 179 Minuten (SD = 156 Minuten) mit moderater bis intensiver körperlicher Aktivität. Im Durchschnitt gaben die Teilnehmenden 2529 MET-Minuten pro Tag (SD = 515 MET-Minuten pro Tag) an, wobei sich wenig Unterschiede zwischen den Altersgruppen, Männern und Frauen zeigten (siehe Abbildung 3). Bei der Darstellung nach unterschiedlichen Kategorien zeigten sich allerdings deutliche Unterschiede, beispielsweise in der Domäne „Beruf, Arbeit“ (siehe Abbildung 4).

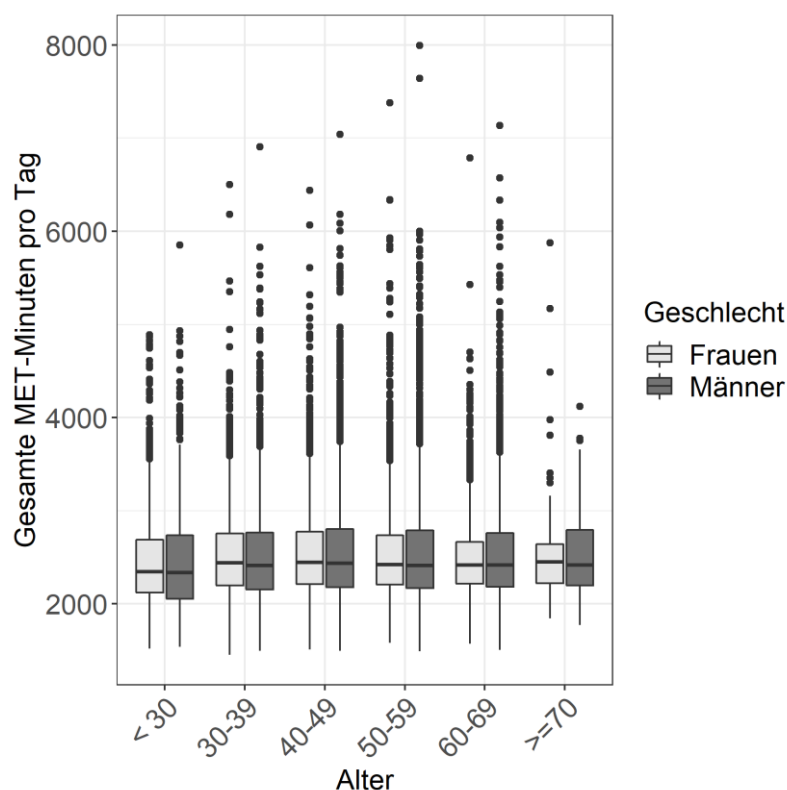


Abbildung 3: MET-Minuten pro Tag (cpar24) stratifiziert nach Alter und Geschlecht

Von 14.056 Teilnehmenden lagen größtenteils positive Benutzerbewertungen vor. Die Kommentare bezogen sich vor allem auf den Wunsch nach einer prospektiven Erhebungsmethode, technischen Problemen beim Ausfüllen oder Bemängelung des zu großen Zeitaufwands. Im Durchschnitt gaben die Teilnehmenden an 27 Minuten für das Ausfüllen des cpar24 benötigt zu haben.

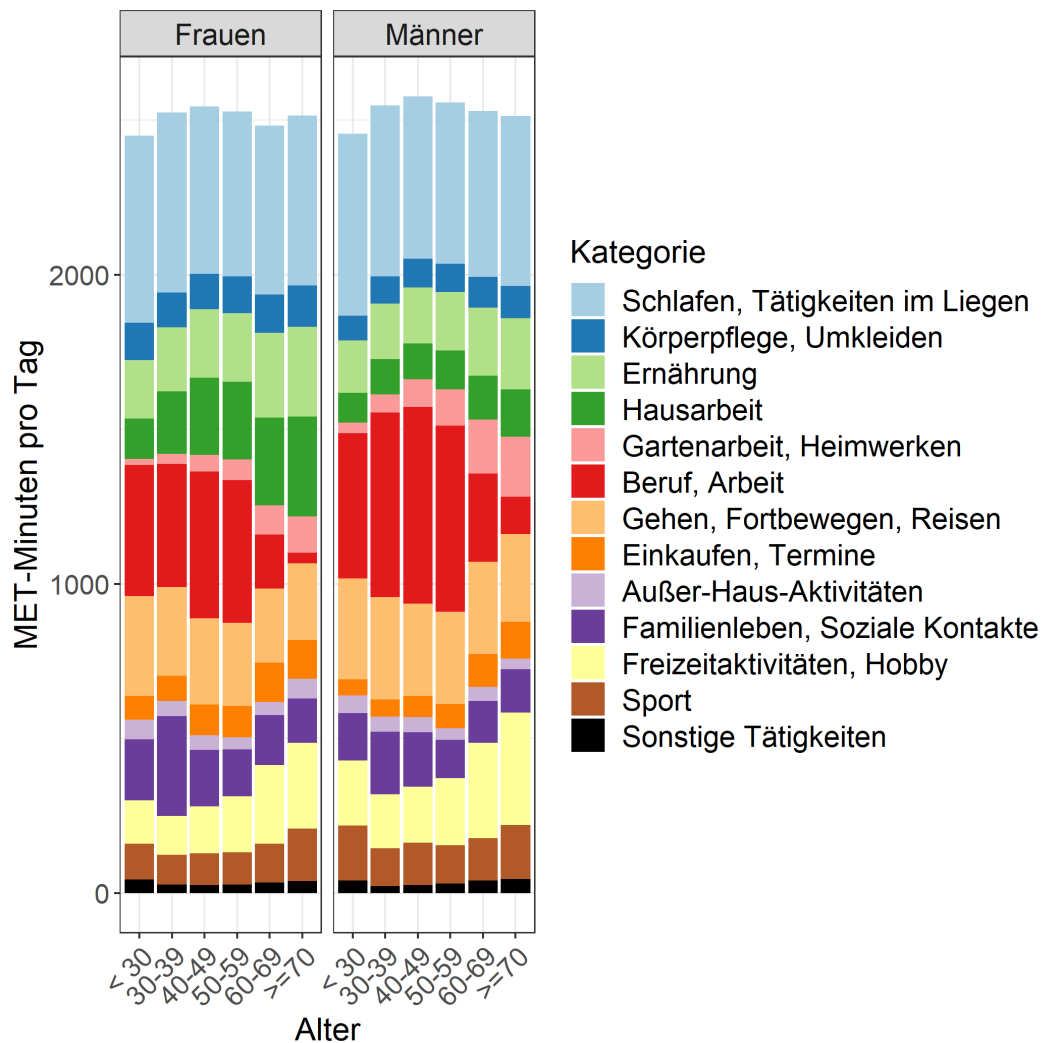


Abbildung 4: MET-Minuten pro Tag in verschiedenen Kontexten (cpar24) stratifiziert nach Alter und Geschlecht

3.3.3 Akzelerometrie

Die folgenden Auswertungen wurden in einem Teildatensatz mit 36.474 Teilnehmenden durchgeführt, welche innerhalb des vorgesehenen Zeitraums das Akzelerometer getragen hatten, sowie einen Kalibrierungsfehler $\leq 0,01$ aufwiesen. Hiervon hatten 27.013 (74 %) der Teilnehmenden an allen 7 Tragetagen mindestens 16 valide Stunden pro Tag generiert. Der Vergleich mittels Intraklassen-Korrelationskoeffizienten ergab, dass mindestens drei Tragetage nötig waren, um weniger als 10 % Abweichung vom tatsächlichen 7-Tage Mittelwert (ENMO) zu erreichen (Abbildung 5).

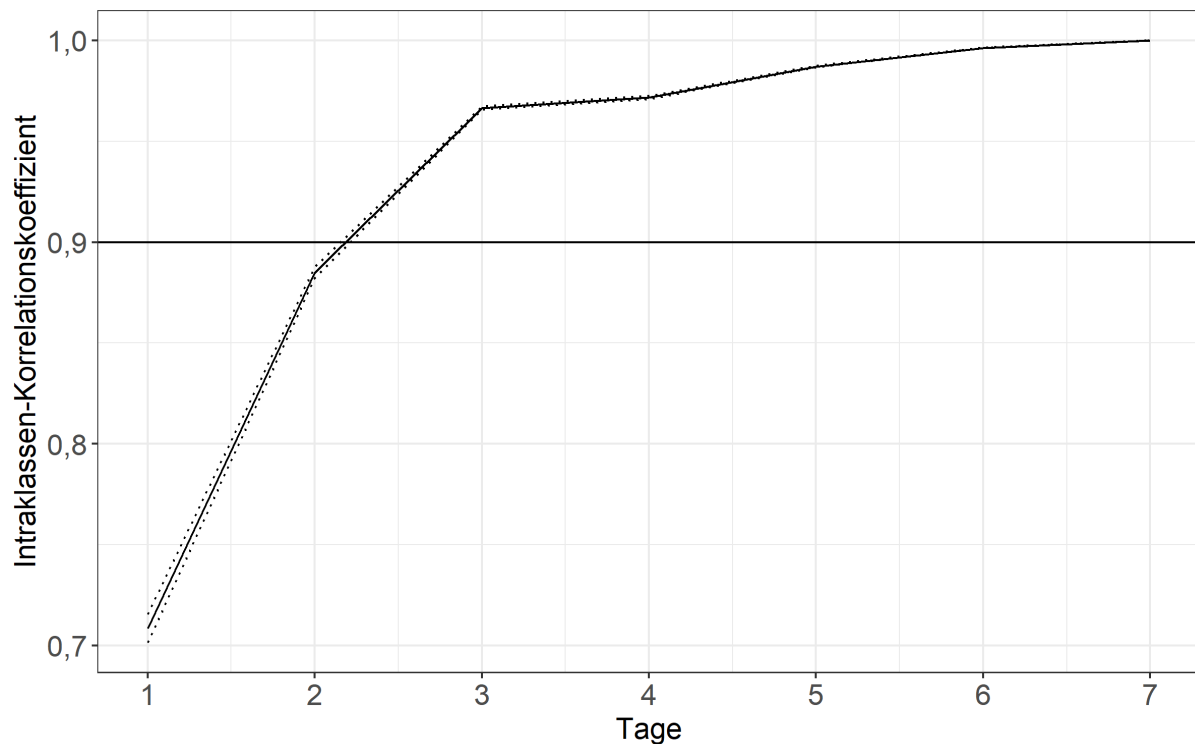


Abbildung 5: Mindestanzahl an validen Akzelerometrie-Tragetagen

Für die weiteren Analysen wurden deshalb nur 33.102 Teilnehmende verwendet, welche mindestens drei valide Tragetage (mindestens zwei Wochentage und einen Wochenendtag) aufwiesen.

Die körperlichen Aktivitätslevel lagen im Durchschnitt bei 11,5 mg (SD = 5,3 mg; ENMO) beziehungsweise 19,6 mg (SD = 6,1 mg; MAD). Die durchschnittliche Beschleunigung (ENMO) war bei den Männern aller Altersklassen etwas höher als bei den Frauen und nahm ab der Altersgruppe der 50-59-Jährigen stetig ab (Abbildung 6). Für die mittlere Amplitudenabweichung ergab sich ein ähnliches Bild (Ergebnisse nicht gezeigt).

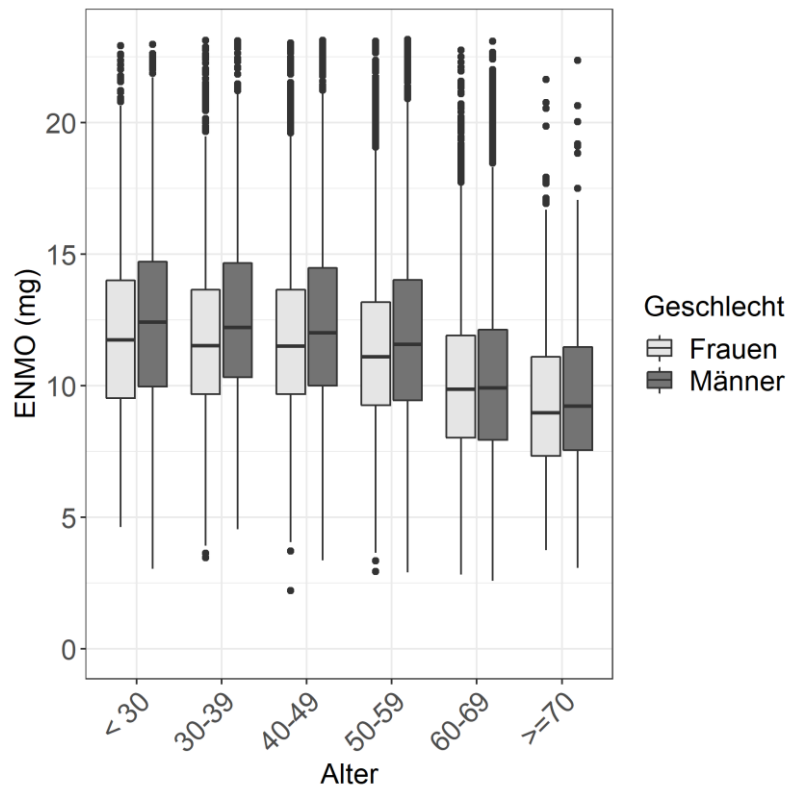


Abbildung 6: Durchschnittliche Beschleunigung (Akzelerometrie) stratifiziert nach Alter und Geschlecht

Zur besseren Darstellbarkeit wurden die Daten an der 99. Perzentile trunziert.

Bei Berücksichtigung der 10 Minuten Mindestdauer verbrachten die Teilnehmenden pro Tag im Durchschnitt 13,3 Minuten (SD = 17,9 Minuten) mit moderater bis intensiver körperlicher Aktivität. Ohne Berücksichtigung der Mindestdauer verbrachten die Teilnehmenden pro Tag im Durchschnitt 56,2 Minuten (SD = 27,2 Minuten) mit moderater bis intensiver körperlicher Aktivität.

Abbildung 7 zeigt die Achsenorientierung der vertikalen Achse über den Tagesverlauf, stratifiziert nach Altersgruppe und Geschlecht. Der durchschnittliche Winkel der vertikalen Achse war in den Nachtstunden deutlich niedriger als in den Tagesstunden.

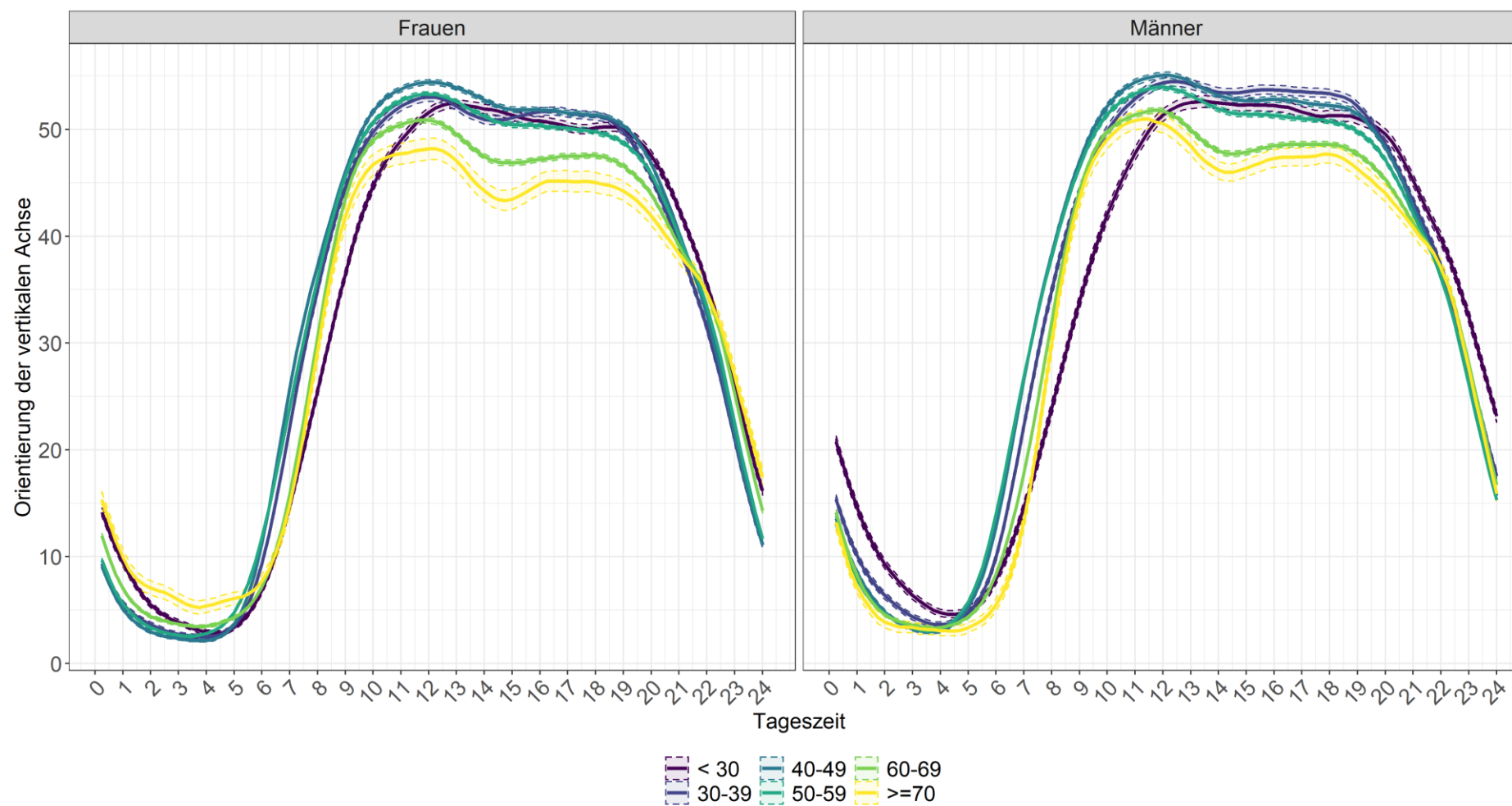


Abbildung 7: Orientierung der vertikalen Achse (Akzelerometrie) über den Tagesverlauf stratifiziert nach Alter und Geschlecht

Die gestrichelten Linien repräsentieren zwei Standardfehler.

3.4 Diskussion

Die Qualitätssicherung der NAKO Module war ein essenzieller Schritt um Vollständigkeit, Plausibilität und Datenintegrität vor der Datenherausgabe und -Analyse zu gewährleisten. Durch das Aufbereiten und Ableiten neuer Variablen wurde sichergestellt, dass die aus der NAKO hervorgehenden Publikationen valide und vergleichbare Daten verwenden.

Der GPAQ zeichnete sich durch seine einfache und zeitsparende Anwendung aus, vor allem durch seine Implementierung als Touchscreen Fragebogen zum Selbstaussfüllen. Da er allen Teilnehmenden angeboten wurde und auch in der Papierversion des Touchscreens verfügbar war, war die Vollständigkeit dieses Moduls sehr hoch. Allerdings wurde im Rahmen der Qualitätssicherung eine stark ausgeprägte Überschätzung der körperlichen Aktivität festgestellt, besonders ersichtlich am überproportional hohen Anteil an Teilnehmenden (88 %), welche die WHO Empfehlungen einhielten (67). Im Gegensatz hierzu ergab die Messung der Level an körperlicher Aktivität mittels Akzelerometrie in einer Follow-up Erhebung der KORA Studie (Kooperative Gesundheitsforschung in der Region Augsburg), dass lediglich 14 % der Teilnehmenden die WHO Empfehlungen zur körperlichen Aktivität erreichten (82). Es ist denkbar, dass in der NAKO Studie ein Selektionsbias vorliegt, bei welchem vor allem aktive und sehr gesundheitsbewusste Personen an einer Studie zum Thema Gesundheit teilnehmen. Allerdings zeigte sich bei einigen GPAQ Domänen, dass Teilnehmende der älteren Altersgruppen mehr körperliche Aktivität sowie weniger sedentäres Verhalten angaben, als die Teilnehmenden der jüngeren Altersgruppen. Während subjektive Erhebungsmethoden hierzu bereits in der Vergangenheit heterogene Ergebnisse lieferten (42, 83), zeigten die ersten Ergebnisse der Akzelerometrie-Messung in der NAKO, dass die Zeit pro Tag in moderater bis intensiver körperlicher Aktivität ab 50 Jahren (bei den Frauen) beziehungsweise ab 60 Jahren (bei den Männern) abnahm (52). Dies deutet auf einen altersabhängigen Informationsbias in den subjektiven GPAQ Daten hin, welcher möglicherweise durch Interpretation der abgefragten Intensitäten (z.B. intensive körperliche Aktivität) nach dem individuellen Leistungsniveau der Teilnehmenden zustande kam. Insgesamt entsprechen diese Beobachtung den Ergebnissen vorangegangener Studien, welche zeigten, dass subjektive Erhebungsmethoden im Vergleich zu gerätebasierten Methoden aufgrund von Informationsbias zu Über- oder Unterschätzung von körperlicher Aktivität und sedentärem Verhalten führen (16, 48, 84).

Die Anwendung des cpar24 war teilweise für die Teilnehmenden zu zeitaufwendig beziehungsweise durch technische Probleme oder fehlenden Internetzugang nicht möglich, was sich in geringerer Vollständigkeit zeigte. Bei der Aufbereitung des cpar24 wurden verschiedene Probleme und Fehlfunktionen bei der Erhebung und Speicherung der cpar24 Daten identifiziert,

welche allerdings alle im Rahmen der Qualitätssicherung behoben werden konnten. Insgesamt waren die im cpar24 generierten Daten plausibel und die neu abgeleiteten Variablen lieferten vielfältige, detaillierte und für viele Fragestellungen relevante Informationen (52). Ein Vergleich mit weiteren internationalen Fragebögen zur Zeitverwendung (Act24, PAR, MARCA) zeigte gute Übereinstimmung und offenbarte wichtige Aspekte, wie die Klassifizierung des sedentären Verhaltens mittels Kombination aus MET-Wert und Körperposition, welche in die Qualitätssicherung einfließen konnten (73). Im Vergleich zu herkömmlichen Fragebögen, wie beispielsweise dem GPAQ, ist die Gefahr des Informationsbias und der Misklassifikation bei 24-Stunden Erinnerungsprotokollen zur Zeitverwendung deutlich geringer, da der Abfragezeitraum kürzer zurückliegt (85). Zudem liefert die chronologische Abfrage aller Aktivitäten des Vortages Hilfestellung für die korrekte und vollständige Erinnerung und wirkt der Angabe sozial erwünschten Verhaltens entgegen, da der Fokus nicht auf die Abfrage körperlicher Aktivität gesetzt ist (86, 87). Falls der cpar24 erneut eingesetzt werden soll, wurden im Zuge der Qualitätssicherung wichtige Punkte identifiziert, welche überarbeitet werden müssen, um eine fehlerfreie und für die Teilnehmenden möglichst angenehme Anwendung des cpar24 zu gewährleisten. Diese beinhalten beispielsweise die Überarbeitung der Webseite zur besseren Darstellbarkeit auf mobilen Endgeräten und allen gängigen Internetbrowsern. Auch sollte die Möglichkeit der Freitexteingaben entfernt werden, da dies keinen ersichtlichen Mehrwert bei gleichzeitigem erheblichem Zeit- und Ressourcenaufwand im Zuge der Qualitätssicherung mit sich bringt. Alternativ könnte die Möglichkeit angeboten werden eine fehlende Aktivität über verschiedene Kategorien zu definieren (beispielsweise Kontext, Intensität, Körperposition).

Das Ziel der Aufbereitung der Akzelerometrie-Rohdaten war es, ein Set an relevanten, detailreichen und sensorbasierten Variablen für die wissenschaftliche Forschung zur Verfügung zu stellen, da nicht alle Forschenden über die Erfahrung oder die technische Infrastruktur verfügen, die notwendig sind, um die Rohdaten zu analysieren. Im Vergleich zu den direkt von ActiLife erzeugten aggregierten Daten war die Aufbereitung der Rohdaten transparenter, flexibler und lieferte auch Informationen zu bestimmten Tagen oder Tagessegmenten, was eine detaillierte Analyse von Aktivitätsmustern erlaubt. Die vorläufige Akzelerometrie-Qualitätssicherung zeigt, dass die Vollständigkeit der Datenerhebung höher war als bei der subjektiven Erhebung mittels cpar24. Im Gegensatz zu den GPAQ Daten, zeigte die Akzelerometer-Messung zudem einen Rückgang an körperlicher Aktivität bei höheren Alterskategorien. Während die unterschiedlichen Tragelokalisationen (Hüfte versus Handgelenk) keinen direkten Vergleich zwischen Akzelerometrie-Daten von NAKO und UK

Biobank ermöglichen, ergab sich bei beiden Studien eine Mindestdauer von drei Tragetagen, um eine valide Messung zu erhalten (80). Bei an der Hüfte getragenen Akzelerometern ist auch die Achsenorientierung eine wichtige Informationsquelle, welche beispielsweise für die Klassifikation von liegenden Tätigkeiten hinzugezogen werden kann. Die ersten Ergebnisse aus der NAKO Qualitätssicherung zeigten einen deutlichen Unterschied im durchschnittlichen Winkel der vertikalen Achse zwischen Tages- und Nachtstunden. Allerdings liefert die Akzelerometrie bislang keine kontextuellen Informationen, wie Art (abgesehen von Schätzungen der im Liegen verbrachten Zeit), Zweck oder Ort der Aktivität.

3.5 Schlussfolgerung

Insgesamt konnten die aus der Literatur bekannten Stärken und Limitationen der verschiedenen Instrumente beobachtet werden. Der Fragebogen GPAQ war zwar einfach und schnell auszufüllen, allerdings gab es hier auch die stärkste Misklassifikation. Das 24-Stunden Erinnerungsprotokoll cpar24 lieferte sehr detaillierte und plausible Daten, war allerdings für einige Teilnehmende zu zeitaufwändig. Während die Akzelerometrie Messung weniger anfällig für Informationsbias ist, liefert diese Methode keine Kontextinformationen. Die unterschiedlichen Instrumente wurden im Rahmen der NAKO Qualitätssicherung aufbereitet und harmonisiert, um die verschiedenen gemessenen Aspekte der körperlichen Aktivität und des sedentären Verhaltens in zukünftigen Studien zu kombinieren.

4 Assoziationen zwischen kardiorespiratorischer Fitness und der Inzidenz von kolorektalem Karzinom in der UK Biobank

4.1 Einleitung

Das kolorektale Karzinom ist, nach Tumoren der Lunge und Brust, der dritthäufigste Krebs weltweit (44). In Industrieländern mit westlichem Lebensstil ist die Inzidenz höher, jedoch haben Erkrankte aus Ländern mit geringem Einkommen geringere Überlebenschancen, möglicherweise aufgrund schlechterer Gesundheitsversorgung (88). In der EU wurden die durch kolorektales Karzinom verursachten Kosten auf 13,1 Milliarden € in 2009 geschätzt (89). Gleichzeitig wird angenommen, dass ein Großteil der Krebsfälle durch einen gesunden Lebensstil verhindert werden könnte (90). Die Identifizierung und Förderung modifizierbarer Schutzfaktoren, wie körperlicher Aktivität oder kardiorespiratorischer Fitness, hat somit eine hohe Relevanz für die Gesundheit der gesamten Bevölkerung. Zudem sind Fitness Tests, beispielsweise durchgeführt beim Allgemeinarzt, eine einfache Möglichkeit um Hochrisikopopulationen zu identifizieren, welchen ein Darmkrebsscreening nahegelegt werden könnte. Bei einer Koloskopie können Kolorektalkrebs-Vorstufen nicht nur erkannt, sondern auch präventiv entfernt werden, was das Krebsrisiko effektiv reduziert (88, 91, 92).

In den vergangenen Jahren wurde subjektiv gemessene körperliche Aktivität bereits in vielen Studien und Metaanalysen als protektiver Faktor für das kolorektale Karzinom identifiziert (93-95). Die Erforschung der Assoziation zwischen kardiorespiratorischer Fitness und dem Risiko an kolorektalem Karzinom zu erkranken erst in den letzten Jahren in den Vordergrund gerückt (33-36, 96-98). Ein protektiver Einfluss der kardiorespiratorischen Fitness auf das Krebsrisiko ist biologisch plausibel und erhöhte Fitnesslevel wurden bereits mit einem verringertem Risiko für Gesamtkrebsmortalität in Verbindung gebracht (31). Dennoch lieferten die bisherigen Studien zur Assoziation zwischen kardiorespiratorischer Fitness und der Inzidenz von kolorektalem Karzinom heterogene Ergebnisse und wiesen größtenteils Limitationen auf (38). Im Gegensatz zur körperlichen Aktivität, welche ein Verhalten darstellt, ist die kardiorespiratorische Fitness die Basisvoraussetzung um körperlich aktiv zu sein (8). Sie wird einerseits zu 25 – 65 % genetisch determiniert (19), andererseits kann die kardiorespiratorische Fitness auch durch körperliche Aktivität modifiziert werden (99). Ein entscheidender Vorteil der kardiorespiratorischen Fitness ist die objektive und damit validere Messung, welche häufig stärkere und konsistentere Zusammenhänge mit Krankheitsendpunkten zeigte als subjektiv gemessene körperliche Aktivität (8, 21, 100).

Die im Folgenden beschriebene prospektive Analyse zur Assoziation zwischen objektiv gemessener kardiorespiratorischer Fitness und Risiko von kolorektalem Karzinom, soll insbesondere durch die Inkludierung von Studienteilnehmerinnen und der Betrachtung anatomischer Sublokalisationen bisherige Forschungslücken füllen.

4.2 Methoden

4.2.1 Studiendesign und Stichprobe

Die Fragestellung wurde mit Daten der UK Biobank bearbeitet (Antrag Nr. 24091). Die groß angelegte populationsbasierte prospektive Kohortenstudie wurde vom North West Multi-centre Research Ethics Committee (MREC) genehmigt und untersuchte zwischen 2006 und 2010 über eine halbe Million Männer und Frauen in 22 Studienzentren (SZ) in Großbritannien (17 SZ in England, drei SZ in Wales, zwei SZ in Schottland) (101). Rekrutiert wurden Teilnehmende zwischen 39 und 70 Jahren aus verschiedenen sozioökonomischen Schichten. Alle Teilnehmenden gaben ihre schriftliche Einwilligung nach einer ausführlichen Aufklärung und wurden zur Basisuntersuchung von geschultem Personal nach standardisierten Arbeitsanweisungen untersucht. Mittels Touchscreen-Fragebögen, persönlichen Interviews, physischen Messungen und Entnahme von Bioproben wurden ausführliche gesundheitsbezogene Informationen und Angaben zum Lebensstil gewonnen (101). In Folgeuntersuchungen wurden die Teilnehmenden regelmäßig erneut zu einem Untersuchungstermin eingeladen um die Veränderung der erhobenen Variablen über die Zeit zu erfassen. Außerdem werden kontinuierlich Informationen zum Vitalstatus der Teilnehmenden mittels Verknüpfung mit Routinedaten des UK National Health Service gewonnen (101).

Eine Teilstichprobe von 79.213 Teilnehmenden wurde für die Messung der kardiorespiratorischen Fitness ausgewählt. Hiervon lieferten 63.535 Teilnehmende valide Werte im Zuge einer inkrementell ansteigenden submaximalen Fahrradergometrie. Nach Ausschluss aller Teilnehmenden mit mindestens einem prävalenten malignen Karzinom zur Basisuntersuchung (ausgeschlossen nichtmelanozytärer Hautkrebs), lag für die Hauptanalysen eine Stichprobe von 59.191 Teilnehmenden vor (siehe Abbildung 8).

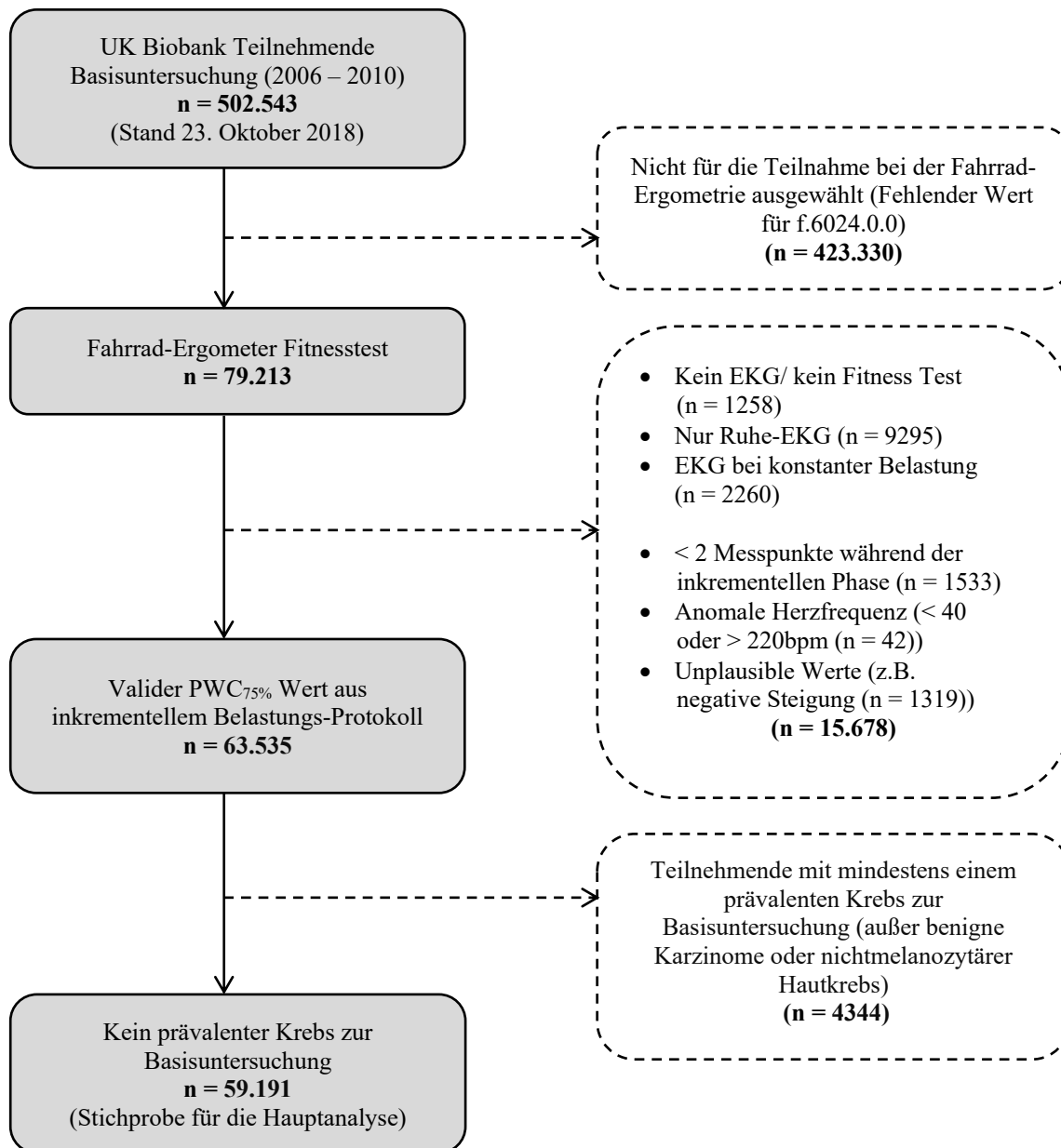


Abbildung 8: Flussdiagramm zum Ein- und Ausschluss der UK Biobank Teilnehmenden für die Analyse der Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom

4.2.2 Messung der kardiorespiratorischen Fitness

Im Rahmen einer Ergänzung des Untersuchungsprogramms wurde ab 2009 in ausgewählten Studienzentren (Sheffield, Liverpool, Hounslow, Croydon, Birmingham) auch ein submaximaler Fitnessstest mittels Fahrradergometrie durchgeführt (102). Der Fitnessstest wurde mit einem stationären eBike (Firmware v1.7) durchgeführt. Für die Aufzeichnung der EKGs wurde ein 4-Kanal-EKG-Gerät (CAM-USB 6.5 ausgestattet mit der Software Cardisoft v6.51) verwendet. Die für die Fahrradergometrie ausgewählte Teilstichprobe wurde basierend auf

Interviewfragen (z.B. Brustschmerz in Ruhe) und körperlichen Untersuchungen (z.B. Blutdruckmessung) in fünf Risikokategorien unterteilt (102):

- Minimales Risiko: Ausbelastung bis 50 % des individuell ermittelten maximalen Tretwiderstands.
- Geringes Risiko: Ausbelastung bis 35 % des individuell ermittelten maximalen Tretwiderstands.
- Moderates Risiko: Ergometertest mit konstantem Tretwiderstand.
- Hohes Risiko: Ruhe-EKG Messung.
- EKG gefährlich oder sinnlos: kein Fitness Test und keine EKG-Messung.

Nur bei Teilnehmenden, mit minimalem oder geringem Risiko wurde ein Fahrradergometer-Fitnesstest mit stufenweise ansteigender Belastung durchgeführt (102). Der maximal zulässige Tretwiderstand wurde vorab anhand von Alter, Geschlecht, Größe, Gewicht und Ruhe-Herzfrequenz ermittelt. Da es sich aus Sicherheitsgründen um einen submaximalen Belastungstest handelte, wurden die Teilnehmenden einem Protokoll zugeteilt, bei welchem die ansteigende Belastung maximal 35 % bzw. 50 % dieses Tretwiderstands erreichte. Der Ablauf des inkrementellen Fahrradergometer-Fitnesstest war dabei wie folgt (102):

- 15 Sekunden Vortest Phase: EKG Ableitung ohne Belastung bzw. Treten.
- 2 Minuten konstante Belastung bei 30 Watt (Frauen) bzw. 40 Watt (Männer).
- 4 Minuten ansteigende Belastung bis 35 % des maximalen Tretwiderstands bei geringem Risiko bzw. 50 % des maximalen Tretwiderstands bei minimalem Risiko.
- 1 Minute Erholungsphase ohne Belastung bzw. Treten.

Abbildung 9 zeigt beispielhaft den Verlauf von Tretwiderstand und Herzfrequenz über die Zeit im Rahmen einer Ergometermessung.

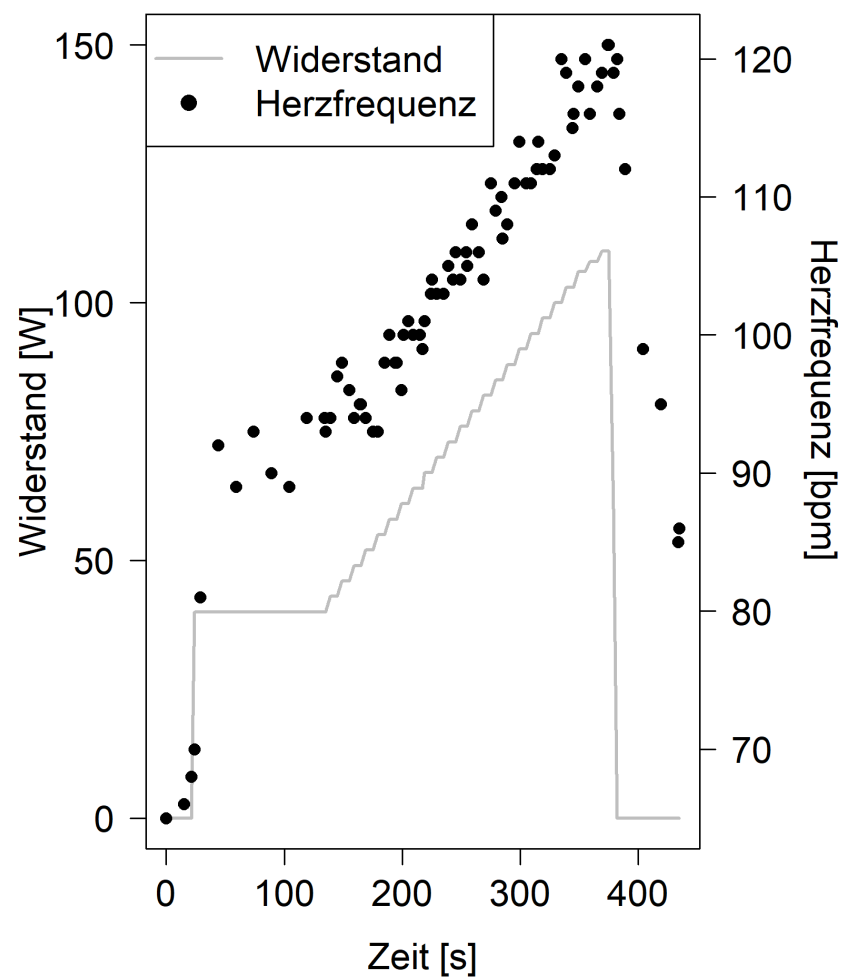


Abbildung 9: Ablauf der Ergometrie mit inkrementell ansteigender Belastung (schematische Darstellung)

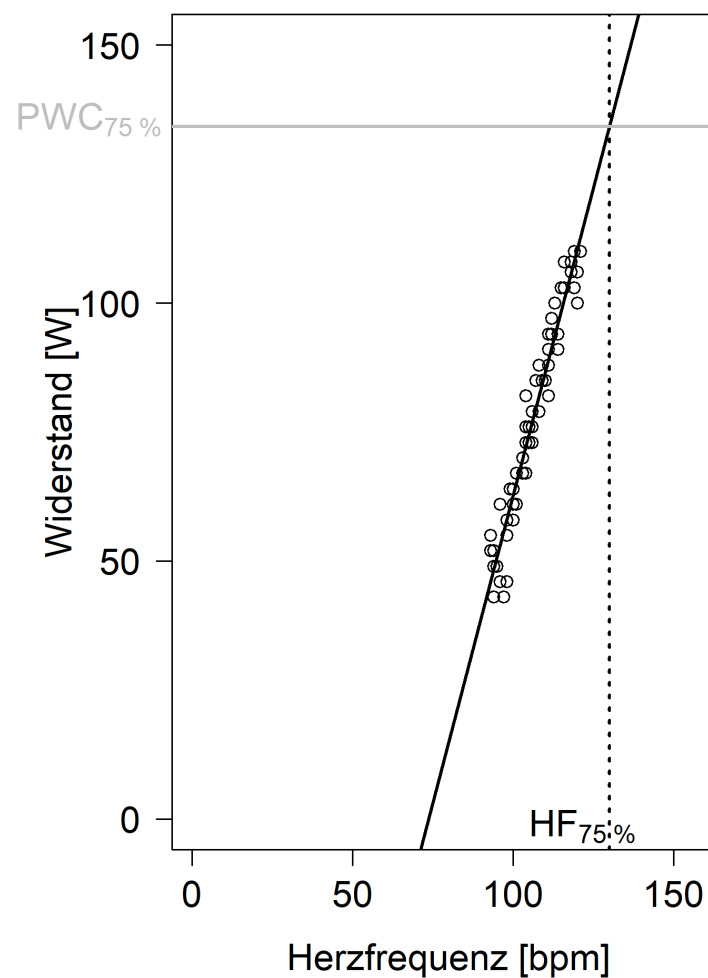


Abbildung 10: Schätzung der $PWC_{75\%}$ während der inkrementellen Phase (schematische Darstellung)

Bei maximalen Ergometertests mit paralleler Spirometrie kann die $\dot{V}O_{2\max}$ (maximale Sauerstoffaufnahme) direkt bestimmt werden (103). Selbst bei submaximalen Ergometer-Tests kann die $\dot{V}O_{2\max}$ durch Extrapolation unter Annahme eines linearen Zusammenhangs zwischen Herzfrequenz und Leistung geschätzt werden. Diese Schätzungen zeigten gute Validität (Korrelationskoeffizienten von 0,69-0,98) und Reliabilität (0,92) (104, 105). Da der Tretwiderstand im UK Biobank Ergometertest aus Sicherheitsgründen aber auf maximal 50 % der individuellen Belastungsgrenze beschränkt wurde und ein linearer Anstieg bis zur maximalen Herzfrequenz vor allem bei älteren Personen nicht gesichert ist (102, 106, 107), wurde die $\dot{V}O_{2\max}$ nicht ermittelt. Stattdessen wurde als Maß für die kardiorespiratorische Fitness die sogenannte Physical Work Capacity bei 75 % der maximalen Herzfrequenz ($PWC_{75\%}$) geschätzt (106, 108). Hierbei wurden die beobachteten Werte nur bis 75 % der altersabhängigen maximalen Herzfrequenz (berechnet als $208 - 0,7 \cdot \text{Alter}$) intra- oder leicht extrapoliert und die Leistung an der entsprechenden Stelle unter Annahme eines linearen Zusammenhangs auf der y-Achse abgelesen (siehe Abbildung 10) (106, 108, 109).

$$PWC_{75\%} \left[\frac{W}{kg} \right] = \frac{(\text{Intercept} + (0,75 \cdot (208 - 0,7 \cdot \text{Alter})) \cdot \text{Steigung})}{\text{Körpergewicht}}$$

Zusätzlich wurde die $PWC_{75\%}$ auf das individuelle Körpergewicht standardisiert, um die Leistung pro Kilogramm Körpergewicht (W/kg) zu erhalten (106, 108). Man erhält so ein zwischen verschiedenen Teilnehmenden vergleichbares Maß, welches den altersabhängigen Abfall der maximalen Herzfrequenz berücksichtigt und unabhängig von der fehleranfälligen Ruhe-Herzfrequenz berechnet werden kann (106, 108). Um unplausible Werte zu entfernen aber gleichzeitig Informationsverlust vorzubeugen wurden die $PWC_{75\%}$ Werte abschließend an der 1. und 99. Perzentile der Frauen und Männer winsorisiert (siehe Abbildung 11) (110).

Im Gegensatz zur $\dot{V}O_{2\max}$ stellt die $PWC_{75\%}$ keine absolute Größe dar, jedoch ist die relative Bewertung der Fitness und deren Assoziation mit der Krebsinzidenz das Ziel der Analysen. In vorangegangenen Studien zeigte die $PWC_{75\%}$ eine gute Augenscheinvalidität sowie eine hohe Korrelation mit einem weiteren Fitnessindex, der die Leistung bei 65% der Herzfrequenzreserve pro Kilogramm Körpergewicht angibt (ICC Koeffizient = 0,96) (106, 108). Außerdem wurde die $PWC_{75\%}$ bereits in verschiedenen Studien zur Messung der körperlichen Fitness verwendet (111-116).

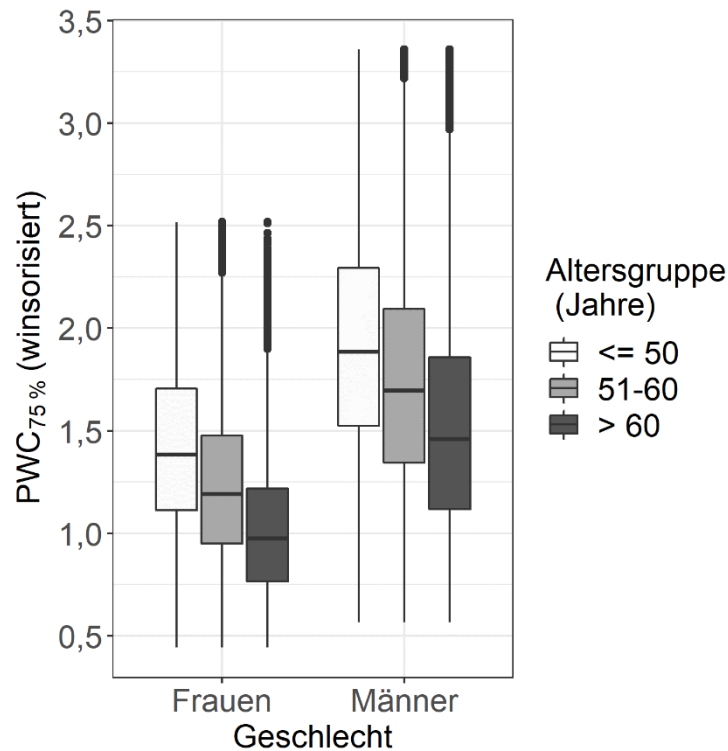


Abbildung 11: Kardiorespiratorische Fitness stratifiziert nach Alter und Geschlecht

4.2.3 Nachbeobachtung und Verknüpfung mit Routinedaten

Informationen zu Krebserkrankungen und Mortalität wurden durch Verknüpfung der UK Biobank Teilnehmenden mit Nationalen Krebs- und Mortalitätsregistern (National Health Service/ National Health Service Scotland) gewonnen (101, 117, 118). Diese erhalten ihre Informationen unter anderem von (privaten) Krankenhäusern, Krebszentren, Krebsscreening Programmen oder Sterbeurkunden (117). Die Vollständigkeit der Fallerfassung in Englischen Krebsregistern beträgt laut einer Studie 98 – 99 % (119). Die Krebslokalisationen wurden mit der Internationalen statistischen Klassifikation der Krankheiten und verwandten Gesundheitsprobleme (ICD) codiert. Die registrierten Krebsfälle der UK Biobank Teilnehmenden reichen bis in die 1970er Jahre zurück, weshalb zunächst die 9. Version und ab ca. 2000 die Version ICD-10 verwendet wurde (117, 120, 121). Tumor-Morphologie und Histologie wurden mittels ICD-O-3 systematisiert (122). Da die Integration der Registerdaten in den Datenbestand der UK Biobank zeitverzögert stattfindet, stehen jeweils nur vollständige Informationen bis zu einem bestimmten Zeitpunkt zur Verfügung (117).

Für die Analysen von kardiorespiratorischer Fitness und kolorektalem Karzinom lagen vollständige Krebsdaten bis zum 30. November 2014 vor. Als Endpunkt wurde die erste Kolorektalkarzinom Diagnose nach der Basisuntersuchung mit primär maligner oder maligner

mikroinvasiver Histologie gezählt. Kolonkarzinom wurde durch den ICD-Code C18 und Rektumkarzinom durch die ICD-Codes C19 und C20 definiert. Die Sublokalisierung proximales Kolonkarzinom wurde durch die ICD-Codes C18.0 - C18.5 definiert, die Sublokalisierung distales Kolonkarzinom durch die ICD-Codes C18.6 - C18.7. Andere Kolonkarzinom-Lokalisationen (C18.8 und C18.9) wurden aufgrund der zu geringen Fallzahl nicht einzeln betrachtet, sondern nur in den übergeordneten Analysen zu Kolonkarzinom (C18) oder kolorektalem Karzinom (C18 - C20). Bei zeitgleich diagnostizierten kolorektalen Karzinomen unterschiedlicher Lokalisation, wurden diese Fälle nur in den übergeordneten Analysen (Kolonkarzinom oder kolorektales Karzinom) gezählt. Als Nachbeobachtungszeit wurde die Dauer von der Basisuntersuchung bis zum Zeitpunkt der ersten Diagnose von kolorektalem Karzinom, dem Zeitpunkt der vollständigen Verknüpfung oder des Todeszeitpunkts gezählt.

4.2.4 Kovariablen

Um die Assoziation zwischen Exposition und Endpunkt ohne den verzerrenden Einfluss von Störgrößen zu berechnen, wurden die Regressionsmodelle für a priori definierte Kovariablen adjustiert. Gerichtete azyklische Graphen (DAG) sind ein immer beliebter werdendes Mittel zur Visualisierung kausaler Strukturen und zur Identifizierung von Störgrößen in Abgrenzung zu beispielsweise Intermediärvariablen und Collidern (123, 124). Das Disjunctive Cause Criterion wurde zudem verwendet, um eine minimal ausreichende Auswahl an Kovariablen zu erhalten (125). Hierbei wurden alle Ursachen von Exposition und Endpunkt (ausschließlich reiner Instrumentvariablen) zur Adjustierung ins Modell aufgenommen. Kovariablen, die nur Ursache des Endpunktes sind, stellen zwar keine Störgröße dar, Adjustierung für solche verbessert aber die Präzision des Ergebnisses (125, 126). Wenn alle kausalen Annahmen korrekt getroffen wurden und für alle notwendigen Variablen adjustiert wurde, kann auch mittels Beobachtungsdaten Kausalität abgeleitet werden (127).

Im Falle der Assoziation zwischen kardiorespiratorischer Fitness und kolorektalem Karzinom, wurden die Basis-Analysen für Alter (kontinuierlich (Jahre); als Zeitachse), Geschlecht (Frauen/Männer) und Studienzentrum (Sheffield/Liverpool/Hounslow/Croydon/Birmingham) adjustiert. Voll adjustierte Modelle berücksichtigten zusätzlich: Bildung (Universitäts-, College-Abschluss/A-Level, AS-Level, NVQ, HND, HNC oder äquivalent oder anderer professioneller Abschluss/O-Level, CSEs oder äquivalent/Keiner der oben genannten Abschlüsse), Haushalts-Netto-Einkommen (< 31.000 £/31.000 – 51.999 £/> 51.999 £), Sedentäres Verhalten (h/d Fernsehschauen), Taillen-Hüft-Verhältnis (Taillen-umfang in cm geteilt durch Hüftumfang in cm), Körpergröße (cm), Alkoholkonsum

(Nie/Vormals/Gegenwärtig), Rauchverhalten (Nie/Vormals/Gegenwärtig), Verzehr von rotem und verarbeitetem Fleisch (< 2 Mal pro Woche/2 – 3 Mal pro Woche/> 3 Mal pro Woche), Darmkrebsscreening in der Vergangenheit (Ja/Nein), Familienhistorie von Darmkrebs (Ja/Nein) und regelmäßige Einnahme von nichtsteroidalen Antiphlogistika (NSAID) (Ja/Nein). Frauenspezifische Subgruppen-Analysen wurden zusätzlich für die Einnahme von Hormonersatzpräparaten (Ja/Nein) adjustiert. In weiteren Sensitivitätsanalysen, wurde zusätzlich für prävalenten Diabetes mellitus (Ja/Nein) und körperliche Aktivität (MET-Minuten pro Woche, erfragt mit einer Kurzform des International Physical Activity Questionnaires (IPAQ-SF)) adjustiert.

Abbildung S 4 im Anhang zeigt ein DAG für die kausalen Zusammenhänge der einzelnen Störgrößen mit Exposition und Endpunkt sowie untereinander (124). Familienhistorie von kolorektalem Karzinom (91, 128), Rauchverhalten (129), Alkoholkonsum (130), Verzehr von rotem und prozessiertem Fleisch (131, 132), Übergewicht (133), Größe (134), steigendes Alter (88, 91), männliches Geschlecht (88, 135), Diabetes (136-138) und sedentäres Verhalten (139) sind Risikofaktoren für die Entwicklung von kolorektalem Karzinom. Im Gegensatz hierzu zeigten die regelmäßige Einnahme von NSAIDs (durch die Reduzierung entzündlicher Prozesse) (140-142) und die Teilnahme an Screening-Untersuchungen (durch Entdeckung und Entfernung von Adenomen oder anderen Vorstufen) (88, 91, 92, 143) einen inversen Zusammenhang mit dem Risiko für kolorektales Karzinom. Bei Frauen zeigte die Einnahme von Hormonersatzpräparaten, möglicherweise durch protektive Effekte der Östrogene, eine inverse Assoziation zum Risiko für kolorektale Karzinome (144, 145). In vorangegangenen Studien zeigte sich zudem ein inverser Zusammenhang zwischen aktivem und passivem Rauchen (146), Übergewicht (147) und sedentärem Verhalten (148) mit kardiorespiratorischer Fitness. Auf der anderen Seite zeigte sich ein positiver Zusammenhang zwischen sozioökonomischem Status, vor allem Bildung, und kardiorespiratorischer Fitness (149). In einer Metaanalyse wurde eine inverse Assoziation zwischen kardiorespiratorischer Fitness und dem Risiko an Diabetes Mellitus Typ 2 zu erkranken herausgestellt (150). Dies deutet daraufhin, dass Diabetes eine Intermediärvariable auf dem Pfad zwischen kardiorespiratorischer Fitness und kolorektalem Karzinom darstellt. Das bedeutet, dass ein Teil des Effekts von kardiorespiratorischer Fitness auf das Krebsrisiko durch den Effekt auf Diabetes vermittelt wird. Um den Gesamteffekt von kardiorespiratorischer Fitness zu erhalten, wurden die Hauptanalysen nicht für Diabetes kontrolliert. Körperliche Aktivität ist möglicherweise die Hauptkomponente für kardiorespiratorische Fitness (2) und

zudem ein protektiver Faktor bei der Entstehung von kolorektalem Karzinom (139). Um jedoch sowohl den genetisch determinierten Teil, als auch den durch körperliche Aktivität bedingten und somit modifizierbaren Teil der kardiorespiratorischen Fitness zu untersuchen, wurden die Hauptanalysen nicht für körperliche Aktivität adjustiert. Die Analysen wurden ebenfalls nicht für Ethnizität adjustiert, da Studienergebnisse darauf hinweisen, dass nicht Ethnizität, sondern Ungleichheiten in der Gesundheitsversorgung mit dem Risiko für kolorektales Karzinom assoziiert sind (151-153). Die Analysen wurden auch nicht für koronare Herzkrankheiten adjustiert, da diese teilweise die gleichen Risikofaktoren wie das kolorektale Karzinom besitzen und auch mit kardiorespiratorischer Fitness assoziiert sind (100, 154). Somit stellen koronare Herzkrankheiten einen Collider dar, dessen Kontrolle zu einer Verzerrung der wahren Assoziation zwischen kardiorespiratorischer Fitness und kolorektalem Karzinom führen würde (123).

4.2.5 Statistische Analysen

Durch direkte Standardisierung auf die Altersverteilung der Kohorte, wurden die Eigenschaften der Studienpopulation zur Basisuntersuchung altersadjustiert und nach geschlechtsspezifischen Quartilen von kardiorespiratorischer Fitness stratifiziert dargestellt.

Um die statistische Teststärke in den Analysen zu kardiorespiratorischer Fitness und kolorektalem Karzinom zu erhöhen und einen potentiellen Selektionsbias zu minimieren, wurden Kovariablen mit fehlenden Werten imputiert. Zur Berücksichtigung statistischer Unsicherheit, wurde eine multiple Imputation mittels chained equations (R package "mice" version 2.03) durchgeführt (155, 156). Als Imputationsmodell wurde für kontinuierliche Variablen predictive mean matching, für binäre Variablen logistische Regression, für ordinale Variablen ein ordered logit Model und für nominale Variablen ein multinomial logit Model verwendet (155). Zusätzlich zu relevanten Prädiktoren wurden der Nelson Aalen Schätzer und ein Endpunkt Indikator verwendet, wie für Überlebenszeitdaten empfohlen (157). Die Imputation wurde mit jeweils fünf Iterationen zehn Mal durchgeführt. Konvergenz und Plausibilität der Imputation wurden mithilfe diagnostischer Plots überprüft (155).

Die Assoziation zwischen kardiorespiratorischer Fitness und dem Risiko für kolorektales Karzinom wurde mittels Cox proportional hazards Regressionsmodellen mit Alter als zugrundeliegender Zeitachse berechnet. Da nicht die Zeit, sondern steigendes Alter das Risiko für Erkrankungen erhöht, wird dieses Vorgehen bei Überlebenszeitmodellen empfohlen (158,

159). Als Startpunkt wurde das Alter in Jahren zur Basisuntersuchung verwendet. Als Endpunkt wurde Alter zum Zeitpunkt der vollständigen Verknüpfung, Alter zur Krebsdiagnose oder Alter zum Todeszeitpunkt verwendet, je nachdem welches der Ereignisse zuerst eintrat. Kardiorespiratorische Fitness ($PWC_{75\%}$ [W/kg]) wurde als kontinuierliche Variable in das Modell aufgenommen und die Log-Linearität des Zusammenhangs mit dem Krebsrisiko mittels restricted cubic splines festgestellt (160). Hazard Ratios (HR) und dazugehörige 95 % Konfidenzintervalle (KI) für einen Anstieg um einen Interquartilsabstand (z.B. 25. zu 75. Quartile) der Expositionsvariable wurden berichtet (160). Die Annahme proportionaler hazards wurde mit Hilfe von Schoenfeld Residuen überprüft und im Falle einer Abweichung bei kategorialen Variablen wurden stratifizierte Cox Modelle gerechnet. Die Cox proportional hazards Modelle wurden separat für jeden Kolorektalkarzinom Endpunkt (kolorektales Karzinom, Kolonkarzinom, distales und proximales Kolonkarzinom, Rektumkarzinom) sowie zusätzlich stratifiziert nach Geschlecht gerechnet. Eine potenzielle Effektmodifikation durch Geschlecht wurde mittels Wald Test überprüft. Die Heterogenität zwischen den Kolorektalkarzinom-Subtypen wurde mit Hilfe eines competing risk Ansatzes und Wald-Tests berechnet (161).

Die Robustheit der getroffenen Annahmen wurde in weiteren zahlreichen Sensitivitätsanalysen überprüft. E-values wurden berechnet, um den potenziellen Einfluss nicht gemessener Störgrößen auf die Ergebnisse zu quantifizieren. Der E-value gibt das Relative Risiko an, welches ein nicht gemessener Confounder mit Exposition und Outcome haben müsste, um die beobachtete Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom zu erklären (162, 163). Wie bereits im Abschnitt 4.2.4 beschrieben, stellt Diabetes mellitus wahrscheinlich eine Intermediärvariable dar, für welche deshalb in den Hauptanalysen nicht adjustiert wurde. Um die Auswirkung alternativer Annahmen auf das Ergebnis zu testen, wurde in einer Sensitivitätsanalyse für Diabetes adjustiert. Um nur den genetisch determinierten Anteil der kardiorespiratorischen Fitness zu betrachten, und potenzielle Verzerrung durch körperliche Aktivität, die Hauptdeterminante von kardiorespiratorischer Fitness (2), auszuschließen, wurde zudem eine Analyse mit Adjustierung für körperliche Aktivität (MET-Minuten pro Woche) gerechnet. Umgekehrte Kausalität kann auftreten, wenn das Kolorektalkarzinom bereits vor der Basisuntersuchung vorlag und somit die kardiorespiratorische Fitness beeinflusste. Um den Einfluss dieses Fehlers zu minimieren, wurden in einer Sensitivitätsanalyse alle Teilnehmenden ausgeschlossen, bei welchen das Kolorektalkarzinom innerhalb von 12 Monate nach der Basisuntersuchung diagnostiziert wurde. Des Weiteren

wurden in einer Sensitivitätsanalyse alle Teilnehmende ausgeschlossen, welche ihren Gesundheitsstatus als „schlecht“ beschrieben. Dies sollte Verzerrung der Ergebnisse durch den Gesundheitsstatus zur Basisuntersuchung beziehungsweise einen nicht diagnostizierten Kolorektalkrebs ausschließen. Um ausreichende Teststärke zu erhalten, wurden alle Teilnehmende mit mindestens zwei Datenpunkten während der inkrementell ansteigenden Belastungsphase der Ergometrie eingeschlossen. Allerdings kann dieses Vorgehen zu einer differenziellen Misklassifizierung der kardiorespiratorischen Fitness führen. Deshalb wurde eine Sensitivitätsanalyse durchgeführt, bei welcher ausschließlich Teilnehmende mit ≥ 30 Sekunden andauernder inkrementell ansteigender Belastungsphase betrachtet wurden. Um die Plausibilität der Imputation zu überprüfen, wurden die Analysen in einer Subgruppe berechnet welche nur beobachtete Werte enthielt (Complete Case Analyse). In einer weiteren Sensitivitätsanalyse wurden zusätzlich zu den imputierten Kovariablen auch imputierte Expositionsvariablen (fehlende oder unplausible PWC_{75%}-Werte aus der Ergometrie-Stichprobe) analysiert. Zur besseren Vergleichbarkeit mit vorherigen Studien, wurde die kardiorespiratorische Fitness in einer zusätzlichen Analyse in alters- und geschlechtsspezifische Terzile (niedrig/mittel/hoch) eingeteilt. Die Analysen wurden zusätzlich stratifiziert nach alters- und geschlechtsspezifischen Terzilen von körperlicher Aktivität berechnet. Multiplikative Effektmodifikation zwischen kardiorespiratorischer Fitness und Studienzentrum, Bildung, Einkommen, körperlicher Aktivität, sedentärem Verhalten, Taillen-Hüft-Verhältnis, Größe, Alkoholkonsum, Rauchverhalten, Verzehr von rotem oder verarbeitetem Fleisch, Darmkrebsscreening, Familienhistorie von Darmkrebs, prävalentem Diabetes, Einnahme von NSAIDs und (bei den Frauen) Einnahme von Hormonersatzpräparaten wurde mittels Wald Tests berechnet und nach Bonferroni für multiples Testen korrigiert. Alle statistischen Analysen wurden mit der Statistik Software R Version 3.3.2 durchgeführt (siehe auch Ausschnitte des R Codes in Abschnitt 7.4.4) (164).

4.3 Ergebnisse

Die durchschnittliche Nachbeobachtungszeit betrug 4,6 Jahre (Standardabweichung 0,3 Jahre). In insgesamt 271.504,5 Personenjahren traten bei den 59.191 Teilnehmenden 232 Kolorektalkarzinome auf, davon 151 Kolonkarzinome (74 im proximalen und 61 im distalen Teil des Kolons) und 79 Rektumkarzinome. In der Subgruppe der 28.483 Männer traten insgesamt 144 Kolorektalkarzinome, bei den 30.708 Frauen traten 88 Kolorektalkarzinome auf. Die Inzidenzrate war somit bei den Männern mit 11,04 pro 10.000 Personenjahren fast doppelt so hoch als bei den Frauen mit 6,24 pro 10.000 Personenjahren.

Die altersadjustierten Charakteristika der Studienpopulation zur Basisuntersuchung zeigten, dass Teilnehmende in der höchsten Quartile von kardiorespiratorischer Fitness zu einem größeren Anteil einen tertiären Schulabschluss und hohes Haushaltseinkommen angaben, als die Teilnehmenden in den anderen Quartilen. Ebenso berichteten die fittesten Teilnehmenden mehr körperliche Aktivität und weniger sedentäres Verhalten. Ein kleinerer Anteil an Teilnehmenden mit hoher kardiorespiratorischer Fitness beschrieb den Gesamt-Gesundheitszustand als „schlecht“. Der Anteil der derzeitigen Raucher und Alkoholkonsumenten war in der Gruppe der fitten Teilnehmenden größer als in den weniger fitten Gruppen (Tabelle 3).

In den für Alter, Geschlecht und Studienzentren adjustierten Basismodellen war die HR für Kolorektalkarzinom pro Interquartilsanstieg an kardiorespiratorischer Fitness 0,79 (95 % KI 0,65-0,98). Diese inverse Assoziation blieb auch nach voller Adjustierung bestehen (HR 0,78; 95 % KI 0,62-0,97). In den Geschlechtsstratifizierten Basis-Modellen war die inverse Assoziation bei der Gruppe der Männer (HR 0,74; 95 % KI 0,57-0,94), aber nicht bei der Gruppe der Frauen (HR 0,99; 95 % KI 0,72-1,36) zu sehen. Auch hier änderte volle Adjustierung die Ergebnisse nur marginal (HR 0,72, 95 % KI 0,55-0,94 bei den Männern; HR 0,99, 95 % KI 0,71-1,38 bei den Frauen, Abbildung 12, S. 50). Die Effektmodifikation durch Geschlecht war nicht signifikant ($p = 0,192$).

Bei der Betrachtung der Krebs-Sublokalisationen (Abbildung 12) zeigte sich eine inverse Assoziation zwischen kardiorespiratorischer Fitness und Kolonkarzinom im Basismodell (HR 0,75, 95 % KI 0,57-0,98), welche auch nach voller Adjustierung erhalten blieb (HR 0,74, 95 % KI 0,56-0,97). Bei den nach Geschlecht stratifizierten Analysen war der Nullwert von 1,0 in allen Konfidenzintervallen enthalten. Dennoch war auch hier die HR für Kolonkarzinom pro Interquartilsanstieg von kardiorespiratorischer Fitness bei den Männern kleiner als bei den Frauen (z.B. im voll adjustierten Modell HR 0,72; 95 % KI 0,51-1,01 bei den Männern und HR 0,84; 95 % KI 0,56-1,26 bei den Frauen; p -Wert für Interaktion durch Geschlecht = 0,802). Bei einer weiteren Unterteilung der Sublokalisation Kolonkarzinom in den distalen und proximalen Teil, enthielten alle Konfidenzintervalle den Nullwert 1,0.

Tabelle 3: Charakteristika der Studienpopulation zur Basisuntersuchung nach direkter Standardisierung auf die Altersverteilung der Kohorte pro Quartile von PWC_{75%}

	Geschlechtsspezifische Quartile von PWC _{75%}			
	Q1	Q2	Q3	Q4
Anzahl an Teilnehmenden	14.798	14.798	14.797	14.798
Kardiorespiratorische Fitness (PWC_{75%} in W/kg)	0,9	1,3	1,6	2,1
Alter zur Basisuntersuchung (Jahre)	56,7	56,3	56,0	55,8
Frauen (%)	52,0	51,7	51,3	49,0
UK Biobank Studienzentrum				
Sheffield	23,5	24,1	25,6	26,2
Liverpool	8,3	8,4	8,1	7,0
Hounslow	16,7	18,3	19,3	20,4
Croydon	22,6	23,2	24,1	26,5
Birmingham	29,0	26,0	23,0	20,0
Bildung (678 NA)				
Universitäts-, College-Abschluss	29,2	34,4	39,1	45,3
A-, AS-Level, NVQ, HND, HNC, äquivalent, anderer professioneller Abschluss	24,4	24,4	23,6	22,6
O-Level, CSEs oder äquivalent	29,9	28,6	26,7	23,5
Keiner der oben genannten Abschlüsse	16,5	12,5	10,6	8,6
Haushalts-Netto-Einkommen (7754 NA)				
< 31,000 £	53,1	45,3	39,8	34,6
31,000 £ - 51,999 £	25,9	26,6	27,6	27,1
> 51,999 £	21,0	28,1	32,6	38,3
Körperliche Aktivität (MET-Minuten/Woche) (11.547 NA)	2512,0	2721,2	2938,6	3330,5
Fernseh schauen (h/d) (531 NA)	3,1	2,7	2,5	2,3
Taillen-Hüft-Verhältnis (3 NA)	0,9	0,9	0,9	0,9
Körpergröße (cm)	167,8	169,0	169,7	170,9
Alkoholkonsum (204 NA)				
Nie	6,9	4,8	3,6	2,5
Vormals	3,5	3,6	2,7	2,4
Gegenwärtig	89,6	91,6	93,7	95,0
Rauchverhalten (333 NA)				
Nie	59,2	56,7	55,6	55,0
Vormals	32,4	33,9	35,2	35,9
Gegenwärtig	8,4	9,4	9,2	9,2
Verzehr von rotem und verarbeitetem Fleisch (168 NA)				
< 2 Mal pro Woche	13,7	14,7	15,1	16,7
2 - 3 Mal pro Woche	41,4	44,1	46,3	47,6
> 3 Mal pro Woche	45,0	41,2	38,6	35,7
Darmkrebsscreening (%) (949 NA)	40,2	41,0	41,7	41,7
Familienhistorie von Darmkrebs (%) (7142 NA)	9,9	10,3	10,7	10,7
Prävalenter Diabetes (%) (318 NA)	6,3	3,4	2,3	2,1
Schlechter Gesundheitszustand (%) (314 NA)	4,5	2,9	1,8	1,3
Einnahme von NSAIDs (%) (796 NA)	25,8	24,0	23,9	23,5
Hormonersatztherapie (%) (28.677 NA)	34,8	36,4	36,6	33,3

Q = Quartile, PWC_{75%} = Körperliche Leistungsfähigkeit bei 75 % der maximalen Herzfrequenz, NA = fehlende Werte, MET = Metabolisches Äquivalent, NSAID = nichtsteroidales Antiphlogistikum

Jedoch waren die stärksten Ausprägungen der inversen Assoziation zwischen kardiorespiratorischer Fitness und proximalem Kolonkarzinom bei den Männern zu beobachten (z.B. voll adjustiertes Modell HR 0,64; 95 % KI 0,38-1,08). Es zeigte sich keine Assoziation zwischen ansteigenden Leveln von kardiorespiratorischer Fitness und Rektumkarzinom in der gesamten Studienpopulation oder bei den geschlechtsstratifizierten Modellen (Abbildung 12).

Während bei den Männern HRs < 1 auftraten (z.B. voll adjustiertes Modell HR 0,76; 95 % KI 0,50-1,15), waren bei den Frauen HRs > 1 zu beobachten (z.B. voll adjustiertes Modell HR 1,47; 95 % KI 0,81-2,65; p-Wert für Interaktion durch Geschlecht = 0,027). Die p-Werte für die Subtyp-Heterogenität von Kolon- versus Rektalkarzinom betragen 0,056 in der gesamten Studienpopulation und 0,683 bei den Männern sowie 0,199 bei den Frauen. Für die Subtyp Heterogenität von proximalem Kolon- versus distalem Kolon- versus Rektumkarzinom betragen die p-Werte in der gesamten Studienpopulation 0,113, bei den Männern 0,659 und bei den Frauen 0,310 (Abbildung 12).

In der Complete Case Analyse (Abbildung 13 A) lagen nach Ausschluss aller Teilnehmenden mit fehlenden Angaben bei den Kovariablen nur noch 187 Kolorektalkarzinome vor. Die Richtungen der Assoziationen zwischen kardiorespiratorischer Fitness und kolorektalem Karzinom waren vergleichbar mit denen des imputierten Modells (z.B. kleinste HRs bei den Männern und beim proximalen Kolonkarzinom). Allerdings beinhalteten alle Konfidenzintervalle den Nullwert 1,0. Nach Imputation aller fehlenden oder implausiblen kardiorespiratorischen Fitness Werte der Ergometer-Subgruppe lagen 300 Kolorektalkarzinome vor (Abbildung 13 B). In der gesamten Stichprobe zeigten sich nach voller Adjustierung HRs pro Interquartilsanstieg an kardiorespiratorischer Fitness von 0,76 (95 % KI 0,62-0,93 für Kolorektalkarzinom), 0,71 (95 % KI 0,54-0,93 für Kolonkarzinom) und 0,66 (95 % KI 0,44-0,99 für proximales Kolonkarzinom). Bei den Männern war die HR pro Interquartilsanstieg an kardiorespiratorischer Fitness 0,70 (95 % KI 0,54-0,90 für Kolorektalkarzinom), 0,67 (95 % KI 0,48-0,94 für Kolonkarzinom) und 0,58 (95 % KI 0,35-0,96 für proximales Kolonkarzinom) nach voller Adjustierung.

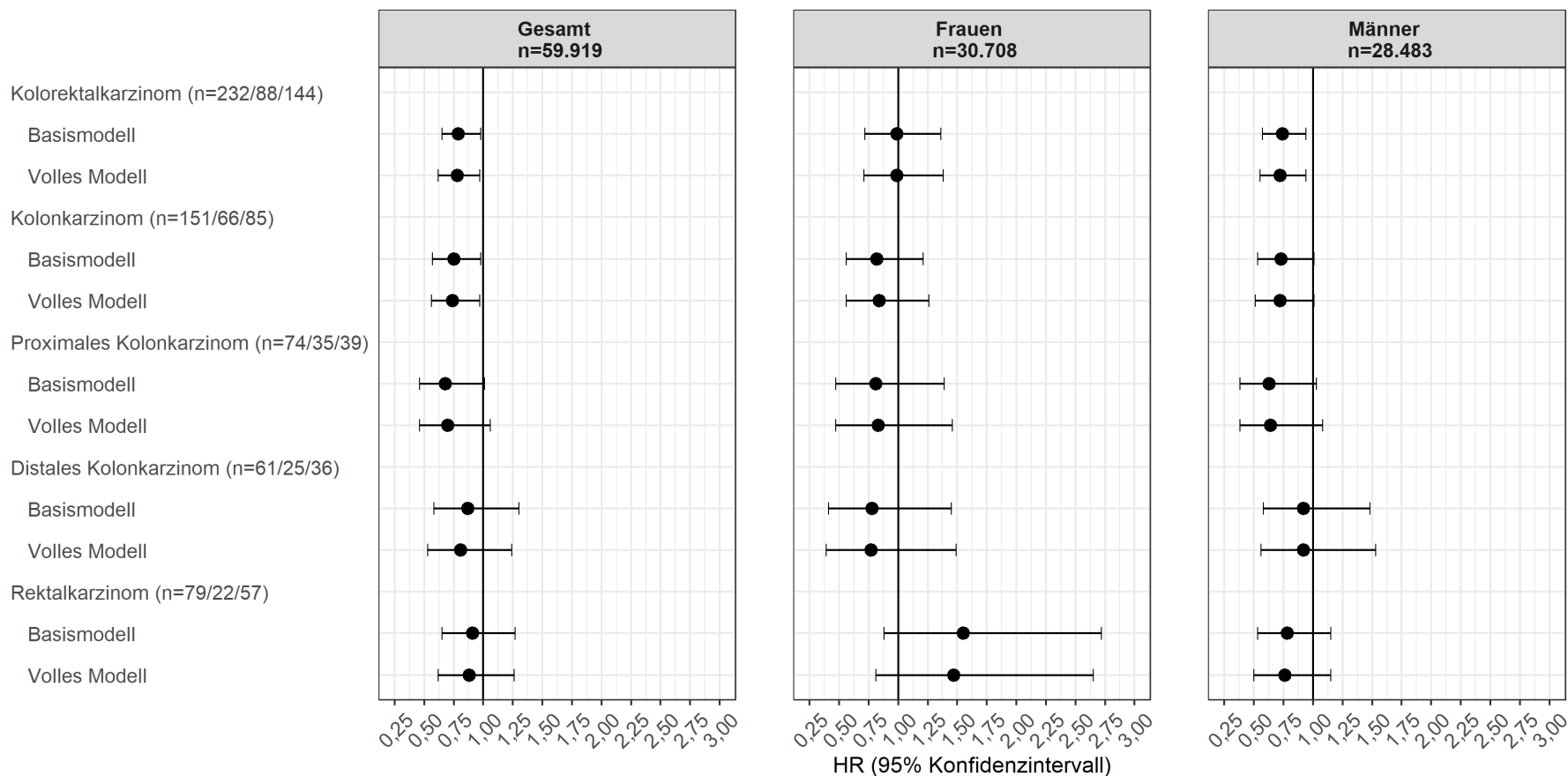
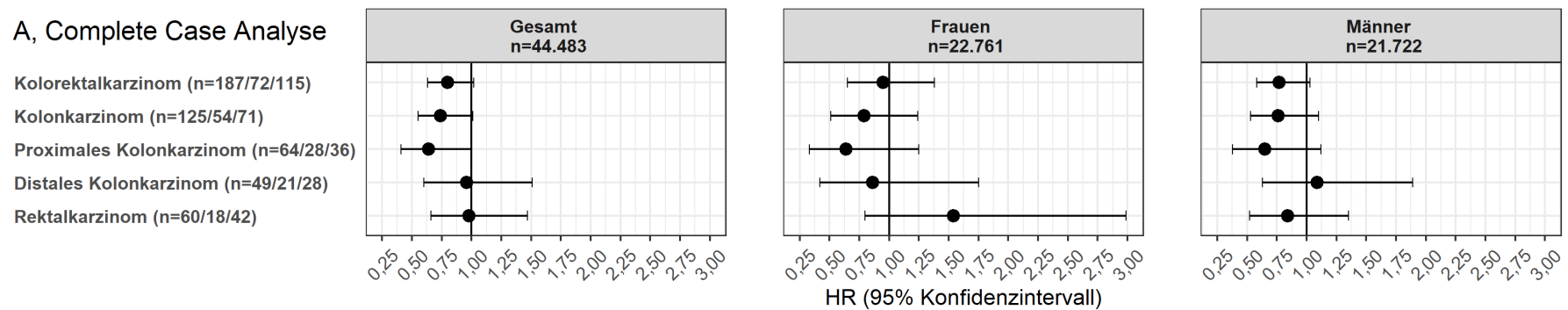


Abbildung 12: Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom

Multivariabel adjustierte Cox Proportional Hazards Regression mit Alter als Zeitachse. Kardiorespiratorische Fitness ($PWC_{75\%}$) wurde als kontinuierliche Variable ins Modell aufgenommen. Assoziationen wurden linear modelliert. Hazard Ratios wurden für einen Anstieg um einen Interquartilsabstand (z.B. 25. bis 75. Quantile) berechnet. Die Stichprobe beinhaltet Teilnehmende mit gemessenen, validen Werten für kardiorespiratorische Fitness. Fehlende Kovariablen wurden imputiert. Anzahl an Krebsfällen pro (Sub-)gruppe sind in Klammern angegeben. Basismodelle wurden für UK Biobank Studienzentrum und in der Gesamtgruppe zusätzlich für Geschlecht adjustiert. Volle Modelle wurden zusätzlich für Bildung, Einkommen, sedentäres Verhalten, Taillen-Hüft-Verhältnis, Körpergröße, Alkoholkonsum, Rauchverhalten, Verzehr von rotem und verarbeitetem Fleisch, Darmkrebsscreening, Familienhistorie von Darmkrebs und regelmäßige Einnahme von NSAIDs adjustiert. Modelle in der Frauen-Gruppe wurden zusätzlich für Hormonersatztherapie adjustiert.

A, Complete Case Analyse



B, Volle Imputation

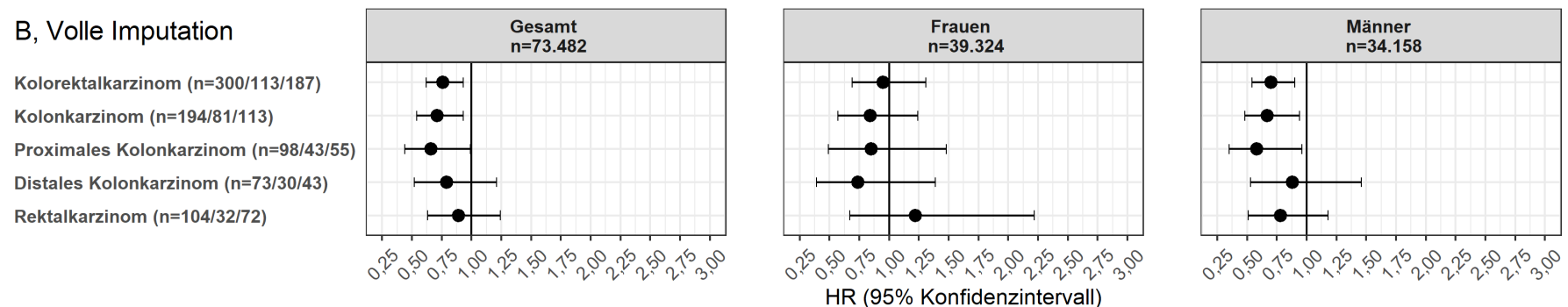


Abbildung 13: Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom bei Complete Case (A) und voller Imputation (B)

Multivariabel adjustierte Cox Proportional Hazards Regression mit Alter als Zeitachse. Kardiorespiratorische Fitness ($PWC_{75\%}$) wurde als kontinuierliche Variable ins Modell aufgenommen. Assoziationen wurden linear modelliert. Hazard Ratios wurden für einen Anstieg um einen Interquartilsabstand (z.B. 25. bis 75. Quantile) berechnet. Anzahl an Krebsfällen pro (Sub-)gruppe sind in Klammern angegeben.

Modelle wurden für UK Biobank Studienzentrum, Bildung, Einkommen, sedentäres Verhalten, Taillen-Hüft-Verhältnis, Körpergröße, Alkoholkonsum, Rauchverhalten, Verzehr von rotem und verarbeitetem Fleisch, Darmkrebscreening, Familienhistorie von Darmkrebs, regelmäßige Einnahme von NSAIDs und in der Gesamtgruppe zusätzlich für Geschlecht adjustiert. Modelle in der Frauen-Gruppe wurden zusätzlich für Hormonersatztherapie adjustiert

Für die Assoziation zwischen kardiorespiratorischer Fitness und kolorektalem Karzinom ergab sich ein E-value von 1,9. Die Assoziation einer nicht berücksichtigten Störgröße mit kardiorespiratorischer Fitness und Kolorektalkarzinom müsste somit ein Relatives Risiko (RR) von 1,9 aufweisen, um die beobachtete Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom von 0,78 zu erklären. Ebenso müsste die nicht berücksichtigte Störgröße mit einem Relativen Risiko von 1,2 mit Exposition und Endpunkt assoziiert sein, um das obere Konfidenzintervall der Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom von 0,97 auf über 1,00 zu verschieben. Zusätzliche Adjustierung der Hauptmodelle um Diabetes mellitus Typ 2 oder körperliche Aktivität veränderte die Ergebnisse quasi nicht (HR 0,78; 95 % KI 0,63-0,97 beziehungsweise HR 0,78; 95 % KI 0,63-0,98).

Nach Ausschluss von 42 Teilnehmenden, welche innerhalb der auf die Basisuntersuchung folgenden 12 Monate eine Kolorektalkarzinomdiagnose erhielten, änderten sich die Ergebnisse nicht essenziell (HR 0,85, 95 % KI 0,67-1,08). Nach Ausschluss von 1714 Teilnehmenden, welche zur Basisuntersuchung einen schlechten Gesundheitszustand berichteten, verblieben noch 228 Kolorektalkarzinomfälle in der Stichprobe. Die Assoziation zwischen kardiorespiratorischer Fitness und kolorektalem Karzinom blieb bestehen (HR 0,78; 95 % KI 0,62-0,97). Um einer differentiellen Misklassifikation der kardiorespiratorischen Fitness entgegen zu wirken, wurden in einer Sensitivitätsanalyse nur Personen betrachtet, welche eine mindestens 30 Sekunden andauernde inkrementelle Belastungsphase während der Ergometrie absolviert hatten. Hierdurch wurden 195 Teilnehmende, davon 1 Teilnehmender mit Kolorektalkarzinomdiagnose, ausgeschlossen. Die Assoziation zwischen kardiorespiratorischer Fitness und kolorektalem Karzinom änderte sich nicht (HR 0,78; 95 % KI 0,62-0,97).

Nach Kategorisierung der kardiorespiratorischen Fitness zeigte sich eine HR von 0,73 (95 % KI 0,51-1,06) für die Gruppe mit hoher versus niedriger Fitness. Die Stratifizierung nach alters- und geschlechtsspezifischen Terzilen von körperlicher Aktivität verringerte die statistische Teststärke und der Nullwert 1,0 war in allen Konfidenzintervallen enthalten. Nach Bonferroni Korrektur ergab sich kein Hinweis auf einen effektmodifizierenden Effekt von Studienzentrum, Bildung, Einkommen, körperliche Aktivität, sedentärem Verhalten, Taillen-Hüft-Verhältnis, Größe, Alkoholkonsum, Rauchverhalten, Verzehr von rotem oder verarbeitetem Fleisch, Darmkrebsscreening, Familienhistorie von Darmkrebs, Diabetes, Einnahme von NSAIDs und, bei den Frauen, Einnahme von Hormonersatzpräparaten auf die Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom.

4.4 Diskussion

In der betrachteten Stichprobe von knapp 60.000 Männern und Frauen, welche bis zu 5 Jahre nachbeobachtet wurden, zeigte sich eine inverse Assoziation zwischen kardiorespiratorischer Fitness und der Inzidenz von Kolorektalkarzinom, sowie Kolonkarzinom. Diese Assoziation schien vor allem in der Subgruppe der Männer und bei der Sublokalisierung proximales Kolonkarzinom stark ausgeprägt zu sein, obgleich Geschlecht kein Effektmodifikator und die Subtyp-Heterogenität nicht statistisch signifikant war.

Bisherige Studien zur Assoziation zwischen kardiorespiratorischer Fitness und dem Risiko von Kolorektalkarzinom zeigten heterogene Ergebnisse und wiesen teilweise Einschränkungen und Limitationen auf. Bis 2018 wurden ausschließlich Studien mit Männern durchgeführt, möglicherweise, da die maximale Fahrradergometrie hier aufgrund eines Schwangerschaftsausschlusses mit einem geringeren Risiko verbunden war. Die zwei ältesten Studien fanden eine inverse Assoziation zwischen kardiorespiratorischer Fitness und dem Risiko von Karzinomen des Gastrointestinaltrakts (HR 0,88; 95 % KI 0,79-0,99 pro 1 MET-Anstieg in der Kupio Ischaemic Heart Disease Risk Factor Study (96)) beziehungsweise der Inzidenz von Kolorektalkarzinomen (HR 0,91; 95 % KI 0,84-0,99 pro 1 MET-Anstieg in der Cooper Center Longitudinal Study (35)). Die Oslo Ischemia Study betrachtete als einzige bisherige Studie Sublokalisierungen des Kolonkarzinoms und fand eine inverse Assoziation zwischen kardiorespiratorischer Fitness und proximalem Kolonkarzinom (HR 0,30; 95 % KI 0,13-0,73 für mittlere versus niedrige kardiorespiratorische Fitness (34)). Diese Beobachtung konnten wir in unseren Analysen insofern bestätigen, dass die inversen Assoziationen bei proximalem Kolonkarzinom am stärksten ausgeprägt schienen. In zwei weiteren Untersuchungen (Veterans Exercise testing study (33) und Copenhagen Male Study (98)) mit ausschließlich männlichen Studienteilnehmern zeigten sich keine Assoziationen zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom. Allerdings wurden die meisten dieser Studien nicht für alle relevanten Kovariablen, wie beispielsweise Familienhistorie von Darmkrebs, Alkoholkonsum oder Sozioökonomischer Status, adjustiert. Eine Metaanalyse fasste die Ergebnisse der Studien (33-35, 98) zusammen und berichtete eine HR von 0,77 (95 % KI 0,62-0,92) für hohe versus niedrige sowie eine HR von 0,74 (95 % KI 0,55-0,93) für moderate versus niedrige kardiorespiratorische Fitness (38), was eine sehr gute Übereinstimmung mit den von uns ermittelten Ergebnissen zeigt. Relativ zeitnah mit unseren Analysen wurden zwei weitere Publikationen veröffentlicht. In der UK Biobank zeigte sich eine inverse Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkrebsinzidenz

(HR 0,96; 95 % KI 0,92-1,00) (97). Die Analysen unterschieden sich von unseren insofern, dass als Maß für die kardiorespiratorische Fitness maximale METs aus dem submaximalen Fitnessstest verwendet wurden, die Kolorektalkarzinom-Diagnosen aus der Verknüpfung mit den Krankenhausdaten ermittelt wurden und die Assoziationen nicht stratifiziert nach Geschlecht oder anatomischer Kolon-Sublokalisierung betrachtet wurden (97). Die Analysen in der Henry Ford Exercise Testing (FIT) Kohorte lieferten, neben unseren Berechnungen, die ersten Ergebnisse zur Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom-Inzidenz in Frauen (36). In der FIT Kohorte wurde die Exposition mittels maximalem Fitnessstest gemessen, allerdings wurden keine Sublokalisierungen des Kolorektalkarzinoms betrachtet. In der gesamten Kohorte zeigte sich eine HR von 0,39 (95 % KI 0,23-0,66) für hohe versus niedrige kardiorespiratorische Fitness. Analog zu unseren Ergebnissen zeigte sich die statistisch signifikante inverse Assoziation zwischen kardiorespiratorischer Fitness und dem Risiko von Kolorektalkarzinomen nur in der Subgruppe der Männer, nicht aber in der Subgruppe der Frauen (36).

In den meisten der bisherigen Studien wurde die kardiorespiratorische Fitness mittels maximaler Ergometrie (Fahrrad oder Laufband) bestimmt (33-36, 96). Ein Vorteil der Messung bis zur subjektiven Ausbelastung besteht darin, dass die maximale Sauerstoffaufnahme valider bestimmt, oder bei paralleler Spirometrie sogar direkt ermittelt werden kann (103). Nachteile dieser Methode sind die größere Belastung und Gefährdung der Teilnehmenden, welche gerade in groß angelegten bevölkerungsbasierten Studien problematisch sind, sowie der starke Einfluss der individuellen Motivation der Teilnehmenden (103, 107). Es gibt Methoden, um die maximale Sauerstoffaufnahme beziehungsweise die maximalen METs aus submaximalen Belastungstests zu schätzen, wie beispielsweise in vorangegangenen UK Biobank Analysen geschehen (97, 165). Allerdings ist dieses Vorgehen gerade bei älteren Personen fehleranfällig, weshalb in den vorliegenden Analysen die $PWC_{75\%}$ als Maß für die kardiorespiratorische Fitness verwendet wurde (106, 107). Zudem ist hier die relative Einordnung in Teilnehmende mit höherer oder niedrigerer kardiorespiratorischer Fitness relevanter als die objektive Beurteilung des Fitnessstatus.

Die beobachtete inverse Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom-Inzidenz ist biologisch plausibel. Zwar sind die exakten Mechanismen noch nicht abschließend geklärt, aber es wird postuliert, dass kardiorespiratorische Fitness die Insulinsensitivität verbessert und den Nüchtern-Insulin-Spiegel senkt (24). Diese Relationen werden teilweise durch Veränderungen der Körperzusammensetzung, vor allem eine Erhöhung

des Anteils an fettfreier Masse, vermittelt (166). Insulinresistenz und Hyperinsulinämie sind möglicherweise Risikofaktoren für das Kolorektalkarzinom (138). Des Weiteren zeigte sich eine inverse Assoziation zwischen maximaler Sauerstoffaufnahme, herbeigeführt durch moderates Fitnesstraining, und zirkulierenden Konzentrationen der Entzündungsmarker IL-6 und TNF- α im Blut, sowie eine positive Assoziation mit IL-10 Konzentrationen (28). Dies wirkt insgesamt entzündungshemmend und so antikanzerogen (28). Eine weitere Hypothese besagt, dass kardiorespiratorische Fitness die mikrobielle Diversität im Dickdarm und damit die Butyrat Produktion erhöht, was mit verringertem Risiko für Kolorektalkarzinom assoziiert ist (167, 168). Zudem soll die kardiorespiratorische Fitness die Regulierung des Zellzyklus und die DNA Reparatur verbessern (169, 170) und durch Hochregulierung antioxidativer Enzyme die Resilienz gegen oxidativen Stress erhöhen (26).

Die inverse Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom war am stärksten im proximalen Teil des Kolons ausgeprägt. Während die Konfidenzintervalle in den Hauptanalysen, aufgrund zu geringer Teststärke, den Nullwert 1,0 enthielten, waren die Assoziationen in der Sensitivitätsanalyse mit imputierten Expositionswerten statistisch signifikant in der gesamten Stichprobe und bei den Männern. Auch in einer vorangegangenen Studie, welche anatomische Sublokalisationen des Kolonkarzinoms untersuchte, zeigte sich die inverse Assoziation ausschließlich im proximalen Kolon (34). Die Gründe hierfür sind spekulativ, aber könnten auf molekulare Heterogenität zwischen den Sublokalisationen zurückzuführen sein. Bei proximalen Kolonkarzinomen zeigte sich häufiger eine Mikrosatelliteninstabilität (Mutationsanfälligkeit), hervorgerufen durch defekte DNA Reparatur, als bei den distalen Kolonkarzinomen (171). Falls die Hypothese stimmt, dass Fitness die DNA Reparatur verbessert, könnte sie vor allem bei proximalen Kolontumoren ihre protektive Wirkung entfalten (170).

Zum Zeitpunkt der Analysen gab es, nach unserem besten Wissen, keine weitere Studie, die den Zusammenhang von kardiorespiratorischer Fitness und Kolorektalkrebsinzidenz in einer Stichprobe von Frauen untersuchte. Mittlerweile wurde die fehlende Assoziation bei den Frauen von einer weiteren Kohortenstudie bestätigt (36). Eine mögliche biologische Erklärung könnte die geschlechtsspezifische Wirkung von Hormonen auf die Karzinogenese sein. So zeigte sich, dass bei postmenopausalen Frauen sowohl exogene als auch endogene Östrogene das Kolorektalkarzinom Risiko verringerten (144, 172). Allerdings ist die Aromataseaktivität und damit die Östrogenproduktion im Fettgewebe fitter postmenopausaler Frauen reduziert (173). Daher könnten die, durch die Fitness herbeigeführten positiven Effekte (z.B.

Verbesserung der Insulinsensitivität (24)), durch den fehlenden protektiven Einfluss von Östrogen ausgeglichen werden. In unserer Stichprobe waren 72 % der Teilnehmerinnen älter als 50 Jahre. Da die Menopause in Großbritannien durchschnittlich im Alter von 51 Jahren einsetzt (174), ist davon auszugehen, dass unsere Analytestichprobe einen großen Anteil an postmenopausalen Frauen beinhaltet. Eine weitere Begründung könnte die geringere Variation der kardiorespiratorischen Fitness, und die damit einhergehende schlechtere Unterscheidbarkeit, bei den Frauen sein. Während die $PWC_{75\%}$ bei den Männern von der 25. bis zur 75. Perzentile 0,79 W/kg umfasste, betrug diese Spannbreite bei den Frauen nur 0,57 W/kg.

Die zahlreichen Sensitivitätsanalysen lieferten weitere Ergebnisse und gaben Rückschlüsse auf die Robustheit der Analysen. Der E-value, welcher angibt wie stark ein unberücksichtigter Störfaktor mit Exposition und Endpunkt assoziiert sein muss um das beobachtete Ergebnis zu erklären, sollte immer im Kontext der jeweiligen Studie interpretiert werden (162, 163). In den vorliegenden Daten zeigten etablierte starke Risikofaktoren von Kolorektalkarzinom, wie männliches Geschlecht oder Familienhistorie von Darmkrebs, HRs von 1,6 beziehungsweise 1,4. Laut des berechneten E-values müsste ein ungemessener Störfaktor mit einem Relativen Risiko von mindestens 1,9 mit Exposition und Endpunkt assoziiert sein. Es ist allerdings unwahrscheinlich, dass ein in den vorliegenden Analysen nicht berücksichtigter Risikofaktor einen größeren Einfluss auf das Kolorektalkarzinomrisiko hat als diese sehr starken Risikofaktoren. Im Zuge der Kovariablenselektion ist es wichtig Störgrößen von Intermediärvariablen zu unterscheiden. Letztere liegen auf dem Pfad zwischen Exposition und Endpunkt und vermitteln so einen Teil des beobachteten Effekts, weshalb man nicht für sie adjustieren sollte. Während körperliche Aktivität einen Einfluss auf das Gewicht und die Körperzusammensetzung hat (175), wird die kardiorespiratorische Fitness selbst von der Adipositas beeinflusst (147). Deshalb stellt der Taillen-Hüft-Quotient höchstwahrscheinlich keinen Intermediärfaktor, sondern eine Störgröße für die Assoziationen zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom dar und wurde deshalb in allen Hauptmodellen als Kovariable aufgenommen. Im Gegensatz hierzu ist die kardiorespiratorische Fitness invers mit Typ 2 Diabetes assoziiert (150) was wiederum einen Risikofaktor für Krebs darstellt (138). Für den Intermediärfaktor Diabetes wurde deshalb nur in einer Sensitivitätsanalyse adjustiert, mit vergleichbaren Ergebnissen. Körperliche Aktivität ist eine wichtige Determinante von kardiorespiratorischer Fitness (2, 176, 177) und ist invers mit dem Risiko für Kolorektalkarzinom assoziiert (93-95). Dennoch wurde nur in einer

Sensitivitätsanalyse für körperliche Aktivität adjustiert, da wir in den Hauptmodellen den Gesamteffekt der kardiorespiratorischen Fitness inklusive des durch körperliche Aktivität modifizierbaren Anteils betrachten wollten. Die Ergebnisse waren in beiden Modellen quasi identisch, was möglicherweise darauf beruht, dass die objektive Messung der kardiorespiratorischen Fitness valider und aussagekräftiger war, als die subjektive und daher fehleranfällige Messung der körperlichen Aktivität (21). Zudem könnte dies ein Hinweis dafür sein, dass der durch Genetik, Körpergewicht, Rauchverhalten, oder Krankheitsstatus beeinflusste Teil der Fitness für den antikanzerogenen Effekt entscheidender ist, als der durch körperliche Aktivität beeinflusste Teil. Auch die nach körperlicher Aktivität stratifizierten Analysen ergaben, möglicherweise aufgrund der verminderten Teststärke, keine aussagekräftigen Assoziationen zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom.

Die zentrale Stärke der vorliegenden Analysen ist die objektive Messung der kardiorespiratorischen Fitness mittels anerkannter Methoden. Dies führte zu einer validen und präzisen Bestimmung der Exposition und verhinderte Misklassifikation. Auch die Verknüpfung der Studiendaten mit Daten aus nationalen Krebsregistern und anderen Routinedaten gewährleistet eine valide und vollständige Erfassung des Krankheitsendpunktes. Zudem konnten durch diese detaillierten Informationen auch Analysen zu anatomischen Sublokalisationen durchgeführt werden. Die relevanten Kovariablen wurden a priori basierend auf vorangegangenen Studien und Expertenwissen ausgewählt. Im Vergleich zu vielen bisherigen Analysen konnten die Ergebnisse so für eine Vielzahl an wichtigen Störgrößen, wie beispielsweise Darmkrebsscreening, Familienhistorie von Darmkrebs, Hormonersatztherapie, Größe oder Verzehr von rotem und verarbeitetem Fleisch, adjustiert werden. Um die Gefahr für Selektionsbias zu minimieren und die Teststärke der Analysen zu erhöhen, wurden alle fehlenden Kovariablen multipel imputiert. Die Annahmen und die Robustheit der Ergebnisse wurden zudem in zahlreichen Sensitivitätsanalysen überprüft. Dennoch liegen auch potenzielle Limitationen vor. Die relativ kurze Nachbeobachtungszeit und die damit einhergehende geringe Zahl an Fällen verringerte die Präzision der Ergebnisse, vor allem in den Subanalysen. Bei Studien mit zu geringer Teststärke ist die Gefahr, dass statistisch signifikante Ergebnisse nicht die wahren Zusammenhänge widerspiegeln größer (178). Allerdings stimmen die vorliegenden Ergebnisse sehr gut mit den Ergebnissen einer kürzlich veröffentlichten Metaanalyse aus vier verschiedenen Kohortenstudien mit unterschiedlicher Methodik überein (38), was dafür spricht, dass die hier berichteten Schätzer nicht übermäßig durch die getroffenen Annahmen

herbeigeführt wurden (178, 179). Eine kurze Nachbeobachtungszeit erhöht zudem das Risiko für umgekehrte Kausalität. Die Sensitivitätsanalysen zeigten jedoch, dass der Ausschluss von allen Teilnehmenden mit als „schlecht“ bewertetem Gesundheitszustand oder von Teilnehmenden, welche eine Kolorektalkarzinom-Diagnose innerhalb von 12 Monaten nach der Basiserhebung erhielten, nicht zu maßgeblich anderen Schlussfolgerungen führte. Die UK Biobank steht teilweise in der Kritik, da sie eine sehr geringe Teilnahmequote (ca. 5,5 %) aufwies. Zusätzlich wurde der Ergometer-Fitnesstest nur in einer Teilstichprobe angeboten und nur bei Personen mit geringem Risiko als inkrementell ansteigender Belastungstest durchgeführt. Nach Imputation aller fehlenden Expositionswerte der Ergometer Stichprobe blieben die inversen Assoziationen zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom allerdings bestehen. Auch kam eine aktuelle Studie zu dem Schluss, dass die Exposition-Endpunkt Assoziationen aus der UK Biobank vergleichbar mit den Ergebnissen repräsentativer Studien waren (180). Insgesamt ist anzumerken, dass das primäre Ziel der Analysen nicht die Übertragbarkeit auf bestimmte Populationen, sondern die Identifizierung ethiologischer Pfade darstellt. Die Latenzperiode zwischen kardiorespiratorischer Fitness und kolorektalem Karzinom beträgt höchstwahrscheinlich viele Jahre bis Jahrzehnte. Die Analysen beruhen damit auf der Annahme, dass sich die kardiorespiratorische Fitness über die Zeit nicht verändert. Da die kardiorespiratorische Fitness auch eine starke genetische Komponente aufweist, und die inversen Assoziationen auch nach Adjustierung für körperliche Aktivität bestehen blieben, ist das Fehlerpotential dieser Annahme möglicherweise vernachlässigbar. Die lange Latenzperiode ist allerdings auch ein Grund, warum die untersuchten Assoziationen nicht mittels randomisierter klinischer Trials, dem Goldstandard für die Feststellung von Kausalität, betrachtet werden können. Beobachtungsstudien, wie die hier durchgeführte sind derzeit notwendig, um die Assoziation zwischen kardiorespiratorischer Fitness und Kolorektalkarzinom aufzuklären. Weitere Studien mit längerer Nachbeobachtungszeit und größerer Teststärke werden weiterhin benötigt um die bislang spärlichen Ergebnisse zu Krebssubtypen oder stratifiziert nach Geschlecht zu untermauern. Möglicherweise wird auch eine Aktualisierung der hier vorliegenden Ergebnisse mit zukünftigen UK Biobank Daten diese Anforderungen liefern.

4.5 Schlussfolgerung

Die Analysen deuten auf eine inverse Assoziation zwischen kardiorespiratorischer Fitness und dem Risiko für Kolorektalkarzinom beziehungsweise Kolonkarzinom hin. Vor allem bei der Subgruppe der Männer und der Sublokalisierung proximales Kolonkarzinom schienen erhöhte

Level an kardiorespiratorischer Fitness einen protektiven Einfluss auf die Krebsinzidenz zu haben. Kardiorespiratorische Fitness stellt einen, durch habituelle körperliche Aktivität modifizierbaren, Gesundheitsfaktor dar, welcher mittels Interventionsprogrammen gefördert werden sollte, um so das individuelle und gesamtgesellschaftliche Krebsrisiko zu minimieren. Zudem ist die Ergometrie, beispielsweise in der Allgemeinarztpraxis, eine gute Messmethode, um Hochrisikopopulationen zu identifizieren, welchen in einem zweiten Schritt bevorzugt ein Darmkrebsscreening angeboten werden kann.

5 Assoziationen zwischen körperlicher Aktivität, Greifkraft, sedentärem Verhalten und der Inzidenz von malignem Melanom in der UK Biobank

5.1 Einleitung

Das maligne Melanom wurde im Jahr 2018 weltweit knapp 288.000 Mal neu diagnostiziert (44). Während dies global nur 1,6 % aller inzidenten Krebsarten ausmachte (44), schwankt die Inzidenz zwischen verschiedenen Populationen stark (181). Regionen mit hoher UV-Belastung und vorwiegend hellhäutiger Bevölkerung (z.B. Australien oder Neuseeland) haben die höchsten Inzidenzen (181). Zudem steigt die Inzidenz des malignen Melanoms weltweit an (182, 183), was möglicherweise an der Zunahme von Fernreisen und der Nutzung von Bräunungsgeräten liegt (181). Die Identifizierung modifizierbarer protektiver Faktoren ist damit ein wichtiges Ziel der öffentlichen Gesundheitsförderung.

Obwohl ein inverser Zusammenhang zwischen körperlicher Aktivität und Krebsinzidenz aufgrund bereits beschriebener Mechanismen biologisch plausibel ist, zeigten bisherige prospektive Kohortenstudien einen positiven Zusammenhang zwischen körperlicher Aktivität und dem Risiko für malignes Melanom (184). Möglicherweise ist diese Beobachtung auf den verzerrenden Effekt der UV-Exposition, welche bei im Freien durchgeführter körperlicher Aktivität auftritt, zurückzuführen (184, 185). Hierfür spricht, dass einige Fall-Kontroll-Studien, welche auf Individuallevel für UV-Exposition oder -Sensitivität kontrolliert hatten, eher auf eine inverse bzw. Nullasoziation zwischen körperlicher Aktivität und der Inzidenz von malignem Melanom hindeuteten (184). Handgreifkraft ist ein objektiv gemessenes Surrogat für die Gesamtmuskelkraft des Körpers, einer Komponente der körperlichen Fitness (8, 186). Während es Hinweise auf eine inverse Assoziation zwischen Greifkraft und dem Risiko von Kolorektal-, Brust- oder Lungenkrebs gibt (39), liegen bislang keine Analysen zur Assoziation zwischen Greifkraft und malignem Melanom vor. Allerdings zeigte sich in einer Studie eine statistisch nicht signifikante positive Assoziation zwischen kardiorespiratorischer Fitness und dem Risiko an malignem Melanom zu erkranken (34). Die Assoziation zwischen sedentärem Verhalten und dem Risiko an malignem Melanom zu erkranken wurde bislang nicht untersucht, allerdings gibt es Hinweise auf einen positiven Zusammenhang mit anderen Krebsentitäten (z.B. Kolon-, Brust- oder Endometriumkarzinom) (40).

Um diese Forschungslücken zu füllen und einen möglichst validen Schätzer für die Assoziation zwischen körperlicher Aktivität und Inzidenz an malignem Melanom zu erhalten, wurden

prospektive Daten der UK Biobank verwendet. Informationen zu UV-Exposition und -Sensitivität auf Individuallevel, sowie weitere relevante, a priori bestimmte, Störgrößen wurden zur Kontrolle in die Analysen aufgenommen.

5.2 Methoden

5.2.1 Studiendesign und Stichprobe

Die Fragestellung wurde ebenfalls mit Daten der populationsbasierten Kohortenstudie UK Biobank bearbeitet (Antrag Nr. 24091). Ethische Beurteilung, Studiendesign und Studiendurchführung wurden ausführlich in Abschnitt 4.2.1 beschrieben. In der Basisuntersuchung wurden zwischen 2006 und 2010 rund 500.000 Männer und Frauen in 22 Studienzentren in Großbritannien untersucht (101). Für die Analysen wurden 36.941 Teilnehmende mit mindestens einem prävalenten malignen Karzinom zur Basisuntersuchung (außer nichtmelanozytärer Hautkrebs) exkludiert. Nach Ausschluss aller Teilnehmenden mit fehlenden oder invaliden Angaben zu den Expositionen körperliche Aktivität, Greifkraft, oder sedentäres Verhalten sowie Teilnehmenden mit erheblichen Beeinträchtigungen in der Bewegungsfähigkeit verblieben 350.512 Personen in der Analysestichprobe (siehe Abbildung 14).

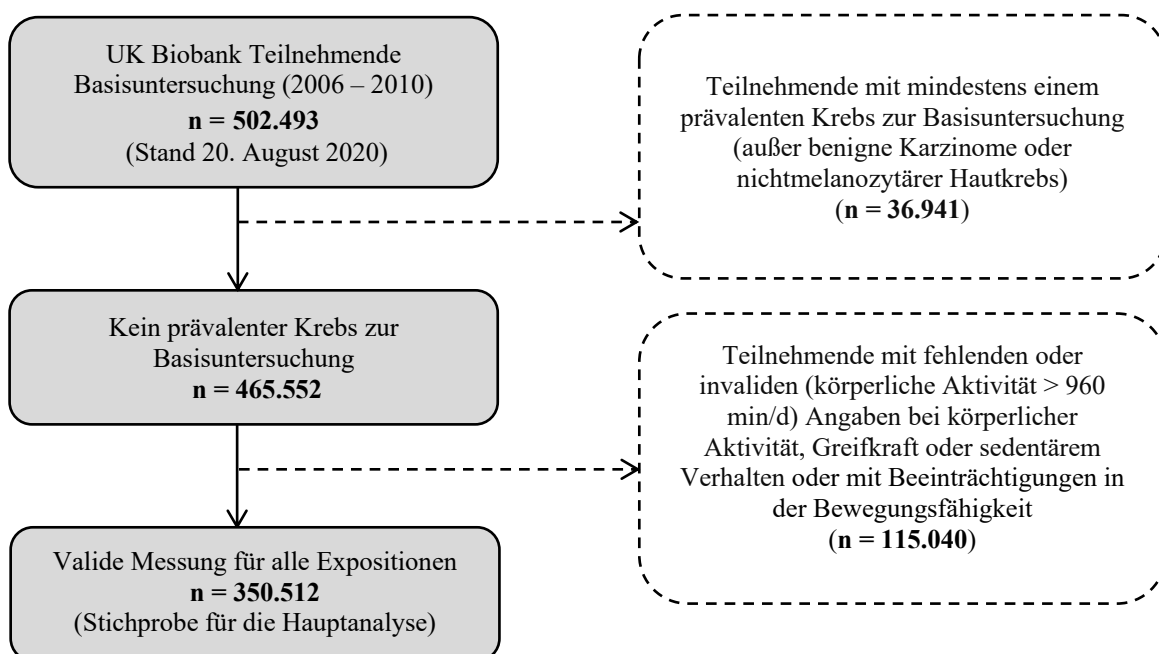


Abbildung 14: Flussdiagramm zum Ein- und Ausschluss der UK Biobank Teilnehmenden für die Analyse der Assoziation zwischen körperlicher Aktivität, Greifkraft, sedentärem Verhalten und malignem Melanom

5.2.2 Erhebung der subjektiven körperlichen Aktivität

Die subjektive körperliche Aktivität wurde zur Basisuntersuchung mit einer abgewandelten Kurzform des International Physical Activity Questionnaires (IPAQ-SF) im Rahmen der Touchscreen-Befragung bei allen Teilnehmenden erhoben. Die Fragen bezogen sich auf die körperlich aktive Zeit der letzten vier Wochen und unterschieden nicht zwischen Arbeit, Hausarbeit oder Freizeit. Explizit gefragt wurden Häufigkeit pro typischer Woche und Dauer pro Tag (mindestens 10 Minuten) von Gehen, moderater und intensiver körperlicher Aktivität. Als Hinweis waren Beispielaktivitäten (z.B. Tragen leichter Gegenstände, schnelles Radfahren) zu jeder Intensität sowie weiche und harte Plausibilitätskontrollen hinterlegt (Tabelle S 2). In vorangegangenen Validierungsstudien zeigte der IPAQ-SF gute Reliabilität (gepoolter Korrelationskoeffizient 0,76; 95% KI 0,73–0,77) und ausreichende Validität gegen Akzelerometrie (gepoolter Korrelationskoeffizient 0,30; 95% KI 0,23–0,36) (15).

Die Aufbereitung der Aktivitäten erfolgte nach dem IPAQ Scoring Protokoll (187). Hierbei wurden zunächst alle Teilnehmenden, deren Summe von Gehen, moderater und intensiver körperlicher Aktivität 960 Minuten pro Tag überstieg aufgrund unplausibel hoher Werte von den Analysen ausgeschlossen. Diesem Vorgehen liegt die Grundannahme zugrunde, dass die Teilnehmenden circa 8 Stunden pro Tag schlafen (187). Auch Teilnehmende mit fehlenden Werten oder der Angabe „weiß nicht“, „ziehe es vor, nicht zu antworten“, oder „nicht in der Lage zu gehen“ wurden von allen Analysen ausgeschlossen. Als nächstes wurden Zeitangaben unter 10 Minuten pro Tag auf 0 umcodiert (187). Dies reflektiert die mittlerweile korrigierte Annahme, dass erst ab 10 Minuten körperlicher Aktivität gesundheitsrelevante Effekte einsetzen (3). Trotz neuer wissenschaftlicher Erkenntnisse wurde dieser Schritt durchgeführt, um Vergleichbarkeit mit anderen Studien zu gewährleisten, und um die Fragestellung („mindestens 10 Minuten“) zu berücksichtigen. Als dritter Schritt wurden die Angaben zur Dauer von Gehen, moderater und intensiver körperlicher Aktivität jeweils auf 180 Minuten pro Tag winsorisiert, um die Verteilung der Aktivitätslevel zu normalisieren (187). Nach der Bereinigung der einzelnen Variablen wurden MET-Minuten pro Woche gebildet, indem die Anzahl an Tagen mit der Dauer pro Tag und einem MET-Wert multipliziert wurde. Hierfür wurden aus dem Kompendiums von Ainsworth abgeleitete durchschnittliche MET-Werte für die Intensitätskategorien Gehen (3,3 METs), moderate (4,0 METs) und intensive (8,0 METs) Aktivität verwendet (69, 187). Die Summe der MET-Minuten pro Woche für diese drei Intensitätskategorien ergibt die gesamte körperliche Aktivität pro Woche (MET-Minuten pro Woche, Abbildung 15) (187).

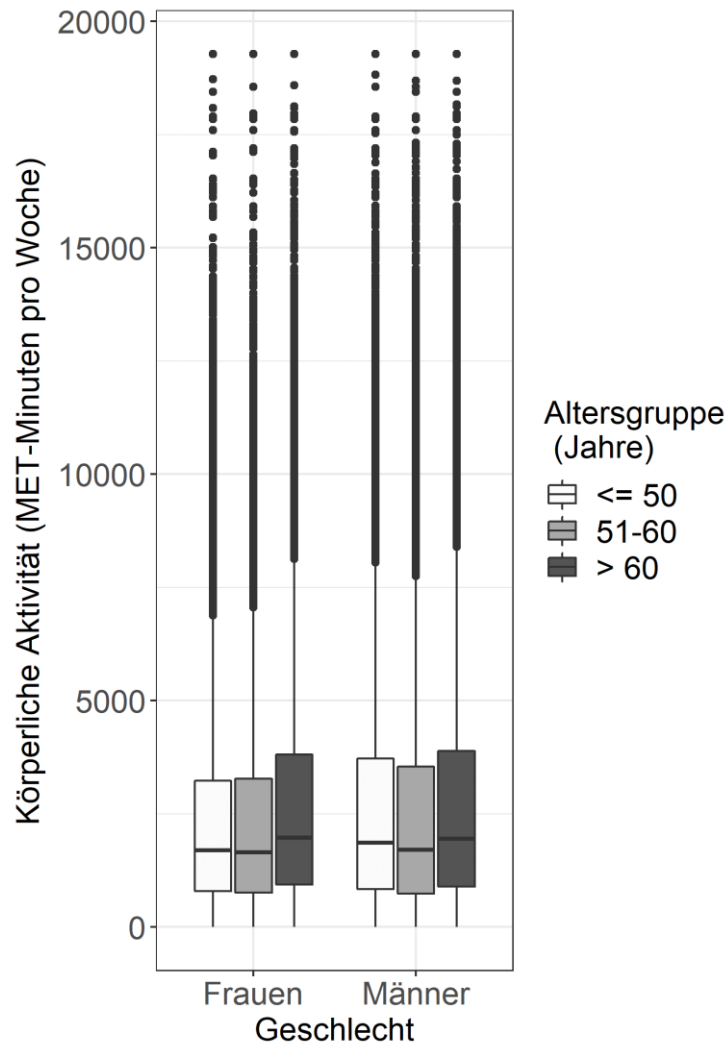


Abbildung 15: MET-Minuten pro Woche (IPAQ) stratifiziert nach Alter und Geschlecht

5.2.3 Messung der Greifkraft

Zur Basisuntersuchung war die Greifkraftmessung für alle UK Biobank Teilnehmenden vorgesehen. Hierfür wurde das Jamar J00105 hydraulic hand dynamometer (Lafayette Instrument USA) verwendet, welches die Greifkraft in kg misst (188). Den Teilnehmenden wurde der Untersuchungsablauf erklärt und sie wurden informiert, dass das Handgreifkraftgerät statisch ist und beim Druck nicht nachgibt. Zudem wurden die Teilnehmenden angewiesen aufrecht zu sitzen, den Ellbogen des Arms, der das Gerät fasst, um 90° anzuwinkeln und für 3 Sekunden das Gerät so fest wie möglich zu drücken. Die Durchführung umfasste jeweils eine Messung der rechten und anschließend der linken Hand. Die Untersuchenden motivierten die Teilnehmenden während der Messung (188). Als Maß für die Greifkraft wurde der Mittelwert aus der Messung der linken und der rechten Hand gebildet (Abbildung 16).

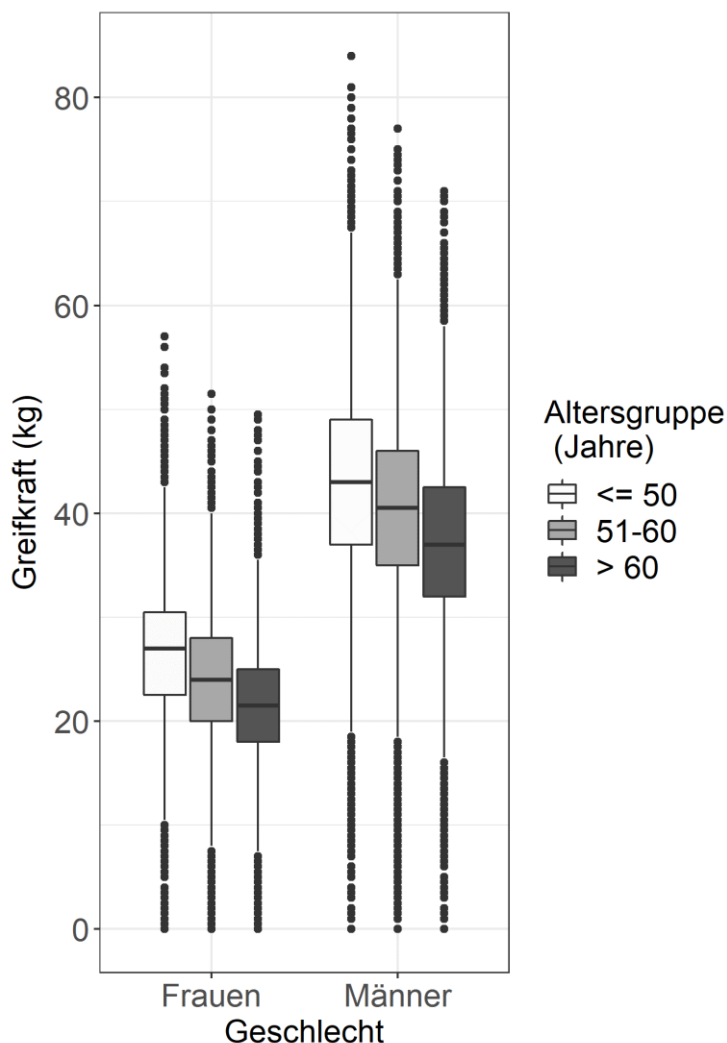


Abbildung 16: Greifkraft stratifiziert nach Alter und Geschlecht

5.2.4 Erhebung des subjektiven sedentären Verhaltens

Alle Teilnehmende wurden in der Touchscreen-Befragung der Basisuntersuchung mittels drei spezifischer Fragen zu ihrem sedentären Verhalten an einem typischen Tag befragt. Erhoben wurde die mit Fernsehschauen, Computernutzung in der Freizeit und sitzender Fortbewegung verbrachte Zeit (Tabelle S 3). Ein Review zur Messung von sedentärem Verhalten fand heraus, dass Fernsehschauen, Bildschirmnutzung und sitzende Fortbewegung einen schädlicheren Einfluss auf die Gesundheit haben könnten, als die gesamte Sitzzeit (189). Daher wurden die drei erfragten sedentären Verhaltensweisen addiert, wie auch in früheren UK Biobank Analysen (39). Die Angabe „weniger als eine Stunde pro Tag“ ging mit 0,5 Stunden in die Berechnung ein. Wenn die Summe der drei Aktivitäten 24 Stunden pro Tag überstieg, wurde diese auf 24 Stunden winsorisiert (Abbildung 17).

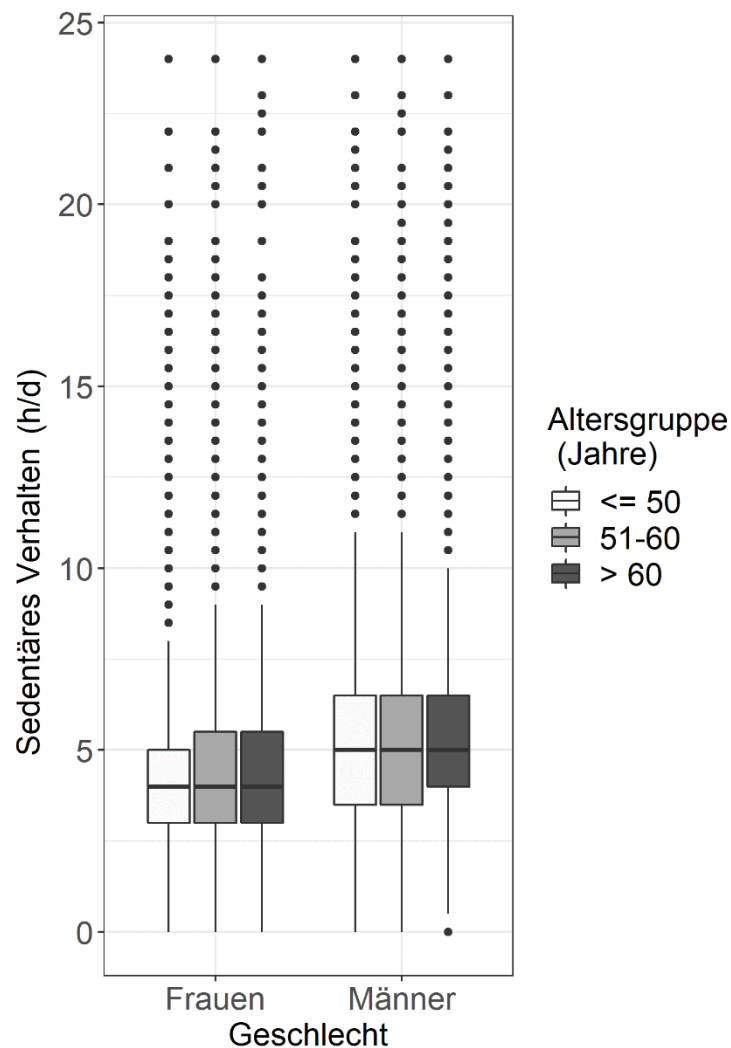


Abbildung 17: Sedentäres Verhalten stratifiziert nach Alter und Geschlecht

5.2.5 Nachbeobachtung und Verknüpfung mit Routinedaten

Wie bereits in Abschnitt 4.2.3 beschrieben, wurden Informationen zu Krebserkrankungen und Mortalität durch Verknüpfung mit Nationalen Krebs- und Mortalitätsregistern gewonnen (117, 118).

Für die Analysen zur Assoziation zwischen körperlicher Aktivität, Greifkraft, sedentärem Verhalten und inzidentem malignem Melanom lagen vollständige Krebsdaten bis zum 31. März 2016 (England/Wales) beziehungsweise bis zum 31. Oktober 2015 (Schottland) vor. Die Diagnose malignes Melanom wurde durch die Kombination von ICD-10 Code C43 und Histologie 8720-8790 identifiziert, basierend auf dem Vorgehen einer großen früheren Studie (94). Die dritte Ziffer des ICD-10 Codes wurde zur Klassifikation chronisch und sporadisch UV-exponierter Hautlokalisationen verwendet (Tabelle 4) (190).

Tabelle 4: ICD-10 Klassifizierung des malignen Melanoms und Einteilung in chronisch und sporadisch UV-exponierte Hautlokalisationen

ICD10	Beschreibung	UV Exposition
C43	Bösartiges Melanom der Haut	
C43.0	Bösartiges Melanom der Lippe	chronisch
C43.1	Bösartiges Melanom des Augenlides, einschließlich Kanthus	chronisch
C43.2	Bösartiges Melanom des Ohres und des äußeren Gehörganges	chronisch
C43.3	Bösartiges Melanom sonstiger und nicht näher bezeichneter Teile des Gesichtes	chronisch
C43.4	Bösartiges Melanom der behaarten Kopfhaut und des Halses	chronisch
C43.5	Bösartiges Melanom des Rumpfes	sporadisch
C43.6	Bösartiges Melanom der oberen Extremität, einschließlich Schulter	chronisch
C43.7	Bösartiges Melanom der unteren Extremität, einschließlich Hüfte	sporadisch
C43.8	Bösartiges Melanom der Haut, mehrere Teilbereiche überlappend	
C43.9	Bösartiges Melanom der Haut, nicht näher bezeichnet	

5.2.6 Kovariablen

Relevante Kovariablen wurden a priori mittels gerichtetem, azyklischem Graphen (DAG) und dem Disjunctive Cause Criterion identifiziert (123, 125). In Abschnitt 4.2.4 findet sich eine genauere Beschreibung dieser Methoden.

Für die Assoziation zwischen körperlicher Aktivität, Greifkraft, sedentärem Verhalten und malignem Melanom wurden die Basis-Analysen für Alter (kontinuierlich (Jahre); als Zeiteinheit), Geschlecht (Frauen/Männer) und Studienregion (England, Schottland, Wales; Stratifizierung des Baseline-Hazards) adjustiert. In den voll adjustierten Modellen, wurde zusätzlich für Bildung (Universitäts-, College-Abschluss/A-Level, AS-Level, NVQ, HND, HNC, äquivalent oder anderer professioneller Abschluss/O-Level, CSEs, äquivalent/Keiner der oben genannten Abschlüsse), Rauchverhalten (Packungsjahre), Alkoholkonsum (g/d; (191)), Körpergröße (cm), Kaffeekonsum (Tassen pro Tag), Nutzung von Solarien (ja/nein), Sonnenbrände in der Kindheit (definiert als vor dem 15. Lebensjahr aufgetretener schmerzhafter Sonnenbrand für mindestens zwei Tage oder Sonnenbrand, der Blasenbildung verursachte; ja/nein), Im Freien verbrachte Zeit im Sommer (Stunden pro typischem Tag) und Saisonalität (Sinus-Kosinus-Terme für den Monat der Untersuchung im Studienzentrum). Die Modelle wurden gegenseitig für körperliche Aktivität (MET-Minuten pro Woche), Greifkraft

(kg) und sedentäres Verhalten (h/d) adjustiert. Aus den Variablen Hautbräunung (stark gebräunt/mäßig gebräunt/leicht oder gelegentlich gebräunt/nie gebräunt, nur Verbrennungen), Hautfarbe ohne Bräunung (schwarz, braun/hell- oder dunkeloliv/hell/sehr hell) und natürliche Haarfarbe vor dem Ergrauen (andere, schwarz, dunkelbraun/hellbraun/blond, rot) wurde eine Indikatorvariable für UV-Sensitivität erstellt (Tabelle S 4 und Tabelle S 5). Die Plausibilität dieses UV-Sensitivitätsindex wurde überprüft, indem die durchschnittliche Anzahl an Sonnenbränden in der Kindheit und die Inzidenz von malignem Melanom (pro 10.000 Personenjahren) in jedem Stratum des Index berechnet wurde. Die voll adjustierten Modelle wurden entweder für diesen UV-Sensitivitätsindex adjustiert oder danach stratifiziert. Eine zusätzliche Analyse in einer Stichprobe an Teilnehmenden, welche aktuell (selbstständig) erwerbstätig waren, wurde zusätzlich für UV-Exposition am Arbeitsplatz (ja/nein) adjustiert. Hierbei wurden die Ergebnisse aus früheren Studien verwendet, um Berufe mit schädlicher natürlicher oder künstlicher UV-Exposition zu identifizieren (Tabelle S 6) (192, 193).

Im Anhang befindet sich ein DAG, welches die kausalen Zusammenhänge der Störgrößen mit Expositionen und Endpunkt darstellt (Abbildung S 5 (124)). Erste Hinweise deuten auf einen protektiven Effekt von Kaffeekonsum auf das maligne Melanom Risiko bei Frauen hin (194). Höhere Bildungslevel (195) und Alkoholkonsum (196) zeigten in früheren Studien eine positive Assoziation mit dem Risiko an malignem Melanom zu erkranken. Während die Studien zu den Auswirkungen des Rauchens auf die Inzidenz von malignem Melanom heterogene Ergebnisse lieferten, beinhaltet Tabakrauch mehr als 60 Karzinogene, welche die Tumorentstehung begünstigen (197). Natürliche und künstliche UV-Exposition, Solariennutzung und Sonnenbrände in der Kindheit sind nachweislich starke Risikofaktoren für das maligne Melanom (198-201). Die Modelle wurden stratifiziert nach Studienregion berechnet, da diese als Surrogat für die bodennahe solare UV-Belastung dient. Die Körpergröße im Erwachsenenalter erwies sich als Marker für genetische, hormonelle, umwelt- und ernährungsbedingte Faktoren, welche Risikofaktoren für das maligne Melanom darstellen (196, 202). In früheren Studien zeigte sich außerdem eine starke Assoziation zwischen Körpergröße und Handgreifkraft (203-205). Ebenso sind körperliche Aktivität, sedentäres Verhalten und Fitness mit der im Freien verbrachten Zeit assoziiert (206). Da die Teilnehmenden über ihre körperliche Aktivität der letzten vier Wochen befragt wurden und das Erinnerungsvermögen, zur im Sommer draußen verbrachten Zeit, je nach Befragungszeitpunkt unterschiedlich sein kann, könnte die Saisonalität diese Variablen beeinflussen. Bedenken bezüglich UV-vermittelter Hautschädigung könnten als Barriere für Aktivitäten im Freien wirken (207). Auch

haben Personen mit rötlichen Haaren, heller Haut und hellen Augen, Sommersprossen, erhöhter Sonnensensitivität und wenig Hautbräunung ein etwa 50 % höheres Risiko an malignem Melanom zu erkranken als Personen ohne diese Merkmale (182). Da Informationen auf Individuallevel zu Hautfarbe, Haarfarbe und Hautbräunung vorlagen, welche aussagekräftiger in Hinblick auf Sonnensensitivität sind, wurden die Modelle nicht für Ethnizität adjustiert. Ebenfalls wurde nicht für die Verwendung von Sonnen- oder UV-Schutz adjustiert, da diese Information in der Vergangenheit zu irreführenden Schlussfolgerungen geführt hatte. So zeigte sich, dass die Verwendung von Sonnencreme bei sporadischer UV-Exposition mit einem erhöhten Risiko für malignes Melanom assoziiert war, möglicherweise, da die Personen hierdurch längerer Zeit schädlicher UV-Belastung ausgesetzt waren (208). Da eine vorangegangene Mendelsche Randomisierungsstudie keine Assoziation zwischen BMI und malignem Melanom fand, wurden die Analysen nicht für Körperzusammensetzung adjustiert (202).

5.2.7 Statistische Analysen

Altersadjustierte Eigenschaften der Studienpopulation zur Basisuntersuchung, berechnet mittels direkter Standardisierung auf die Altersverteilung der Kohorte, wurden nach geschlechtsspezifischen Quartilen von körperlicher Aktivität, Greifkraft und sedentärem Verhalten dargestellt.

Kovariablen mit fehlenden Werten wurden mittels chained equations multipel imputiert (zehn Datensätze mit jeweils fünf Iterationen) um die statistische Teststärke zu erhöhen (155). Als Imputationsmodell wurde für kontinuierliche Variablen predictive mean matching, für binäre Variablen logistische Regression, für ordinale Variablen ein ordered logit Model und für nominale Variablen ein multinomial logit Model verwendet (155). Konvergenz und Plausibilität der Imputation wurden mithilfe diagnostischer Plots überprüft.

Cox proportional hazards Regressionsmodelle wurden verwendet um die Assoziationen zwischen körperlicher Aktivität (MET-Minuten pro Woche), Greifkraft (kg) sowie sedentärem Verhalten (h/d) mit der Inzidenz von malignem Melanom zu berechnen. Alter wurde als zugrundeliegende Zeitachse verwendet (Alter zur Basisuntersuchung als Startpunkt; Alter zum Zeitpunkt der vollständigen Verknüpfung, zur Krebsdiagnose oder zum Todeszeitpunkt, je nachdem welches Ereignis zuerst eintrat, als Endpunkt (158, 159). Nicht-Linearität der Assoziation zwischen Exposition und Endpunkt wurde mittels restricted cubic splines überprüft (160). Hierbei wurden für körperliche Aktivität und Greifkraft vier Knoten (an der 5.,

35., 65. und 95. Quantile der jeweiligen Verteilung) und für sedentäres Verhalten, aufgrund der geringeren Anzahl an Ausprägungen, drei Knoten (an der 10., 50. und 95. Quantile der Verteilung) gesetzt (160). Da die körperliche Aktivität (MET-Minuten pro Woche) viele Ausprägungen am Nullpunkt aufwies und dies bei restricted cubic splines problematisch sein kann (209), wurde für diese Modelle eine Dummy Variable (körperliche Aktivität ja/nein) inkludiert (210). Da beinahe alle Modelle Log-Linearität aufwiesen, wurden die HRs mit zugehörigen 95 % KIs pro Anstieg um einen Interquartilsabstand (z.B. 25. zu 75. Quartile) der Exposition berechnet, was den Vorteil hat, dass die Ergebnisse über die verschiedenen Expositionen vergleichbar sind (160). Für die wenigen Modelle mit Hinweisen auf einen nicht-linearen Zusammenhang, wurden die Assoziationen zusätzlich mittels restricted cubic splines graphisch dargestellt (160). Die proportional hazards Annahme wurde mit Schoenfeld Residuen überprüft und, falls es Hinweise auf eine Verletzung der Annahme gab, wurden stratifizierte Cox Modelle berechnet.

Die Regressionen wurden für die gesamte Studienpopulation, sowie stratifiziert nach Geschlecht berechnet. Zusätzlich wurde als Endpunkt gesamtes malignes Melanom, sowie maligne Melanome an spezifischen anatomischen Lokalisationen (chronisch UV-exponierte Haut und sporadisch UV-exponierte Haut betrachtet (190). In weiteren Analysen wurden die Modelle nach UV-Sensitivität stratifiziert. Effektmodifikation durch Geschlecht oder UV-Sensitivität wurde mittels Wald-Test überprüft.

Verschiedene Sensitivitätsanalysen wurden durchgeführt, um die Robustheit der getroffenen Annahmen zu überprüfen. Um die Gefahr von umgekehrter Kausalität zu minimieren, wurden alle Teilnehmenden ausgeschlossen, welchen innerhalb von zwei Jahren nach der Basisuntersuchung ein malignes Melanom diagnostiziert wurde. In einer weiteren Analyse wurden alle Teilnehmenden ausgeschlossen, welche ihren Gesundheitszustand als „schlecht“ beschrieben oder nicht auf diese Frage antworteten. Zur besseren Interpretierbarkeit wurden die Expositionen körperliche Aktivität, Greifkraft und sedentäres Verhalten zusätzlich in alters- und geschlechtsspezifische Terzile (niedrig/mittel/hoch) eingeteilt, und die HRs für hohe versus niedrige Level berechnet. Die Plausibilität der multiplen Imputation wurde durch eine Complete Case Analyse in einer Stichprobe von Teilnehmenden, welche keine fehlenden Werte in Kovariablen aufwiesen, überprüft. Alle Analysen wurden mit der Statistik Software R Version 3.6.1 durchgeführt (Auszüge R Code im Abschnitt 7.4.5) (164).

5.3 Ergebnisse

5.3.1 Deskriptive Statistik

Die Nachbeobachtungszeit betrug im Durchschnitt 7,1 Jahre (Standardabweichung 1,0 Jahr). Während 2.472.298 Personenjahren traten bei 350.512 Teilnehmenden 1239 inzidente maligne Melanome auf (507 (40,9 %) an anatomischen Lokalisationen mit chronischer UV-Exposition (Klassifizierung siehe Tabelle 4), 712 (57,7 %) an anatomischen Lokalisationen mit sporadischer UV-Exposition und 20 an überlappenden oder nicht näher bezeichneten Lokalisationen). In der Subgruppe der 172.172 Männer traten 680 maligne Melanome auf (Inzidenzrate von 5,6 pro 10.000 Personenjahren). In der Subgruppe der 178.340 Frauen traten 559 maligne Melanome auf (Inzidenzrate von 4,4 pro 10.000 Personenjahren).

Tabelle 5, Tabelle 6 und Tabelle 7 zeigen die altersadjustierten Charakteristika der Studienpopulation zur Basisuntersuchung nach geschlechtsspezifischen Quartilen von körperlicher Aktivität, Greifkraft und sedentärem Verhalten. Teilnehmende mit hohen Leveln an körperlicher Aktivität wiesen größere Greifkraft auf und berichteten weniger sedentäres Verhalten als Teilnehmende mit niedrigen Leveln an körperlicher Aktivität. Der geringste Anteil an Teilnehmenden mit tertiärem Bildungsabschluss (Universität/ College) war in der Quartile mit höchster körperlicher Aktivität, höchstem sedentärem Verhalten oder niedrigster Greifkraft zu sehen. Der größte Anteil an Teilnehmenden mit „sehr gutem“ Gesundheitszustand fand sich in der Quartile mit der höchsten körperlichen Aktivität oder der größten Greifkraft. Mit steigenden Quartilen der Greifkraft, stieg auch die durchschnittliche Körpergröße stetig an. Teilnehmende in der höchsten Quartile von körperlicher Aktivität, Greifkraft oder sedentärem Verhalten berichteten im Durchschnitt mehr draußen verbrachte Zeit im Sommer und wiesen den größten Anteil an Personen mit UV-Exposition am Arbeitsplatz auf.

Tabelle 5: Charakteristika der Studienpopulation zur Basisuntersuchung nach direkter Standardisierung auf die Altersverteilung der Kohorte pro Quartile von körperlicher Aktivität

	Geschlechtsspezifische Quartile von körperlicher Aktivität (MET-Minuten pro Woche)			
	Q1	Q2	Q3	Q4
Anzahl an Teilnehmenden	87.705	87.680	87.632	87.495
Frauen (%)	50,6	50,9	50,8	51,0
Alter zur Basisuntersuchung (Jahre)	55,9	56,0	56,0	56,0
MET-Minuten pro Woche	419,0	1286,2	2571,8	6362,2
Greifkraft (kg)	30,9	31,6	32,0	32,2
Sedentäres Verhalten (h/d)	5,2	4,8	4,6	4,6
Region				
England	88,3	88,3	88,6	89,5
Wales	4,6	4,0	3,9	3,9
Schottland	7,1	7,6	7,5	6,6
Bildung (1999 NA)				
Universitäts-, College-Abschluss	37,3	41,3	39,0	28,1
A-, AS-Level, NVQ, HND, HNC, äquivalent, anderer professioneller Abschluss	23,4	23,2	23,2	24,6
O-Level, CSEs oder äquivalent	26,0	24,4	25,6	30,0
Keiner der oben genannten Abschlüsse	13,3	11,1	12,2	17,3
Rauchverhalten (Packungsjahre; 6452 NA)	7,7	6,1	6,0	6,9
Alkoholkonsum (g/d; 49.841 NA)	17,5	17,5	17,9	18,7
Kaffeekonsum (Tassen/d; 153 NA)	2,1	2,0	2,0	2,0
Körpergröße (cm; 309 NA)	169,3	169,5	169,4	168,9
Allgemeiner Gesundheitszustand (812 NA)				
Sehr gut	12,1	17,5	22,0	22,7
Gut	55,1	60,8	60,3	59,3
Mäßig	25,5	18,9	15,5	16,0
Schlecht	7,3	2,8	2,2	2,0
UV-Sensitivitäts-Kategorie (9757 NA)				
Niedrige UV-Sensitivität	42,3	43,0	44,0	46,0
Hohe UV-Sensitivität	57,7	57,1	56,0	54,0
Draußen verbrachte Zeit im Sommer (h/d; 7497 NA)	3,0	3,4	3,8	4,7
Sonnenbrände in der Kindheit (75.150 NA)	1,7	1,8	1,8	1,7
Solariumnutzung (%; 1675 NA)	9,0	9,2	9,6	10,2
Jahreszeit der Basisuntersuchung				
Frühling	28,7	28,6	28,3	28,7
Sommer	24,8	25,6	26,7	28,1
Herbst	23,7	24,1	24,4	24,1
Winter	22,8	21,7	20,7	19,1
UV Exposition am Arbeitsplatz (%; 5 NA)^a	4,7	4,9	6,8	16,5

Q = Quartile, NA = fehlende Werte, MET = Metabolisches Äquivalent, UV = Ultraviolettstrahlung.

^a In einer Teilstichprobe von 212.700 aktuell (selbstständig) erwerbstätigen Teilnehmenden.

Tabelle 6: Charakteristika der Studienpopulation zur Basisuntersuchung nach direkter Standardisierung auf die Altersverteilung der Kohorte pro Quartile von Greifkraft

	Geschlechtsspezifische Quartile von Greifkraft (kg)			
	Q1	Q2	Q3	Q4
Anzahl an Teilnehmenden	96.289	91.035	79.557	83.631
Frauen (%)	53,8	48,8	51,9	46,6
Alter zur Basisuntersuchung (Jahre)	56,3	56,1	55,9	55,5
MET-Minuten pro Woche	2448,8	2642,1	2722,3	2821,7
Greifkraft (kg)	22,4	30,2	34,3	41,8
Sedentäres Verhalten (h/d)	4,9	4,8	4,8	4,8
Region				
England	91,3	89,7	87,9	85,1
Wales	3,8	4,0	4,3	4,3
Schottland	4,9	6,3	7,8	10,7
Bildung (1999 NA)				
Universitäts-, College-Abschluss	33,3	36,1	38,1	39,8
A-, AS-Level, NVQ, HND, HNC, äquivalent, anderer professioneller Abschluss	22,9	23,6	24,0	24,6
O-Level, CSEs oder äquivalent	27,3	27,0	26,0	25,0
Keiner der oben genannten Abschlüsse	16,6	13,3	11,9	10,7
Rauchverhalten (Packungsjahre; 6452 NA)	6,9	6,7	6,5	6,6
Alkoholkonsum (g/d; 49.841 NA)	16,1	18,0	18,3	19,5
Kaffeekonsum (Tassen/d; 153 NA)	2,0	2,0	2,0	2,1
Körpergröße (cm; 309 NA)	166,6	168,9	169,8	172,5
Allgemeiner Gesundheitszustand (812 NA)				
Sehr gut	13,7	18,1	20,6	22,7
Gut	55,3	59,8	60,6	60,1
Mäßig	24,2	18,9	16,5	15,2
Schlecht	6,9	3,2	2,3	1,9
UV-Sensitivitäts-Kategorie (9757 NA)				
Niedrige UV-Sensitivität	43,1	43,6	43,7	45,3
Hohe UV-Sensitivität	57,0	56,4	56,3	54,7
Draußen verbrachte Zeit im Sommer (h/d; 7497 NA)	3,6	3,7	3,8	3,9
Sonnenbrände in der Kindheit (75.150 NA)	1,6	1,7	1,8	2,0
Solariumnutzung (%; 1675 NA)	9,1	9,4	9,7	9,8
Jahreszeit der Basisuntersuchung				
Frühling	30,1	28,8	28,1	27,0
Sommer	27,7	26,7	25,9	24,4
Herbst	22,6	23,7	24,7	25,5
Winter	19,6	20,8	21,3	23,2
UV Exposition am Arbeitsplatz (%; 5 NA)^a	7,0	7,4	8,4	10,0

Q = Quartile, NA = fehlende Werte, MET = Metabolisches Äquivalent, UV = Ultraviolettstrahlung.

^a In einer Teilstichprobe von 212.700 aktuell (selbstständig) erwerbstätigen Teilnehmenden.

Tabelle 7: Charakteristika der Studienpopulation zur Basisuntersuchung nach direkter Standardisierung auf die Altersverteilung der Kohorte pro Quartile von sedentärem Verhalten

	Geschlechtsspezifische Quartile von sedentärem Verhalten (h/d)			
	Q1	Q2	Q3	Q4
Anzahl an Teilnehmenden	106.899	93.201	70.949	79.463
Frauen (%)	55,0	43,0	57,4	48,9
Alter zur Basisuntersuchung (Jahre)	55,9	56,0	56,1	56,0
MET-Minuten pro Woche	2793,3	2760,2	2581,4	2430,7
Greifkraft (kg)	31,1	33,1	30,7	31,6
Sedentäres Verhalten (h/d)	2,5	4,2	5,4	8,2
Region				
England	88,0	88,5	89,1	89,6
Wales	3,8	4,3	4,2	4,1
Schottland	8,2	7,2	6,8	6,4
Bildung (1999 NA)				
Universitäts-, College-Abschluss	46,4	36,2	32,1	27,3
A-, AS-Level, NVQ, HND, HNC, äquivalent, anderer professioneller Abschluss	22,0	24,4	24,3	24,3
O-Level, CSEs oder äquivalent	21,5	26,4	30,0	30,2
Keiner der oben genannten Abschlüsse	10,1	13,0	13,6	18,2
Rauchverhalten (Packungsjahre; 6452 NA)	4,9	6,4	6,8	9,2
Alkoholkonsum (g/d; 49.841 NA)	16,5	19,0	17,6	18,8
Kaffeekonsum (Tassen/d; 153 NA)	1,9	2,0	2,1	2,1
Körpergröße (cm; 309 NA)	169,0	170,3	168,3	169,1
Allgemeiner Gesundheitszustand (812 NA)				
Sehr gut	23,6	18,4	16,8	13,5
Gut	59,3	60,4	60,4	55,4
Mäßig	14,8	18,4	19,5	24,9
Schlecht	2,4	2,9	3,3	6,3
UV-Sensitivitäts-Kategorie (9757 NA)				
Niedrige UV-Sensitivität	43,0	44,3	42,8	45,2
Hohe UV-Sensitivität	57,0	55,7	57,2	54,8
Draußen verbrachte Zeit im Sommer (h/d; 7497 NA)	3,4	3,7	3,8	4,2
Sonnenbrände in der Kindheit (75.150 NA)	1,8	1,8	1,7	1,8
Solariumnutzung (%; 1675 NA)	8,8	9,2	10,2	10,3
Jahreszeit der Basisuntersuchung				
Frühling	27,8	28,5	29,1	29,4
Sommer	27,9	26,4	25,3	24,9
Herbst	24,3	24,1	24,2	23,5
Winter	20,0	21,0	21,5	22,1
UV Exposition am Arbeitsplatz (%; 5 NA)^a	4,8	7,8	7,5	14,1

Q = Quartile, NA = fehlende Werte, MET = Metabolisches Äquivalent, UV = Ultraviolettstrahlung.

^a In einer Teilstichprobe von 212.700 aktuell (selbstständig) erwerbstätigen Teilnehmenden.

5.3.2 Körperliche Aktivität

Zwischen körperlicher Aktivität und malignem Melanom, malignem Melanom an chronisch UV-exponierten Körperlokalisationen oder malignem Melanom an sporadisch UV-exponierten Körperlokalisationen zeigten sich keine Assoziationen, weder im Basismodell, noch im voll adjustierten Modell (Abbildung 18). Auch in den nach Geschlecht oder UV-Sensitivität stratifizierten Modellen war der Nullwert 1,0 in allen Konfidenzintervallen enthalten (Abbildung 18). Im voll adjustierten Modell zur Assoziation zwischen körperlicher Aktivität und malignem Melanom an chronisch UV-exponierten anatomischen Lokalisationen gab es in der Subgruppe der Frauen Hinweise auf einen nichtlinearen Zusammenhang (Abbildung S 6).

Auch die zusätzliche Adjustierung für UV-Exposition am Arbeitsplatz in einer Teilstichprobe von 212.700 Teilnehmenden (mit 655 inzident aufgetretenen malignen Melanomen), die aktuell (selbstständig) erwerbstätig waren, veränderte die Ergebnisse nur marginal. Der Ausschluss von 308 malignen Melanom Fällen, welche innerhalb der ersten zwei Jahre der Nachbeobachtungszeit aufgetreten waren, sowie der Ausschluss von 13.353 Teilnehmenden (1196 maligne Melanom Fälle verblieben im Datensatz), welche ihren Gesundheitszustand als „schlecht“ beschrieben oder diese Frage nicht beantworteten, veränderte die Ergebnisse nicht maßgeblich. Nach Kategorisierung der Exposition zeigte sich eine HR von 1,04 (95 % KI 0,90-1,20) für hohe versus niedrige Level an körperlicher Aktivität. Die Analyse in einem Teildatensatz von 222.000 Teilnehmenden mit 776 malignen Melanomen, welche vollständige Informationen zu allen Kovariablen gegeben hatten, lieferte vergleichbare Ergebnisse wie die multipel imputierten Analysen (Tabelle S 7).

5.3.3 Greifkraft

In den für Alter, Geschlecht und Studienregion adjustierten Basismodellen war die HR für malignes Melanom pro Interquartilsanstieg an Greifkraft 1,23 (95 % KI 1,08-1,40) (Abbildung 19 A). Während Adjustierung für Körpergröße diese Assoziation auf 1,09 (95 % KI 0,95-1,25) abschwächte, verstärkte Adjustierung für mit UV-Exposition in Zusammenhang stehenden Variablen die Assoziation leicht (HR 1,24; 95 % KI 1,09-1,42). Die nach Geschlecht stratifizierten Basis-Modelle zeigten eine positive Assoziation bei den Männern (HR 1,16; 95 % KI 1,04-1,29), aber nicht bei den Frauen (Abbildung 19 A), obwohl keine Effektmodifikation durch Geschlecht beobachtet werden konnte (p-Wert für Interaktion durch Geschlecht = 0.710). Nach voller Adjustierung beinhalteten alle Konfidenzintervalle den Nullwert von 1,0

(Abbildung 19 A). Betrachtung nach unterschiedlich UV-exponierten Körperlokalisationen zeigte keine Assoziation zwischen Greifkraft und malignen Melanomen an chronisch UV-exponierten Stellen. Allerdings zeigte sich im Basismodell eine positive Assoziation zwischen Greifkraft und malignem Melanom an sporadisch UV-exponierten Hautstellen in der gesamten Kohorte (HR 1,39; 95 % KI 1,17-1,66), in der Stichprobe der Männer (HR 1,25; 95 % KI 1,08-1,44) und in der Stichprobe der Frauen (HR 1,17; 95 % KI 1,01-1,36) (Abbildung 19 A). Diese positive Assoziation blieb nach voller Adjustierung in der gesamten Gruppe und in der Stichprobe der Männer bestehen (Abbildung 19). Bei Stratifizierung nach UV-Sensitivität zeigte sich eine positive Assoziation zwischen Greifkraft und malignem Melanom bei Personen mit niedriger UV-Sensitivität (HR 1,29; 95 % KI 1,01-1,64) (Abbildung 19 B).

Nach zusätzlicher Adjustierung für UV-Exposition am Arbeitsplatz ($n = 212,700$; 655 Fälle), verstärkten sich die Assoziationen zwischen Greifkraft und malignem Melanom in der gesamten Kohorte (HR 1,20; 95 % KI 1,00-1,44) und in der Stichprobe der Frauen (HR 1,15; 95 % KI 0,97-1,37). Der Ausschluss von 308 Fällen malignes Melanom, welche innerhalb der ersten zwei Jahre nach der Basisuntersuchung aufgetreten waren, verstärkte die Assoziation zwischen Greifkraft und malignem Melanom in der gesamten Gruppe (HR 1,17; 95 % KI 1,00-1,37) und bei den Männern (HR 1,15; 95 % KI 1,02-1,31). Die Ergebnisse wurden durch den Ausschluss von Teilnehmenden mit schlechtem Gesundheitszustand oder ohne Angabe (Ausschluss von $n = 13,353$ Teilnehmenden) nur marginal verändert. Die voll adjustierte HR für hohe versus niedrige Level an Greifkraft war 1,10 (95 % KI 0,95-1,27). Bei der Analyse in einem Teildatensatz mit beobachteten Kovariablen ($n = 222,000$; 776 Fälle) beinhalteten alle Konfidenzintervalle den Nullwert 1,0 (Tabelle S 7).

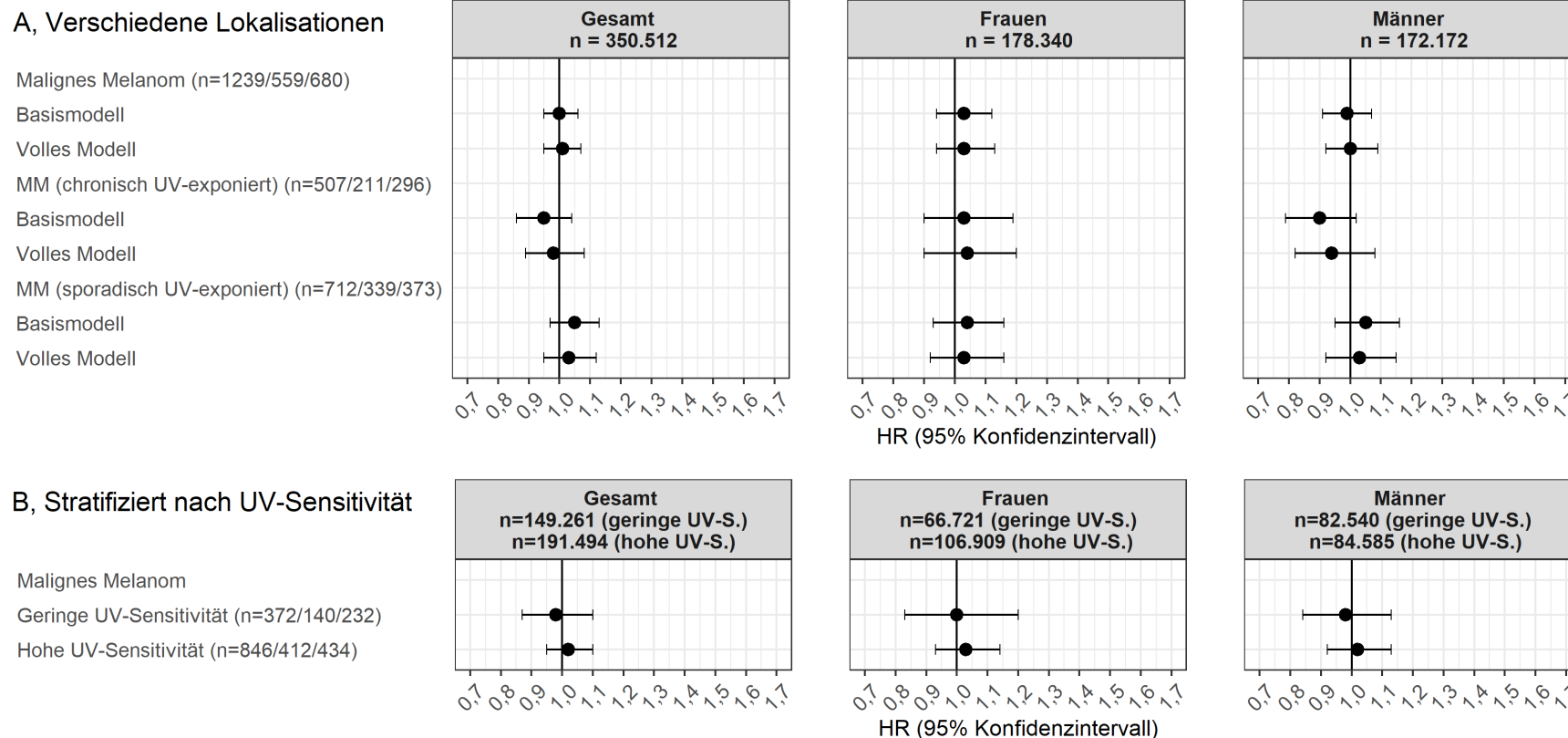


Abbildung 18: Assoziation zwischen körperlicher Aktivität und malignem Melanom an verschiedenen Lokalisationen (A) und stratifiziert nach UV-Sensitivität (B)

Cox proportional hazards Regression mit beobachteter Exposition (und UV-Sensitivitäts-Kategorie (B)) und multipel imputierten Kovariablen. Körperliche Aktivität (MET-Minuten pro Woche) wurde als kontinuierliche Variable ins Modell aufgenommen. Assoziationen wurden linear modelliert und HRs wurden für einen Anstieg um einen Interquartilsabstand (z.B. 25. bis 75. Quantile) berechnet.

Basismodelle wurden für Alter (als Zeitachse), Studienregion (Stratifikation der Baseline-Hazard) und in der gesamten Kohorte zusätzlich für Geschlecht adjustiert.

Volle Modelle wurden zusätzlich für Bildung, Packungsjahre, Alkoholkonsum, Kaffeekonsum, Körpergröße, draußen verbrachte Zeit im Sommer, Sonnenbrände in der Kindheit, Solariumnutzung, Saisonalität, sedentäres Verhalten und Greifkraft adjustiert. Die Modelle wurden für den UV-Sensitivitäts-Score adjustiert (A) oder danach stratifiziert (B): Geringe UV-Sensitivität: UV-Sensitivitäts-Index < 4; Hohe UV-Sensitivität: UV-Sensitivitäts-Index ≥ 4.

A, Verschiedene Lokalisationen

Malignes Melanom (n=1239/559/680)

Basismodell

Volles Modell

MM (chronisch UV-exponiert) (n=507/211/296)

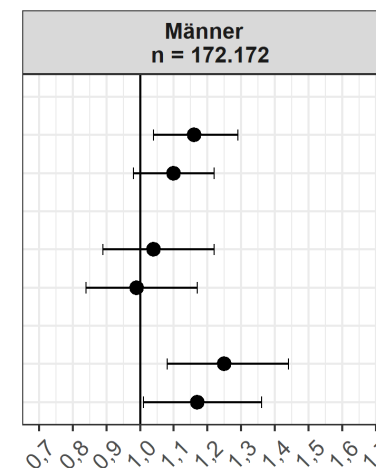
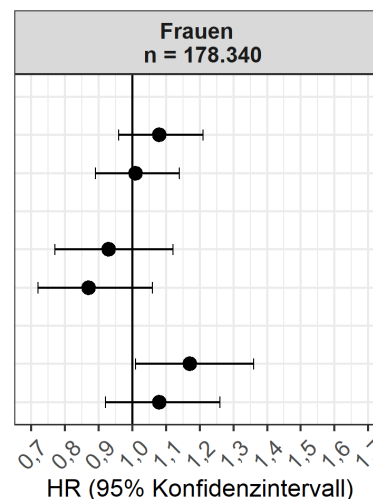
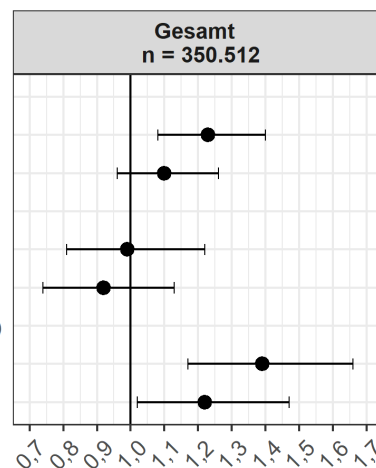
Basismodell

Volles Modell

MM (sporadisch UV-exponiert) (n=712/339/373)

Basismodell

Volles Modell



B, Stratifiziert nach UV-Sensitivität

Malignes Melanom

Geringe UV-Sensitivität (n=372/140/232)

Hohe UV-Sensitivität (n=846/412/434)

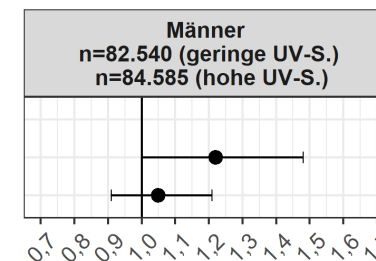
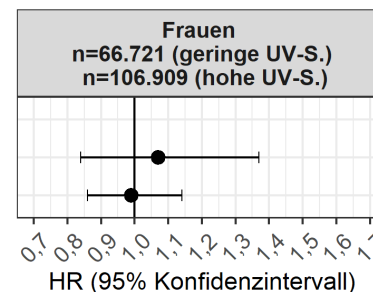
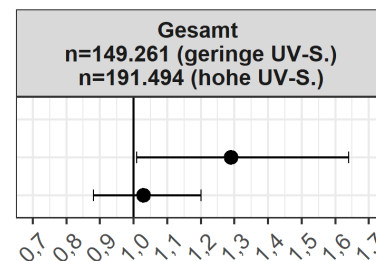


Abbildung 19: Assoziation zwischen Greifkraft und malignem Melanom an verschiedenen Lokalisationen (A) und stratifiziert nach UV-Sensitivität (B)

Cox proportional hazards Regression mit beobachteter Exposition (und UV-Sensitivitäts-Kategorie, (B)) und multipel imputierten Kovariablen. Greifkraft (kg) wurde als kontinuierliche Variable ins Modell aufgenommen. Assoziationen wurden linear modelliert und HRs wurden für einen Anstieg um einen Interquartilsabstand (z.B. 25. bis 75. Quantile) berechnet.

Basismodelle wurden für Alter (als Zeitachse), Studienregion (Stratifizierung der Baseline-Hazard) und in der gesamten Kohorte zusätzlich für Geschlecht adjustiert.

Volle Modelle wurden zusätzlich für Bildung, Packungsjahre, Alkoholkonsum, Kaffeekonsum, Körpergröße, draußen verbrachte Zeit im Sommer, Sonnenbrände in der Kindheit, Solariumnutzung, Saisonalität, sedentäres Verhalten und körperliche Aktivität adjustiert. Die Modelle wurden für den UV-Sensitivitäts-Score adjustiert (A) oder danach stratifiziert (B): Geringe UV-Sensitivität: UV-Sensitivitäts-Index < 4; Hohe UV-Sensitivität: UV-Sensitivitäts-Index \geq 4.

A, Verschiedene Lokalisationen

Malignes Melanom (n=1239/559/680)

Basismodell

Volles Modell

MM (chronisch UV-exponiert) (n=507/211/296)

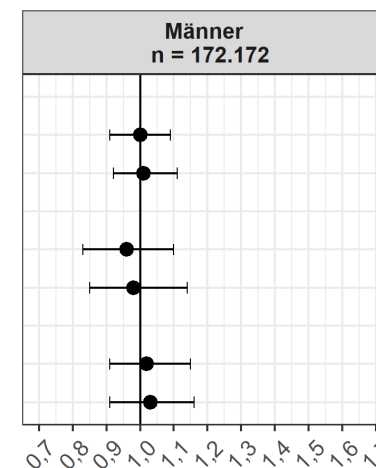
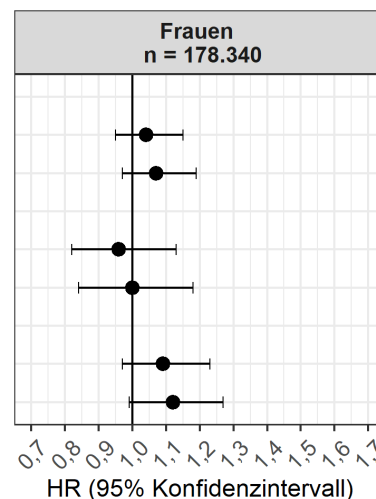
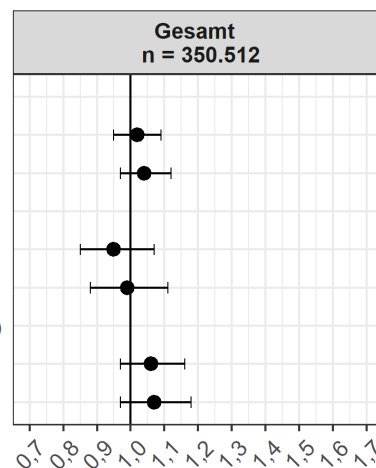
Basismodell

Volles Modell

MM (sporadisch UV-exponiert) (n=712/339/373)

Basismodell

Volles Modell



B, Stratifiziert nach UV-Sensitivität

Malignes Melanom

Geringe UV-Sensitivität (n=372/140/232)

Hohe UV-Sensitivität (n=846/412/434)

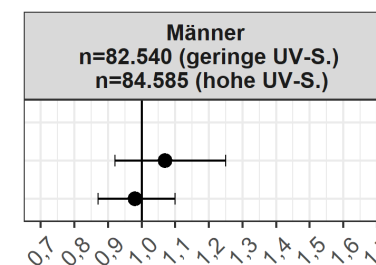
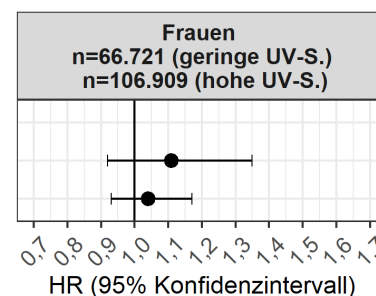
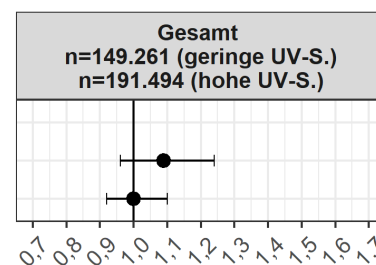


Abbildung 20: Assoziation zwischen sedentärem Verhalten und malignem Melanom an verschiedenen Lokalisationen (A) und stratifiziert nach UV-Sensitivität (B)

Cox proportional hazards Regression mit beobachteter Exposition (und UV-Sensitivitäts-Kategorie, (B)) und multipel imputierten Kovariablen. Sedentäres Verhalten (h/d) wurde als kontinuierliche Variable ins Modell aufgenommen. Assoziationen wurden linear modelliert und HRs wurden für einen Anstieg um einen Interquartilsabstand (z.B. 25. bis 75. Quantile) berechnet.

Basismodelle wurden für Alter (als Zeitachse), Studienregion (Stratifizierung der Baseline-Hazard) und in der gesamten Kohorte zusätzlich für Geschlecht adjustiert.

Volle Modelle wurden zusätzlich für Bildung, Packungsjahre, Alkoholkonsum, Kaffeeconsum, Körpergröße, draußen verbrachte Zeit im Sommer, Sonnenbrände in der Kindheit, Solariumnutzung, Saisonalität, körperliche Aktivität und Greifkraft adjustiert. Die Modelle wurden für den UV-Sensitivitäts-Score adjustiert (A) oder danach stratifiziert (B): Geringe UV-Sensitivität: UV-Sensitivitäts-Index < 4; Hohe UV-Sensitivität: UV-Sensitivitäts-Index ≥ 4.

5.3.4 Sedentäres Verhalten

Alle Konfidenzintervalle der Basis- und vollständig adjustierten Modelle zwischen sedentärem Verhalten und Inzidenz von malignem Melanom, malignen Melanom bei chronisch UV-exponierter Hautlokalisationen oder sporadisch UV-exponierter Hautlokalisationen, sowie in nach Geschlecht oder UV-Sensitivität stratifizierten Modellen beinhalteten den Nullwert von 1,0 (Abbildung 20). In der gesamten Population und bei den Frauen zeigten sich Hinweise auf einen nicht-linearen Zusammenhang zwischen sedentärem Verhalten und malignem Melanom an sporadisch UV-exponierten Hautlokalisationen (Abbildung S 6).

Zusätzliche Adjustierung für UV-Exposition am Arbeitsplatz, Ausschluss von malignen Melanom Fällen, welche innerhalb der ersten zwei Jahre Nachbeobachtungszeit aufgetreten waren oder der Ausschluss von 13.353 Teilnehmenden mit schlechtem Gesundheitszustand oder ohne Angabe veränderten die Ergebnisse nicht maßgeblich. Die HR für hohe versus niedrige Level an sedentärem Verhalten war 1,10 (95 % KI 0,96-1,27) im voll adjustierten Modell. Die Analysen im Teildatensatz mit 222.000 Teilnehmenden, welche vollständige Kovariablendaten lieferten, zeigten vergleichbare Ergebnisse wie die Analysen im imputierten Datensatz (Tabelle S 7).

5.4 Diskussion

Die prospektive Beobachtung von über 350.000 Männern und Frauen für bis zu 9 Jahre, zeigte keine Assoziation zwischen körperlicher Aktivität, sedentärem Verhalten und der Inzidenz von malignem Melanom. Die positive Assoziation zwischen Greifkraft und malignem Melanom im Basismodell wurde durch die Kontrolle a priori identifizierter Kovariablen, vor allem Körpergröße, deutlich abgeschwächt.

Die Assoziation zwischen körperlicher Aktivität und malignem Melanom war bereits in früheren Studien betrachtet worden, allerdings mit heterogenen Ergebnissen. In der UK Biobank zeigte sich zwischen dem, aus körperlicher Aktivität, geschätzten Energieverbrauch und malignem Melanom eine HR von 1,05 (95 % KI 0,93-1,19), jedoch war diese Analyse nicht für mit UV-Exposition und -Sensitivität in Zusammenhang stehenden Variablen adjustiert worden (211). Eine Metaanalyse fasste drei Kohortenstudien, darunter eine große gepoolte Studie, und fünf Fall-Kontroll Studien zusammen, um die Assoziation zwischen körperlicher Aktivität und malignem Melanom zu quantifizieren (184). Während die Fall-Kontroll Studien, welche größtenteils für UV-Exposition auf dem Individuallevel adjustiert waren, eine HR von

0,85 (95 % KI 0,63-1,14) zeigten, errechnete die Metaanalyse für die Kohortenstudien eine HR von 1,27 (95 % KI 1,16-1,40). Die Kohortenstudien waren größtenteils in Nordamerika durchgeführt worden, wo extreme UV-Expositionen auftreten können (UV-Index Werte ≥ 11 (212)), jedoch wurde keine davon für UV-Exposition auf dem Individuallevel adjustiert (184). Im Gegensatz hierzu zeigte sich in der vorliegenden prospektiven Analyse keine Assoziation zwischen körperlicher Aktivität und malignem Melanom, weder mit noch ohne Adjustierung für UV-Exposition und -Sensitivität. Möglicherweise ist dies auf die geringere UV-Belastung in Großbritannien zurückzuführen (UV-Index ≤ 8 (213)), wodurch die verzerrende Assoziation zwischen körperlicher Aktivität und UV-vermittelter Hautschädigung abgeschwächt wird. Diese Hypothese wird durch eine frühere Untersuchung gestützt, welche zeigte, dass die positive Assoziation zwischen körperlicher Aktivität und malignem Melanom nicht in Regionen mit niedrigen UV-Leveln zu sehen war (94). Auch könnten die Wetterverhältnisse in Großbritannien Sport in geschlossenen Räumen begünstigen (214). Neben diesen strukturellen Erklärungen ist auch denkbar, dass die UK-Biobank Frage zur im Sommer draußen verbrachten Zeit eher die chronische UV-Exposition abbildet und wenig aussagekräftig in Bezug auf die sporadische UV-Exposition (z.B. Sonnenbaden im Urlaub) ist. Gerade die sporadische UV-Exposition ist ein starker Risikofaktor für die Entstehung des malignen Melanoms, möglicherweise da die betroffenen Hautregionen weniger Möglichkeit zur Adaption (z.B. Bräunung und Verdickung der Haut) haben, als chronisch exponierte Stellen (201, 215, 216)

Die Assoziation zwischen muskulärer Fitness und malignem Melanom wurde nach unserem besten Wissen bislang nicht untersucht. Eine bisherige Studie betrachtete die Assoziation zwischen kardiorespiratorischer Fitness und malignem Melanom und fand eine HR von 2,19 (95 % KI 0,99-4,96) (34). In den vorliegenden Basismodellen zeigten sich positive Assoziationen, die aber nach Adjustierung für z.B. Körpergröße verschwanden. Tatsächlich ist die Körpergröße eine starke Störgröße für die betrachtete Assoziation, da sie ein Surrogat für genetische, hormonelle, umwelt- und ernährungsbedingte Faktoren darstellt, welche das Risiko für malignes Melanom erhöhen (196, 202) und gleichzeitig stark mit Greifkraft assoziiert ist (203-205). Wie Körpergröße ist auch Geburtsgewicht ein möglicher Risikofaktor für malignes Melanom (196, 217) und mit Greifkraft assoziiert (218). Da die gezeigten Analysen nicht für Geburtsgewicht kontrolliert wurden, sind die Ergebnisse möglicherweise durch ungemessene Konfundierung verzerrt. Allerdings zeigte eine frühere Studie, dass der Einfluss von Geburtsgewicht auf das maligne Melanom nach Kontrolle für Körpergröße

vernachlässigbar ist (217). In einigen Subanalysen (z.B. malignes Melanom an sporadisch UV-exponierten Lokalisationen, Ausschluss von Fällen innerhalb der ersten zwei Jahre der Nachbeobachtungszeit) war die positive Assoziation zwischen Greifkraft und dem betrachteten Endpunkt auch bei voller Adjustierung zu sehen. Da diese Adjustierung körperliche Aktivität und sedentäres Verhalten beinhaltete, ist die positive Assoziation möglicherweise auf den genetisch determinierten, nicht modifizierbaren Anteil der Greifkraft zurückzuführen. Diese Beobachtung sollte in weiteren Analysen, beispielsweise mittels Mendelscher Randomisierung, überprüft werden.

Die hier beschriebenen Analysen liefern nach unserem Wissen die ersten Ergebnisse zur Assoziation zwischen sedentärem Verhalten und malignem Melanom. Generell ist denkbar, dass die negativen Effekte des sedentären Verhaltens durch die geringere UV-Exposition aufgrund des häufigen Sitzens (vor dem Fernseher, Computer) ausgeglichen werden. Allerdings zeigte die deskriptive Betrachtung der Studienpopulation, dass die Teilnehmenden mit dem höchsten Level an sedentärem Verhalten auch die meiste Zeit im Freien berichteten. Zusätzlich legten unsere Ergebnisse nahe, dass die verzerrende Wirkung der UV-Exposition keinen Einfluss auf die Analysen hatte.

Eine der Stärken der vorliegenden Analysen ist die prospektive Beobachtung und die Adjustierung für eine Vielzahl an a priori identifizierten, relevanten, und auf dem Individuallevel gemessenen Störgrößen. Durch den großen Umfang der UK Biobank und die Vielfalt an Datenerhebungen war es möglich stratifizierte Analysen durchzuführen und den Endpunkt malignes Melanom getrennt nach anatomischen Lokalisationen zu betrachten. Zudem lieferten die objektive Messung der Greifkraft und die Verknüpfung der Studiendaten mit Krebsregistern valide und vollständige Informationen. Im Gegensatz hierzu stellt die subjektive Messung von körperlicher Aktivität und sedentärem Verhalten eine potenzielle Limitation dar, da diese Methodik anfällig für einen Informationsbias ist (16). Während die gerätebasierte Messung der körperlichen Aktivität, beispielsweise mittels Akzelerometrie, die bevorzugte Methode darstellt, ist der derzeit vorliegende Nachbeobachtungszeitraum für die zwischen 2013 und 2015 durchgeführte Akzelerometermessung in der UK Biobank zu gering. Auch die meisten Kovariablen, insbesondere Informationen zu UV-Exposition und -Sensitivität wurden subjektiv erhoben. Diese wurden unter der Annahme verwendet, dass die Fehlklassifizierung aufgrund des prospektiven Designs nicht differenziell ist. Eine weitere Limitation ist die Beschränkung des sedentären Verhaltens auf die Domänen Freizeit und Fortbewegung. Allerdings haben möglicherweise eben diese Domänen den schädlichsten

Einfluss auf die Gesundheit (189). Die statistische Präzision und Teststärke wären durch eine längere Nachbeobachtungszeit und eine größere Anzahl an Fällen höher. Zusätzlich begünstigt die Annahme, dass zur Basisuntersuchung gemessene Expositionen und Störgrößen über die gesamte Beobachtungszeit hindurch konstant bleiben Regression dilution bias (219). Jedoch zeigte eine frühere Metaanalyse, dass sich zumindest die UV-Exposition nicht über die Zeit änderte (201). Wie bereits erwähnt, zeigte eine Studie, dass die mittels UK Biobank Daten berechneten Assoziationen trotz der geringen Teilnahmequote (5,5 %) im Großen und Ganzen auf die Allgemeinbevölkerung übertragen werden können (180).

5.5 Schlussfolgerung

Die Ergebnisse deuten darauf hin, dass keine Assoziationen zwischen körperlicher Aktivität, muskulärer Fitness, sedentärem Verhalten und der Inzidenz von malignem Melanom bestehen. Die initial beobachteten positiven Assoziationen zwischen Greifkraft und malignem Melanom wurden durch Adjustierung für Körpergröße stark abgeschwächt.

6 Zusammenfassung

In der vorliegenden Arbeit wurde die Durchführung und Qualitätssicherung von verschiedenen subjektiven und gerätebasierten körperlichen Aktivitätsmessungen in der deutschen epidemiologischen Gesundheitsstudie NAKO beschrieben. Zusätzlich wurden die longitudinalen Assoziationen von körperlicher Aktivität, sedentärem Verhalten, muskulärer und kardiorespiratorischer Fitness mit der Inzidenz wenig erforschter Krebsentitäten (Kolorektalkarzinom und malignes Melanom) in der britischen epidemiologischen Kohortenstudie UK Biobank dargelegt.

Die subjektive Erhebungsmethode GPAQ wurde bereits in vielen Studien angewendet und validiert. In der NAKO Gesundheitsstudie lieferte der Einsatz als Touchscreen-Fragebogen zum Selbstauffüllen allerdings teilweise unplausible Angaben und deutete auf einen altersabhängigen Informationsbias hin, welcher möglicherweise durch die subjektive Einschätzung der Intensitäten zustande kam. Die Daten wurden nach dem GPAQ Analyse Guide aufbereitet und neue Variablen wurden abgeleitet, welche für die weitere Verwendung in epidemiologischen Studien gegen die Akzelerometrie kalibriert werden können. Die Qualitätssicherung ergab, dass die subjektive Erfassung mittels 24-Stunden Protokoll cpar24 detailreiche und plausible Daten zu Art, Dauer, Intensität und Kontext körperlicher Aktivität und sedentärem Verhalten lieferte, welche auch für die Aufbereitung und Analyse der Akzelerometrie-Daten verwendet werden können. Für den weiteren Einsatz des Instrumentes wurde allerdings auch Überarbeitungsbedarf der Software in Hinblick auf Benutzerfreundlichkeit identifiziert. Die Aufbereitung der Akzelerometrie-Rohdaten stellt eine wichtige Ergänzung zu den aggregierten ActiLife-Variablen dar, da hier größere Transparenz und Flexibilität, sowie ein größerer Detailgrad (z.B. Information pro Tag oder Stunde, Körperposition) erreicht wurde. Gleichzeitig ist die Bearbeitung von Rohdaten generell rechenaufwändiger, weshalb die Bereitstellung abgeleiteter Variablen auch Wissenschaftlern und Wissenschaftlerinnen ohne entsprechende Kenntnisse oder Kapazitäten den Zugang zu Akzelerometerdaten liefert. Die Qualitätssicherung und Harmonisierung der verschiedenen NAKO-Instrumente, sowie die Ableitung verschiedener Variablen zu Typ und Domäne körperlicher Aktivität und sedentären Verhaltens ist der entscheidende Schritt für die parallele Verwendung der subjektiven und gerätebasierten Daten in epidemiologischen Analysen.

Die longitudinalen Analysen der UK Biobank Daten zeigten eine inverse Assoziation zwischen kardiorespiratorischer Fitness und dem Risiko für Kolorektalkarzinom bei Männern, jedoch nicht bei Frauen. Die Assoziation schien beim proximalen Kolonkarzinom besonders stark ausgeprägt zu sein, was aufgrund der geringen Teststärke allerdings in weiteren Studien verifiziert werden sollte. Während die kardiorespiratorische Fitness durch einen gesunden Lebensstil positiv beeinflusst werden kann, ist auch die Messung in der klinischen Praxis zur Identifizierung von Hochrisikogruppen denkbar.

In weiteren UK Biobank Analysen wurden die Assoziationen zwischen subjektiv gemessener körperlicher Aktivität, sedentärem Verhalten, objektiv gemessener Greifkraft und Inzidenz von malignem Melanom untersucht. In den Hauptmodellen zeigten sich nach voller Adjustierung (vor allem für Körpergröße) keine Assoziationen. Dies deutet darauf hin, dass die in früheren Studien gefundenen positiven Assoziationen zwischen körperlicher Aktivität und malignem Melanom auf nicht berücksichtigte Konfundierung zurückzuführen sind. In einigen Subanalysen zeigte sich eine positive Assoziation zwischen Greifkraft und malignem Melanom (beispielsweise an sporadisch UV-exponierten Hautstellen) bei gleichzeitiger Adjustierung für körperliche Aktivität. Die Effekte der genetisch determinierten Komponente von Greifkraft sollten daher in Mendelschen Randomisierungsstudien weiter untersucht werden.

Die Ergebnisse dieser longitudinalen Analysen können in Metaanalysen mit anderen Studien gepoolt werden und so als Grundlage für weitere Empfehlungen, Maßnahmen zur Förderung der Öffentlichen Gesundheit oder Evidenzbasierten Medizin dienen.

7 Anhang

7.1 Subjektive Instrumente zur Erfassung der körperlichen Aktivität

7.1.1 GPAQ Fragebogen in der NAKO

Tabelle S 1: GPAQ Fragen und Antwortmöglichkeiten in der NAKO

Frage	Antwortoptionen
<p>Im Folgenden möchten wir Sie gerne dazu befragen, wie viel Zeit Sie in einer gewöhnlichen Woche mit verschiedenen körperlichen Aktivitäten verbringen. Bitte beantworten Sie diese Fragen selbst dann, wenn Sie sich selbst nicht für eine körperlich aktive Person halten. Denken Sie zuerst über die Zeit nach, während der Sie arbeiten. Schließen Sie dabei all die Aufgaben ein, die Sie erledigen müssen, wie bezahlte und unbezahlte Arbeit, studieren/lernen, Aufgaben im Haushalt, Arbeitssuche, und auch ernten, fischen oder jagen. Wenn Sie die Fragen beantworten, denken Sie daran, dass "intensive körperliche Aktivitäten" diejenigen Aktivitäten sind, die große Anstrengung erfordern und daher Atmung und Puls stark zunehmen. "Moderate körperliche Aktivitäten" sind solche, die moderate Anstrengung erfordern und zu einer leichten Erhöhung der Atmung und des Pulses führen.</p>	
<p>Beinhaltet Ihre Arbeit intensive körperliche Aktivität, bei der Atmung und Puls stark zunehmen, wie schwere Lasten tragen oder heben, graben oder Bauarbeiten mit einer Dauer von mindestens zehn Minuten?</p>	<p>ja / nein / (keine Angabe / weiß nicht)</p>
<p>An wie vielen Tagen in einer gewöhnlichen Woche betätigen Sie sich bei der Arbeit körperlich intensiv?</p>	<p>0 - 7 Tage (Keine Angabe / weiß nicht)</p>
<p>Wie viel Zeit verbringen Sie an einem gewöhnlichen Tag bei der Arbeit mit intensiver körperlicher Aktivität?</p>	<p>0 - 1080 Minuten (keine Angabe / weiß nicht)</p>
<p>Beinhaltet Ihre Arbeit moderate körperliche Aktivität, bei der Atmung und Puls leicht zunehmen, wie flottes Gehen oder Tragen leichter Lasten mit einer Dauer von mindestens zehn Minuten?</p>	<p>ja / nein / (keine Angabe / weiß nicht)</p>
<p>An wie vielen Tagen in einer gewöhnlichen Woche führen Sie bei der Arbeit moderate körperliche Aktivität aus?</p>	<p>0 - 7 Tage (Keine Angabe/ weiß nicht)</p>
<p>Wie viel Zeit verbringen Sie an einem gewöhnlichen Tag bei der Arbeit mit moderater körperlicher Aktivität?</p>	<p>0 - 1080 Minuten (keine Angabe / weiß nicht)</p>
<p>Die nächsten Fragen schließen die körperliche Aktivität bei der Arbeit, die Sie bereits erwähnt haben, aus. Wir möchten Sie nun dazu befragen, wie Sie sich von Ort zu Ort fortbewegen, beispielsweise von zu Hause zur Arbeitsstelle, zum Einkaufen, zum Markt oder zur Kirche.</p>	





Gehen Sie zu Fuß oder fahren Sie mit dem Fahrrad, um von einem Ort zum anderen zu kommen, mit einer Dauer von mindestens zehn Minuten?	ja / nein / (keine Angabe / weiß nicht)
An wie vielen Tagen in einer gewöhnlichen Woche gehen Sie zu Fuß oder fahren Sie mit dem Fahrrad, um von einem Ort zum anderen zu kommen, mit einer Dauer von mindestens zehn Minuten?	0 - 7 Tage (Keine Angabe/ weiß nicht)
Wie viel Zeit investieren Sie an einem gewöhnlichen Tag, um zu Fuß oder mit dem Fahrrad von einem Ort zum anderen zu kommen?	0 - 1080 Minuten (keine Angabe / weiß nicht)
Die nächsten Fragen schließen die körperliche Aktivität bei der Arbeit und zur Fortbewegung, die Sie bereits erwähnt haben, aus. Wir möchten Sie nun zu Ihrer körperlichen Aktivität und Ihrem Sport während der Freizeit befragen.	
Betreiben Sie in der Freizeit intensive körperliche Aktivität oder Sport, bei dem Atmung und Puls stark zunehmen, wie Laufen oder Fußball mit einer Dauer von mindestens zehn Minuten?	ja / nein / (keine Angabe / weiß nicht)
An wie vielen Tagen in einer gewöhnlichen Woche betreiben Sie intensive körperliche Aktivität oder Sport in der Freizeit?	0 - 7 Tage (Keine Angabe/ weiß nicht)
Wie viel Zeit investieren Sie an einem gewöhnlichen Tag in der Freizeit in intensive körperliche Aktivität oder Sport?	0 - 1080 Minuten (keine Angabe / weiß nicht)
Betreiben Sie in der Freizeit moderate körperliche Aktivität oder Sport, bei der Atmung und Puls leicht zunehmen, wie flottes Gehen, Fahrrad fahren, Schwimmen oder Volleyball mit einer Dauer von mindestens zehn Minuten?	ja / nein / (keine Angabe / weiß nicht)
An wie vielen Tagen in einer gewöhnlichen Woche betreiben Sie moderate körperliche Aktivität oder Sport in der Freizeit?	0 - 7 Tage (Keine Angabe/ weiß nicht)
Wie viel Zeit investieren Sie an einem gewöhnlichen Tag in der Freizeit in moderate körperliche Aktivität oder Sport?	0 - 1080 Minuten (keine Angabe / weiß nicht)
Bei der nächsten Frage geht es um die Zeit, die Sie mit Sitzen oder Ruhen verbringen, bei der Arbeit, zu Hause, zur Fortbewegung oder mit Freunden, zum Beispiel am Schreibtisch sitzen, Auto, Bus, Zug fahren, Karten spielen oder fernsehen. Die Zeit, die Sie mit schlafen verbringen, sollte ausgeschlossen werden.	
Wie viel Zeit verbringen Sie an einem gewöhnlichen Tag mit Sitzen oder Ruhen?	0 – 1439 Minuten (keine Angabe / weiß nicht)

7.1.2 cpar24 Protokoll in der NAKO

The screenshot shows the cpar24 software interface. On the left, a vertical menu lists activity categories: ALLE TÄTIGKEITEN VON A BIS Z, SCHLAFEN, TÄTIGKEITEN IM LIEGEN, KÖRPERPFLEGE, UMKLEIDEN, ERNÄHRUNG, GEHEN, FORTBEWEGEN, REISEN, HAUSARBEIT, BERUF, ARBEIT, EINKAUFEN, TERMINE, FREIZEITAKTIVITÄTEN, HOBBY (highlighted), SPORT, FAMILIENLEBEN, SOZIALE KONTAKTE, AUßER-HAUS-AKTIVITÄTEN, GARTENARBEIT, HEIMWERKEN, and SONSTIGE TÄTIGKEITEN. The 'Freizeitaktivitäten, Hobby' category is selected, and a search window titled 'Tätigkeit suchen' displays a list of activities. 'Gitarre spielen' is highlighted in blue. To the right of the search window, a large box contains the text 'Gitarre spielen'. Below the search window, a question asks: 'Fand die Tätigkeit im Sitzen, Stehen oder im Sitzen UND Stehen statt?'. Three options are shown with icons: 'Sitzen' (person sitting), 'Stehen' (person standing), and 'Sitzen und Stehen' (person sitting and standing with a plus sign). The 'Sitzen und Stehen' option is checked. Below this, a question asks: 'Wie viel Prozent fand im Sitzen statt?'. A horizontal slider scale from 0% to 100% in 10% increments is shown, with a blue arrow pointing to the 50% mark. Below the slider, a question asks: 'Bitte bestimmen Sie BEGINN und ENDE der Tätigkeit'. Two time selection fields are shown: 'BEGINN' with values 06, 00 and 'Uhr', and 'ENDE' with values 07, 00 and 'Uhr'. At the bottom, there are two blue buttons: 'Tätigkeit speichern' and 'Abbrechen'.

Abbildung S 1: Screenshot cpar24: Auswahl von Aktivitäten, Spezifikation von Körperposition und Angabe von Start- und Endzeit

Wie anstrengend war die Tätigkeit für Sie?


 leicht
  mittel
  anstrengend

Bitte bestimmen Sie BEGINN und ENDE der Tätigkeit

BEGINN Uhr

ENDE Uhr

Abbildung S 2: Screenshot cpar24: Spezifikation von Intensität























cpar24		Tätigkeit hinzufügen +		Klein / Mittel / Groß		Hilfe	
						Menu	
						Abmelden	
Dösen, schlummern	00:00 Uhr bis 00:10 Uhr						
Essen, trinken	00:10 Uhr bis 00:25 Uhr						
Aqua-Jogging	00:10 Uhr bis 00:25 Uhr						
Auto be- und entladen	00:10 Uhr bis 00:25 Uhr						
Aufgaben im Haushalt erledigen	00:25 Uhr bis 11:25 Uhr						
Bitte ergänzen Sie fehlende Tätigkeiten!	11:25 Uhr bis 20:25 Uhr						
Besuch einer Veranstaltung, dabei gemütlich gehen (z.B. Ausstellung, Museum)	20:25 Uhr bis 21:25 Uhr						
Bitte ergänzen Sie fehlende Tätigkeiten!	21:25 Uhr bis 24:00 Uhr						

Abbildung S 3: Screenshot cpar24: Übersicht aller eingetragenen Aktivitäten und Markierung von fehlenden Zeiträumen

7.1.3 IPAQ Fragebogen in der UK Biobank

Tabelle S 2: IPAQ-SF Fragen und Antwortmöglichkeiten in der UK Biobank

Frage	Harte Plausibilitätskontrolle	Weiche Plausibilitätskontrolle
<p>Als nächstes folgen einige Fragen zu Ihren alltäglichen Aktivitäten. Wir wissen, dass die exakte Beantwortung einiger dieser Fragen schwierig sein kann, aber eine ungefähre Antwort ist besser als keine. Die ersten paar Fragen beziehen sich auf die Zeit, die Sie in den letzten 4 Wochen damit verbracht haben, körperlich aktiv zu sein. Um diese Fragen zu beantworten, denken Sie an alle Aktivitäten, die Sie bei der Arbeit, als Teil Ihrer Hausarbeit, um von Ort zu Ort zu kommen und in Ihrer Freizeit ausführen.</p>		
<p>An wie vielen Tagen einer typischen WOCHE gingen Sie mindestens 10 Minuten am Stück? (Inklusive Gehen in der Arbeit, am Arbeitsweg, im Sport oder in der Freizeit)</p>	<p>≥ 0 Tage und ≤ 7 Tage</p>	
<p>Wie viele Minuten sind Sie üblicherweise an einem typischen TAG gegangen?</p>	<p>≥ 0 Minuten und ≤ 1440 Minuten</p>	<p>≥ 10 Minuten und ≤ 300 Minuten</p>
<p>Wie viele Tage einer typischen WOCHE haben Sie mindestens 10 Minuten am Stück moderate körperliche Aktivitäten wie das Tragen leichter Lasten, oder Radfahren in normaler Geschwindigkeit durchgeführt? (nicht Gehen)</p>	<p>≥ 0 Tage und ≤ 7 Tage</p>	
<p>Wie viele Minuten haben Sie üblicherweise mit moderater körperlicher Aktivität an einem typischen TAG verbracht?</p>	<p>≥ 0 Minuten und ≤ 1440 Minuten</p>	<p>≥ 10 Minuten und ≤ 300 Minuten</p>
<p>Wie viele Tage einer typischen WOCHE haben Sie mindestens 10 Minuten am Stück intensive körperliche Aktivitäten durchgeführt (Dies sind Aktivitäten, die Sie ins Schwitzen bringen oder Ihre Atmung beschleunigen, wie beispielsweise schnelles Radfahren, Aerobics oder Heben schwerer Gegenstände)</p>	<p>≥ 0 Tage und ≤ 7 Tage</p>	
<p>Wie viele Minuten haben Sie üblicherweise mit intensiver körperlicher Aktivität an einem typischen TAG verbracht?</p>	<p>≥ 0 Minuten und ≤ 1440 Minuten</p>	<p>≥ 10 Minuten und ≤ 300 Minuten</p>

Der Fragebogen wurde aus dem Englischen ins Deutsche übersetzt.

7.1.4 Fragen zu sedentärem Verhalten in der UK Biobank

Tabelle S 3: Fragen und Antwortmöglichkeiten zum sedentären Verhalten in der UK Biobank

Frage	Harte Plausibilitätskontrolle	Weiche Plausibilitätskontrolle
Wie viele Stunden verbringen Sie an einem typischen TAG mit Fernsehschauen?	≥ 0 Stunden und ≤ 24 Stunden	≤ 8 Stunden
Wie viele Stunden verbringen Sie an einem typischen TAG vor dem Computer? (Nicht Computer-Nutzung in der Arbeit)	≥ 0 Stunden und ≤ 24 Stunden	≤ 6 Stunden
Wie viele Stunden verbringen Sie an einem typischen TAG mit sitzender Fortbewegung?	≥ 0 Stunden und ≤ 24 Stunden	≤ 6 Stunden

Der Fragebogen wurde aus dem Englischen ins Deutsche übersetzt.

7.2 Zusätzliche Informationen

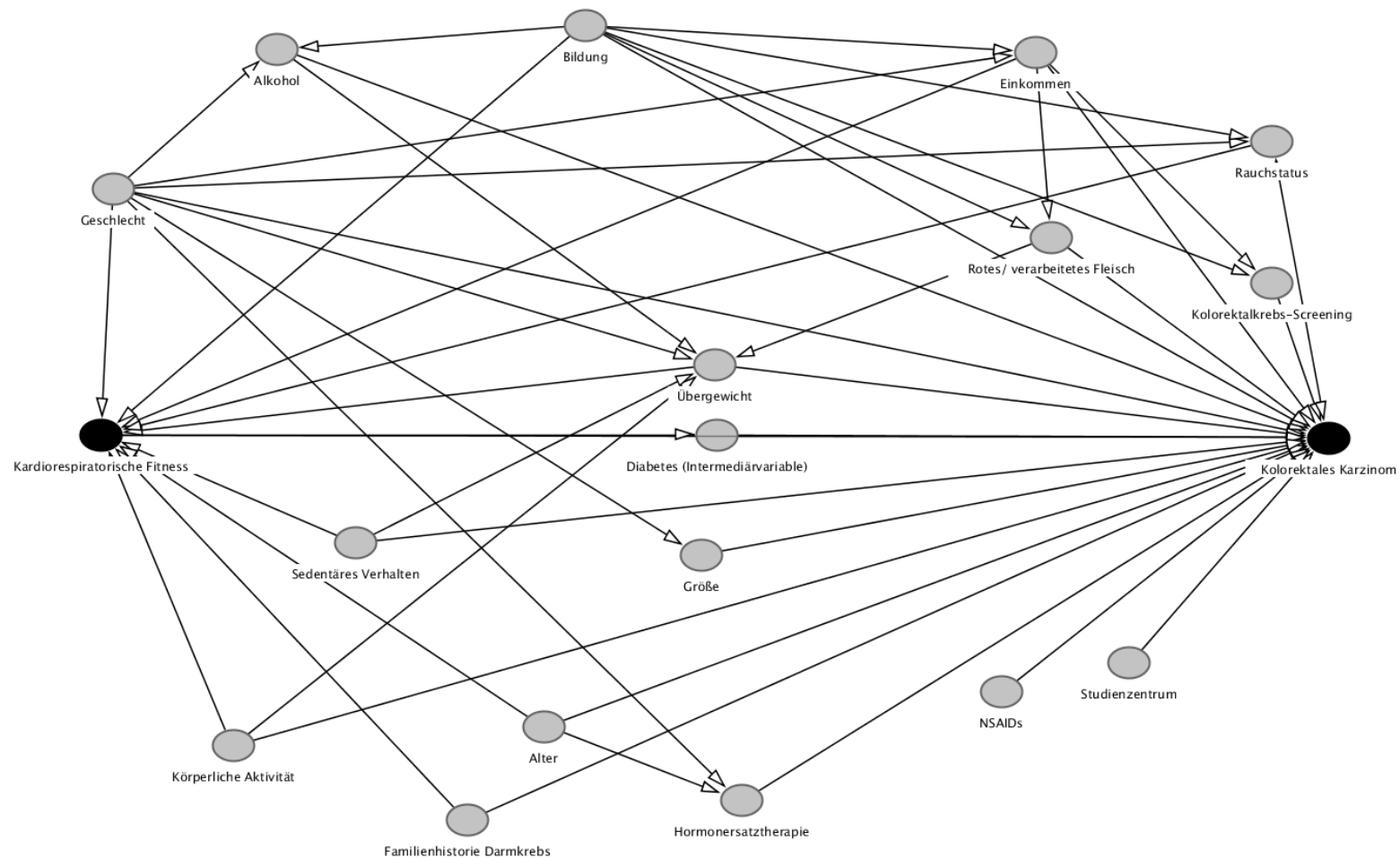


Abbildung S 4: Gerichteter Azyklischer Graph (DAG) zur Repräsentation von kausalen und verzerrenden Pfaden zwischen kardiorespiratorischer Fitness und kolorektalem Karzinom

NSAID = nichtsteroidales Antiphlogistikum

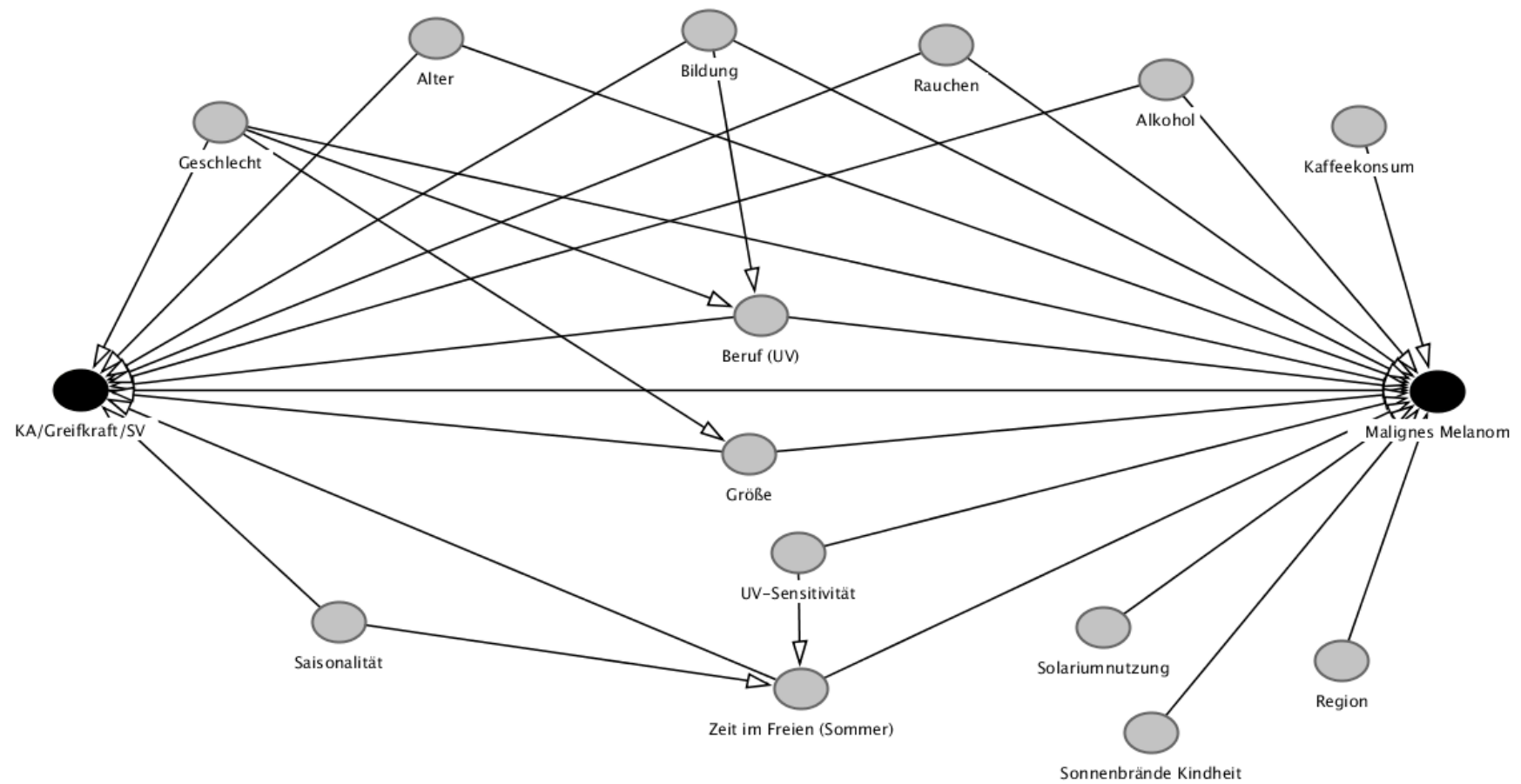


Abbildung S 5: Gerichteter Azyklischer Graph (DAG) zur Repräsentation von kausalen und verzerrenden Pfaden zwischen körperlicher Aktivität, Greifkraft, sedentärem Verhalten und malignem Melanom

KA = Körperliche Aktivität, SV = Sedentäres Verhalten, UV = Ultraviolettstrahlung

Tabelle S 4: Gewichtungsschema für die Erstellung des UV-Sensitivitätsindex

Hautfarbe ohne Bräunung	
Schwarz, braun	0
Hell-, dunkeloliv	1
Hell	2
Sehr hell	3
Natürliche Haarfarbe vor dem Ergrauen	
Schwarz, dunkelbraun, andere	0
Hellbraun	1
Blond, rot	2
Hautbräunung	
Stark gebräunt	0
Mäßig gebräunt	1
Leicht oder gelegentlich gebräunt	2
Nie gebräunt, nur Verbrennungen	3

Tabelle S 5: Plausibilität des UV-Sensitivitätsindex

UV-Sensitivitäts- index	Durchschnittliche Anzahl an Sonnenbränden in der Kindheit	Inzidenz malignes Melanom pro 10.000 Personenjahre
0	0,3	0,7
1	1,0	2,0
2	1,3	2,9
3	1,6	4,8
4	1,7	5,6
5	2,0	6,5
6	2,5	6,4
7	3,1	7,4
8	4,4	8,0

Tabelle S 6: Liste von Berufen mit natürlicher oder künstlicher UV-Exposition

SOC	Berufsbezeichnung	SOC	Berufsbezeichnung
	Konstruktion		Transport
5214	Blechbearbeiter, Schiffszimmerleute, Niettechniker	3513	Schiffs- und Hovercraft-Offiziere
5311	Stahlbauschlosser	8211	Lkw-Fahrer
5312	Maurer, Steinmetze	8212	Transporter-Fahrer
5313	Dachdecker, Schieferdecker	8213	Busfahrer und Reisebusfahrer
5315	Schreiner und Tischler	8214	Taxifahrer und Chauffeure
5316	Glaser, Fensterbauer und -monteure	8215	Fahrlehrer
5319	Bauhandwerk a.n.g.	8216	Mitarbeiter im Schienenverkehr
5321	Stuckateure	8217	Seeleute (Handelsmarine)
5322	Bodenleger und Wandverleger	8218	Mitarbeiter im Luftverkehr
5323	Maler und Dekorateure	8219	Transportarbeiter, a.n.g.
8141	Gerüstbauer, Monteure, Takler		Sonstige
8142	Arbeiter im Straßenbau	3449	Sport- und Fitnessberufe, a.n.g.
8143	Arbeiter im Eisenbahnbau und -instandhaltung	3552	Landschaftspfleger und Parkwächter
8149	Bauarbeiter, a.n.g.	5243	Leitungsinstandsetzer und Kabellöter
8221	Kranführer	6139	Tierpflegeberufe, a.n.g.
8229	Fahrer und Bediener von mobilen Maschinen, a.n.g.	6211	Sport- und Freizeitassistenten
9121	Arbeiter im Baugewerbe und in holzverarbeitenden Berufen	7124	Markt- und Straßenhändler und Assistenten
9129	Arbeiter in anderen Bauberufen, a.n.g.	8123	Steinbrucharbeiter und verwandte Berufe
	Landwirtschaft/Gartenbau/Forstwirtschaft	8222	Gabelstaplerfahrer
5111	Landwirte	9211	Postangestellte, -sortierer, -boten, Kuriere
5112	Gärtnerische Berufe	9226	Freizeit- und Themenparkmitarbeiter
5113	Gärtner	9231	Fensterputzer
5119	Landwirtschaftliche und fischereiliche Berufe, a.n.g.	9232	Straßenkehrer
8223	Landmaschinenführer	9235	Berufe in Müll- und Abfallverwertung
9111	Landarbeiter		Künstliche UV-Exposition
9112	Forstwirtschaftliche Berufe	5211	Schmiede und Schmiedemeister
9119	Fischerei- und landwirtschaftsbezogene Berufe a.n.g.	5213	Blechbearbeiter
	Wehrdienst/ Polizei/ Feuerwehr	5215	Schweißtechniker
3311	Unteroffiziere und andere Dienstgrade	5221	Einrichter und Bediener in der Metallverarbeitung
3312	Polizeibeamte (Wachtmeister und darunter)	5222	Werkzeugmacher, Werkzeugmonteure und Auszeichner
3313	Feuerwehrbeamte (leitender Feuerwehrbeamter und darunter)	5223	Produktions- und Wartungsmonteure in der Metallverarbeitung
		5314	Installateure, Heizungs- und Lüftungstechniker
		8132	Monteure (Fahrzeuge und Metallwaren)

SOC = Standardisierte Berufsklassifikation; a.n.g. = anderweitig nicht genannt

Berufsbezeichnungen wurden aus dem Englischen ins Deutsche übersetzt und sind zur besseren Lesbarkeit nur in der männlichen Form angeführt. Dies soll keinesfalls eine Geschlechterdiskriminierung oder Verletzung des Gleichheitsgrundsatzes zum Ausdruck bringen.

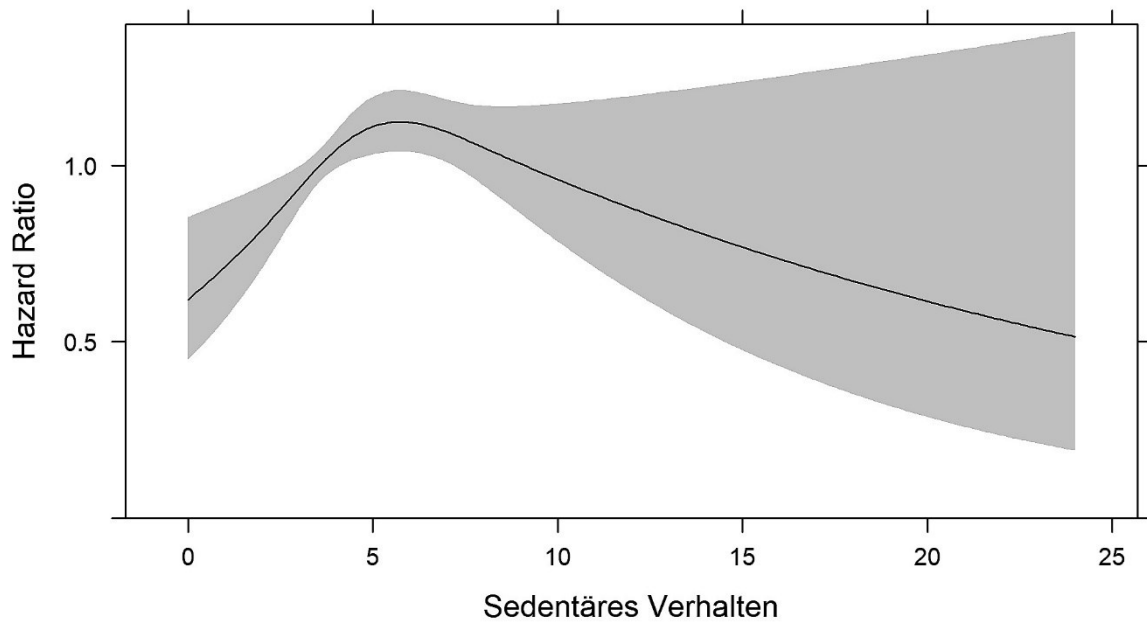
7.3 Zusätzliche Analysen

Tabelle S 7: Complete Case Analyse: Assoziation zwischen körperlicher Aktivität, Greifkraft, sedentärem Verhalten und der Inzidenz von malignem Melanom

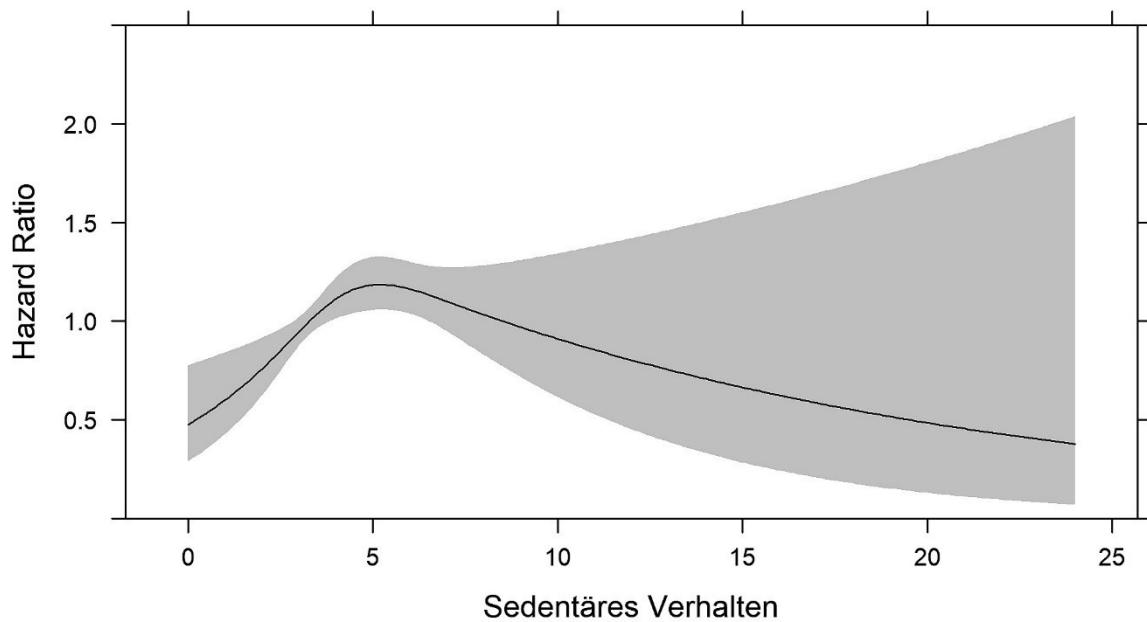
	Malignes Melanom			Malignes Melanom (chronisch UV-exponierte Haut)			Malignes Melanom (sporadisch UV-exponierte Haut)		
	Fälle	HR (95 % KI)	p-Wert	Fälle	HR (95 % KI)	p-Wert	Fälle	HR (95 % KI)	p-Wert
Körperliche Aktivität									
Gesamt	776	1,06 (0,99-1,15)	0,096	314	1,04 (0,92-1,17)	0,571	445	1,09 (0,99-1,19)	0,082
Frauen	346	1,09 (0,97-1,21)	0,147	134	1,09 (0,91-1,30)	0,369	205	1,11 (0,96-1,27)	0,159
Männer	430	1,05 (0,95-1,17)	0,328	180	1,00 (0,85-1,19)	0,957	240	1,08 (0,95-1,23)	0,242
Greifkraft									
Gesamt	776	1,15 (0,97-1,37)	0,098	314	1,08 (0,83-1,40)	0,586	445	1,17 (0,93-1,46)	0,170
Frauen	346	1,00 (0,85-1,17)	0,978	134	0,92 (0,72-1,18)	0,522	205	1,03 (0,84-1,26)	0,780
Männer	430	1,15 (1,00-1,33)	0,044	180	1,13 (0,91-1,40)	0,266	240	1,16 (0,96-1,40)	0,121
Sedentäres Verhalten									
Gesamt	776	1,02 (0,93-1,12)	0,674	314	1,04 (0,89-1,20)	0,640	445	1,01 (0,89-1,14)	0,887
Frauen	346	1,08 (0,95-1,22)	0,256	134	1,15 (0,94-1,40)	0,177	205	1,04 (0,88-1,23)	0,664
Männer	430	0,98 (0,87-1,11)	0,771	180	0,96 (0,80-1,17)	0,705	240	0,99 (0,85-1,16)	0,929

Cox proportional hazards Regression mit beobachteten Expositionen und Kovariablen (Complete-Case Analyse; Gesamt n = 222.000; Frauen n = 111.444; Männer n = 110.556). Expositionen (Körperliche Aktivität, Greifkraft, sedentäres Verhalten) wurden als kontinuierliche Variablen ins Modell aufgenommen. Assoziationen wurden linear modelliert und HRs wurden für einen Anstieg um einen Interquartilsabstand (z.B. 25. bis 75. Quantile) berechnet.

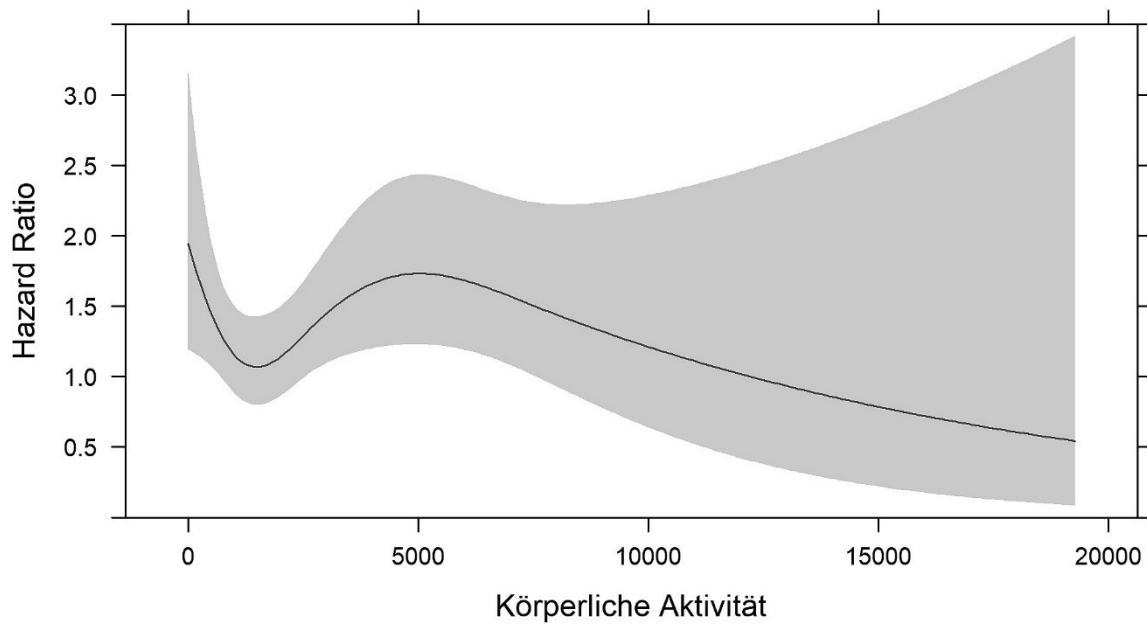
Die Modelle wurden für Alter (als Zeitachse), Studienregion (Stratifikation der Baseline-Hazard), Geschlecht (in der gesamten Kohorte), Bildung, Packungsjahre, Alkoholkonsum, Kaffeekonsum, Körpergröße, draußen verbrachte Zeit im Sommer, Sonnenbrände in der Kindheit, Solariumnutzung, Saisonalität und UV-Sensitivitäts-Score adjustiert. Die Modelle wurden zusätzlich gegenseitig für körperliche Aktivität, Greifkraft und sedentäres Verhalten adjustiert.



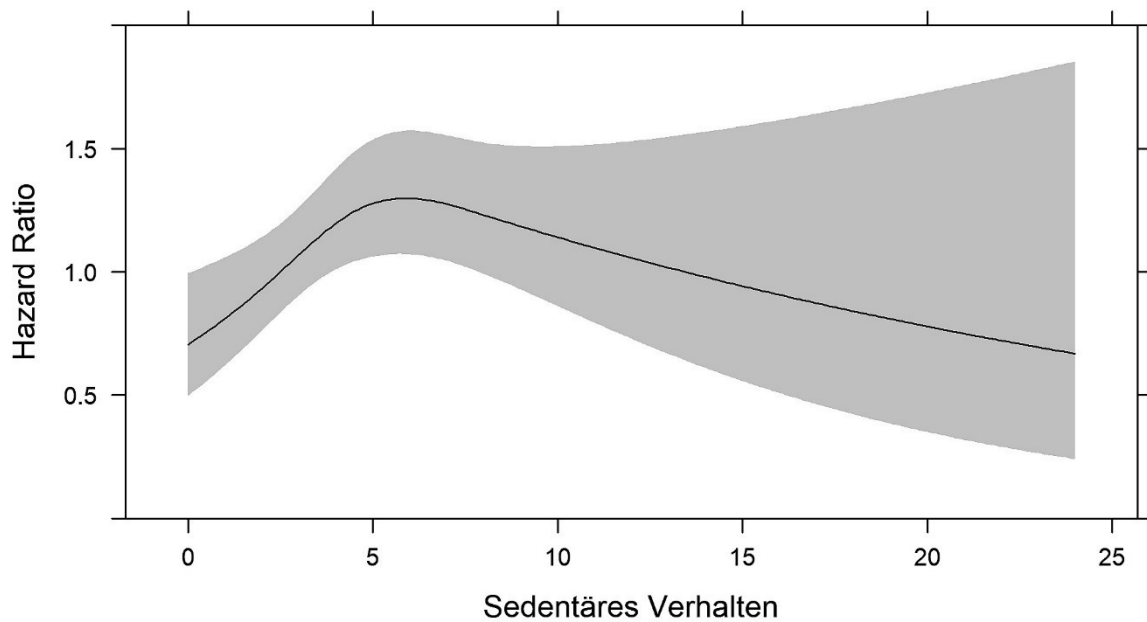
A, Basismodell: Sedentäres Verhalten und Risiko von malignem Melanom an sporadisch UV-exponierten Hautstellen in der gesamten Gruppe (n = 350.512)



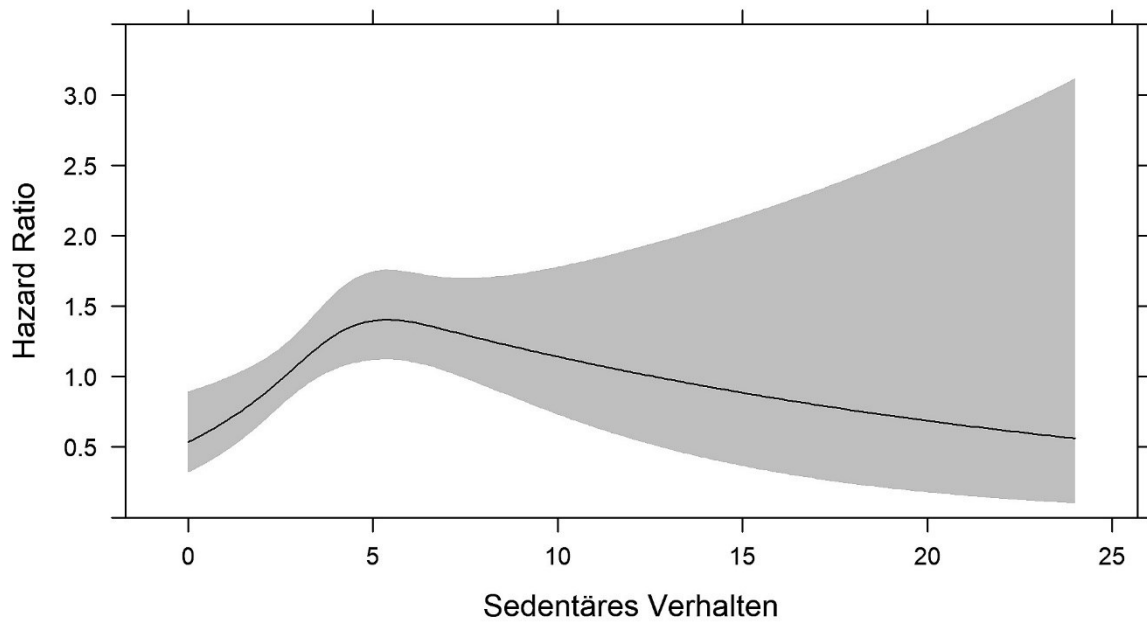
B, Basismodell: Sedentäres Verhalten und Risiko von malignem Melanom an sporadisch UV-exponierten Hautstellen in der Teilstichprobe der Frauen (n = 178.340)



C, Volles Modell: Körperliche Aktivität und Risiko von malignem Melanom an chronisch UV-exponierten Hautstellen in der Teilstichprobe der Frauen (n = 178.340)



D, Volles Modell: Sedentäres Verhalten und Risiko von malignem Melanom an sporadisch UV-exponierten Hautstellen in der gesamten Gruppe (n = 350.512)



D, Volles Modell: Sedentäres Verhalten und Risiko von malignem Melanom an sporadisch UV-exponierten Hautstellen in der gesamten Gruppe (n = 350.512)

Abbildung S 6: Nichtlineare Zusammenhänge zwischen körperlicher Aktivität (C), sedentärem Verhalten (A, B, D, E) und malignem Melanom

7.4 Auszüge R-Code

7.4.1 Auszüge R Code – Aufbereitung GPAQ

```
# 1. Identifizierung invalider GPAQ Einträge nach dem GPAQ Analyse Guide

# Durchzuführen für alle Subdomänen, beispielhaft für intensive körperliche
# Aktivität bei der Arbeit:
GPAQ$d_p2_d[GPAQ$d_p2_d %in% c(8888, 9999, 8880, 9990) | is.na(GPAQ$d_p2_d)] <- 0
GPAQ$d_p3[GPAQ$d_p3 %in% c(8888, 9999, 8880, 9990) | is.na(GPAQ$d_p3)] <- 0
GPAQ$p2c1n <- 2
GPAQ$p2c1n[(GPAQ$d_p1==1 & GPAQ$d_p2_d %in% c(1:7)) | (GPAQ$d_p1==2 &
GPAQ$d_p2_d==0)] <- 1
GPAQ$p3c1n <- 2
GPAQ$p3c1n[(GPAQ$p2c1n==1 & GPAQ$d_p2_d %in% c(1:7) & GPAQ$d_p3 %in% c(10:960)) |
(GPAQ$p2c1n==1 & GPAQ$d_p2_d==0 & GPAQ$d_p3==0)] <- 1
GPAQ$p1t3c1n <- 2
GPAQ$p1t3c1n[GPAQ$p3c1n==1 | (is.na(GPAQ$d_p1) & GPAQ$d_p2_d==0 & GPAQ$d_p3==0)]
<- 1

# Falls eine Subdomäne invalide ist oder alle Subdomänen fehlen, ist der gesamte
# Eintrag zur körperlichen Aktivität invalide und von den Analysen auszuschließen
GPAQ$c1n <- 2
GPAQ$c1n[(GPAQ$p1t3c1n==1 & GPAQ$p4t6c1n==1 & GPAQ$p7t9c1n==1 & GPAQ$p10t12c1n==1
& GPAQ$p13t15c1n==1) & (!is.na(GPAQ$d_p1) | !is.na(GPAQ$d_p4) | !is.na(GPAQ$d_p7)
| !is.na(GPAQ$d_p10) | !is.na(GPAQ$d_p13))] <- 1

# Sedentäres Verhalten
GPAQ$p16c1n <- 2
GPAQ$p16c1n[GPAQ$d_p16>=0 & GPAQ$d_p16<1441 & !is.na(GPAQ$d_p16)] <- 1

# 2. Ableitung neuer Variablen nach dem GPAQ Analyse Guide

# Minuten pro Woche je Subdomäne, beispielhaft für intensive körperliche Aktivität
# bei der Arbeit:
GPAQ[which(GPAQ$p1t3c1n==1), "p1t3"] <- GPAQ[which(GPAQ$p1t3c1n==1), "d_p2_d"] *
GPAQ[which(GPAQ$p1t3c1n==1), "d_p3"]

# Gesamte Minuten körperliche Aktivität pro Woche und pro Tag
GPAQ$ptotal <- GPAQ$p1t3 + GPAQ$p4t6 + GPAQ$p7t9 + GPAQ$p10t12 + GPAQ$p13t15
GPAQ$ptotalday <-GPAQ$ptotal/7

# MET-Minuten pro Woche je Subdomäne, beispielhaft für intensive körperliche
# Aktivität bei der Arbeit:
GPAQ[which(GPAQ$p1t3c1n==1), "p1t3met"] <- GPAQ[which(GPAQ$p1t3c1n==1), "p1t3"] *
8

# Gesamte MET-Minuten pro Woche und Einhaltung der WHO-Empfehlungen
GPAQ$ptotalmet <- GPAQ$p1t3met + GPAQ$p4t6met + GPAQ$p7t9met + GPAQ$p10t12met +
GPAQ$p13t15met
GPAQ$meetrec[GPAQ$ptotalmet >= 600] <- 1
GPAQ$meetrec[GPAQ$ptotalmet < 600] <- 2

# Vorkommen von körperlicher Aktivität in Domänen und Anteil an Gesamtaktivität,
# beispielhaft für körperliche Aktivität bei der Arbeit
GPAQ$work[GPAQ$d_p1 == 1 | GPAQ$d_p4 == 1] <- 1
GPAQ$percent_work <- (GPAQ$p1t3 + GPAQ$p4t6)/GPAQ$ptotal*100
```

```
# Bei Teilnehmenden mit insgesamt invaliden Angaben werden alle abgeleiteten
Variablen auf NA gesetzt
GPAQ[which(GPAQ$c1n==2), 38:61] <- NA
```

```
# Invalide Einträge von sedentärem Verhalten werden auf NA gesetzt
GPAQ$p16 <- GPAQ$d_p16
GPAQ$p16[GPAQ$p16c1n == 2] <- NA
```

7.4.2 Auszüge R Code – Aufbereitung cpar24

```
# 1. Ausschluss von Teilnehmenden mit cpar24 Kennung == NULL (27356), mehrfach
vergebenen Zugängen und ohne Einträge
ds1 <- subset(ds0, Users.login != 27356)
duplpcpar$count <- ave(duplpcpar$ID, duplpcpar$Users.login, FUN=seq_along)
ds2 <- anti_join(ds1, duplpcpar[duplpcpar$count==2, c("Users.login", "count")],
by="Users.login")
ds3 <- subset(ds2, Recallentries.start != "\\N")
```

```
# 2. Markierung wenn cpar24 ausserhalb Akzelerometer Zeitraum (2-8 Tage nach SZ
Besuch)
ds4$cpardate <- substring(as.character(ds4$Recalls.DOM), 1,10)
plotzeitnachs$diff <- plotzeitnachs$cpardate - plotzeitnachs$SZdateall
cpar24zeitraum <- subset(plotzeitnachs, ausschluss8days %in% c(2:8))
```

```
# 3. Bei Vergabe mehrerer Zugänge nur den Aktuellsten behalten:
mult$orderdate <- substring(as.character(mult$Recalls.DOM), 1,10)
mult1 <- mult[order(mult$ID, mult$orderdate, decreasing = TRUE) , ]
mult1$count <- ave(mult1$Users.login, mult1$ID, FUN=seq_along)
mult2 <- left_join(mult1[mult1$count==2, c("ID", "count")], mult1, by="ID")
```

```
# 4. Positionsangaben umcodieren
dsfin1$position <- NA
dsfin1$position[dsfin1$Scaledata.value == "0 %" | dsfin1$Scaledata.value ==
"Stehen"] <- 0
dsfin1$position[dsfin1$Scaledata.value == "100 %" | dsfin1$Scaledata.value ==
"Sitzen"] <- 100
dsfin1$position[dsfin1$Scaledata.value == "20 %"] <- 20
# wenn keine Info zu Position: Annahme niedrigster MET Wert!
dsfin1$position[is.na(dsfin1$position)] <- 100
dsfin1$position <- as.character(dsfin1$position)
```

```
# 5. Zuordnung der korrekten MET-Werte bei Aktivitäten mit und ohne Intensitäts-/
Positionsabfrage
```

```
dsfinpos <- subset(dsfin1, Activities.pk_activities %in%
c(107,81,105,363,18,321,101,235,120,110,364,91))
dsfinint <- subset(dsfin1, Activities.pk_activities %in% c(84,70,19,42,67,
227,71,121,217,111,147,54,353,223,318,177,126,179,123,32, 190,178))
dsfinohne <- dsfin1[!(dsfin1$Activities.pk_activities %in% c(107,81,105,
363,18,321,101,235,120,110,364,91,84,70,19,42,67,227,71,121,217,111,147,
54,353,223,318,177,126,179,123,32,190,178)), ]
dsfinohne_met <- left_join(dsfinohne, actlist24V2[ , c("Activities.pk_activities",
"met")], by="Activities.pk_activities")
dsfinpos_met <- left_join(dsfinpos, actlist24[ , c("Activities.pk_activities",
"met", "position")], by=c("Activities.pk_activities", "position"))
dsfinint_met <- left_join(dsfinint, actlist24[ , c("Activities.pk_activities",
"met", "intensity")], by=c("Activities.pk_activities", "intensity"))
dsfin2 <- rbind(dsfinohne_met, dsfinpos_met, dsfinint_met)
```



```
# 6. Ableitung von Variablen für das Data Dictionary

# Berechnung MET-Minuten
checkds2a$a_metmin_final <- checkds2a$a_met_final * checkds2a$a_durkorr
# Einteilung der Intensitätskategorien
checkds2a$a_intkat[checkds2a$a_met_final <= 1.5] <- 1
checkds2a$a_intkat[checkds2a$a_met_final > 1.5 & checkds2a$a_met_final < 3.0] <- 2
checkds2a$a_intkat[checkds2a$a_met_final >= 3.0 & checkds2a$a_met_final < 6.0] <-
3
checkds2a$a_intkat[checkds2a$a_met_final >= 6.0] <- 4
checkds2a$a_intkat[checkds2a$a_pk_final %in% c(2,5)] <- 0
checkds2a$a_intkat[checkds2a$a_pk_final %in% c(78, 216)] <- 2

# Plausibilitätskontrolle fuer Eintraege mit sehr wenigen Aktivitaeten
checkds3$one <- 1
response1 <- aggregate(checkds3[ , "one"], by=list(Users.login=
checkds3$Users.login), FUN=sum)

# Vollständigkeit (Ausgefüllte Minuten pro Tag)
vollstaendigkeit1 <- aggregate(checkds3[ , "a_durkorr"],
by=list(Users.login=checkds3$Users.login), FUN=sum)
```

7.4.3 Auszüge R Code – Deskriptive Plots Akzelerometrie

```
# Laden von R Paketen

library("irr")
library("dplyr")
library(tidyr)
library(ggplot2)
library(RColorBrewer)

#####

# Mindestanzahl an validen Tagen

# 1. Teildatensatz mit 7 konsekutiven validen Tragetagern (> 16 h/d) als
"Grundwahrheit"

day1$calendar_date1 <- as.Date(as.character(day1$calendar_date))
day2 <- day1[with(day1, order(ID, calendar_date1)), ]
day2$count <- ave(day2$ID, day2$ID, FUN=seq_along)
ID7 <- day2[day2$count==8, "ID"]
day2a <- day2[day2$ID %in% ID7 & day2$count %in% c(2:8), ]
day2b <- day2a[day2a$N.valid.hours >= 16 | !is.na(day2a$mean_ENMO_mg_0.24hr) |
!is.na(day2a$mean_MAD_mg_0.24hr), ]
day2b$count1 <- ave(day2b$ID, day2b$ID, FUN=seq_along)
IDvalid7 <- day2b[day2b$count1==7, "ID"]
day3 <- day2b[day2b$ID %in% IDvalid7, ]

# 2. Berechnung des durchschnittlichen ENMO für zufällig gezogene Kombinationen
aus 1-6 Tagen

day3$daynr <- paste0("day", day3$count1)
day4 <- day3[ , c("ID", "daynr", "mean_ENMO_mg_0.24hr")]
```

```

day4w <- spread(day4, daynr, mean_ENMO_mg_0.24hr)
IDlist <- day4w$ID
testDS <- data.frame(day4w$ID)
testDS$Mean2 <- NA; testDS$Mean3 <- NA; testDS$Mean4 <- NA; testDS$Mean5 <-
NA; testDS$Mean6 <- NA

set.seed(22222)
for (j in 2:6){
  for(i in IDlist){
    M <- day4w[day4w$ID == i, sample(2:8, 6)]
    day4w[day4w$ID == i, "random1"] <- M[ , 1]
    day4w[day4w$ID == i, "random2"] <- M[ , 2]
    day4w[day4w$ID == i, "random3"] <- M[ , 3]
    day4w[day4w$ID == i, "random4"] <- M[ , 4]
    day4w[day4w$ID == i, "random5"] <- M[ , 5]
    day4w[day4w$ID == i, "random6"] <- M[ , 6]
  }
  rm(M)
  testDS[ , j] <- rowMeans(day4w[ , 9:(8+j)])
  day4w$random1 <- NA; day4w$random2 <- NA; day4w$random3 <- NA; day4w$random4 <-
NA; day4w$random5 <- NA; day4w$random6 <- NA
}

# ein zufällig gezogener Tagesmittelwert
testDS$Mean1 <- NA
for(i in IDlist){
  testDS[testDS$day4w.ID == i, "Mean1"] <- day4w[day4w$ID == i, sample(2:8, 1)]
}

# Mittelwert der 7 Tage ("Grundwahrheit")
testDS$Mean7 <- NA
testDS[ , "Mean7"] <- rowMeans(day4w[ , 2:8])

# 3. Berechnung des Intraklassenkorrelations-Koeffizienten für die Kombinationen
aus 1-6 Tagen versus den "wahren" 7-Tages-Mittelwert
icc17 <- icc(cbind(testDS$Mean7, testDS$Mean1), model = "oneway", type =
"consistency", unit = "average", r0 = 0, conf.level = 0.95)
icc127 <- icc(cbind(testDS$Mean7, testDS$Mean2), model = "oneway", type =
"consistency", unit = "average", r0 = 0, conf.level = 0.95)
icc137 <- icc(cbind(testDS$Mean7, testDS$Mean3), model = "oneway", type =
"consistency", unit = "average", r0 = 0, conf.level = 0.95)
icc147 <- icc(cbind(testDS$Mean7, testDS$Mean4), model = "oneway", type =
"consistency", unit = "average", r0 = 0, conf.level = 0.95)
icc157 <- icc(cbind(testDS$Mean7, testDS$Mean5), model = "oneway", type =
"consistency", unit = "average", r0 = 0, conf.level = 0.95)
icc167 <- icc(cbind(testDS$Mean7, testDS$Mean6), model = "oneway", type =
"consistency", unit = "average", r0 = 0, conf.level = 0.95)
icc177 <- icc(cbind(testDS$Mean7, testDS$Mean7), model = "oneway", type =
"consistency", unit = "average", r0 = 0, conf.level = 0.95)

days <- c(1,2,3,4,5,6,7)
icc <- c(icc17$value, icc127$value, icc137$value, icc147$value, icc157$value,
icc167$value ,icc177$value )
ll <- c(icc17$lbound, icc127$lbound, icc137$lbound, icc147$lbound, icc157$lbound,
icc167$lbound ,icc177$lbound)
ul <- c(icc17$ubound, icc127$ubound, icc137$ubound, icc147$ubound, icc157$ubound,
icc167$ubound ,icc177$ubound)
ab <- data.frame(days, icc, ll, ul )

```

```

# 4. Graphische Darstellung
options(OutDec= ",")
ggplot(ab, aes(days, icc)) + geom_line() + theme_bw() + geom_hline(yintercept=0.9,
col="black") +
  labs(x="Tage", y="Intraklassen-Korrelationskoeffizient") +
  theme(axis.text=element_text(size=14), axis.title=element_text(size=14),
legend.text=element_text(size=14), legend.title=element_text(size=14)) +
  scale_x_continuous(breaks = round(seq(1,7, by = 1),1)) +
  geom_line(aes(days, ll), linetype="dotted") + geom_line(aes(days, ul),
linetype="dotted")

#####

# Orientierung der vertikalen Achse über den Tag

# Anzahl pro Zeitpunkt, Alterskategorie und Geschlecht (für Berechnung des
Standardfehlers)
timeseries2$one <- 1
timeseries_number <- aggregate(one ~ time * gender_d * age_d, timeseries2, sum)

# Mittelwert und Standardabweichung der Achsenorientierung pro Zeitpunkt,
Alterskategorie und Geschlecht
timeseries_aggAngle <- aggregate(angle_long_axis ~ time * gender_d * age_d,
timeseries2, mean)
timeseries_sdaxis <- aggregate(angle_long_axis ~ time * gender_d * age_d,
timeseries2, sd)
colnames(timeseries_sdaxis)[4] <- "angle_SD"
timeseries_aggAngle1 <- left_join(timeseries_aggAngle, timeseries_sdaxis,
by=c("time", "age_d", "gender_d"))
timeseries_aggAngle2 <- left_join(timeseries_aggAngle1, timeseries_number,
by=c("time", "age_d", "gender_d"))

# SE = SD / sqrt(n)
timeseries_aggAngle2$angle_SE <- timeseries_aggAngle2$angle_SD /
sqrt(timeseries_aggAngle2$one)

# Graphische Darstellung:
ggplot(timeseries_aggAngle2, aes(time, angle_long_axis, group=age_d, color=age_d,
ymin=angle_long_axis-angle_SE, ymax=angle_long_axis+angle_SE)) +
geom_ribbon(aes(fill=age_d), alpha=0.1, linetype="dashed") +
  geom_line(size=1.2) + theme_bw() + facet_grid(.~gender_d) +
  labs(x="Tageszeit", y="Orientierung der vertikalen Achse") + theme(legend.title
= element_blank()) +
  theme(axis.text=element_text(size=15), strip.text.x = element_text(size = 15),
axis.title=element_text(size=15), legend.text=element_text(size=15)) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) +
  scale_x_continuous(breaks = seq(0 , 24, by = 1)) + theme(legend.position="bottom")

```

7.4.4 Auszüge R Code – Kardiorespiratorische Fitness und Risiko von kolorektalem Karzinom in der UK Biobank

```
#####

# Laden von R Paketen

library("dplyr")
library("tidyr")
library(reshape)
library(data.table)
library(mice)
library(VIM)
library(lattice)
library(ggplot2)
library(gridExtra)
library(rms)
library(parallel)

#####

###
# Aufbereitung der Kardiorespiratorischen Fitness - Auszüge
###

# Transponieren in Langes Format am Beispiel der Herzfrequenz (Variablen
f.5983.0.0-f.5983.0.111)
sub1 <- check3[ , 1:113]
shapesub1 <- melt(sub1, id=c("f.eid"))
shapesub1a <- shapesub1[!is.na(shapesub1$value) , ]
shapesub1a$no_occ <- substring(shapesub1a$variable, 8)
colnames(shapesub1a)[3] <- "heartrate"
shape_heartrate <- shapesub1a[ , c(1,3,4)]
row.names(shape_heartrate) <- NULL

# Durchgeführt für alle relevanten Ergometrie Variablen (Herzfrequenz, Widerstand,
Trendphase, Stagename, Phasenzeit, Tretgeschwindigkeit) und Merge der entstehenden
Datensätze

# Ausschluss von Teilnehmenden ohne inkrementelle Fahrradergometrie und mit
fehlenden Werten bei Herzfrequenz oder Widerstand
ergoall <- subset(shape_merge6, substring(shape_merge6$stagename, 1,4) == "Ramp")
ergoallA <- ergoall[!is.na(ergoall$heartrate) & !is.na(ergoall$load), ]

# Berechnung maximale Herzfrequenz mit Tanaka Formel (208 - (0.7 * weight [kg]))
ergoallD$HR_max_age <- 208 - (0.7 * ergoallD$f.21003.0.0)

# Erstellung einer Liste an Datensätzen für die Schleife
list_increment <- split(ergoallD, f=ergoallD$id)

# Schleife für Regressionsgleichung  $lm(y \sim x)$  mit  $y = \text{Widerstand [W]}$  und  $x = \text{Herzfrequenz [bpm]}$ :  $PWC75\% = \text{intercept} + \text{slope} * \text{Heartrate75\%}$ 
```

```

laenge <- length(result$f.eid)
for (i in 1:laenge){
model <- lm(list_increment[[i]]$load ~ list_increment[[i]]$heartrate)
reg <- summary(model)
result$lm_intercept[result$id == i] <- reg$coef[1]
result$lm_slope[result$id == i] <- reg$coef[2]
Watt75P <- reg$coef[1] + result[result$id==i, "HR_75P"] * reg$coef[2]
result$Watt75P[result$id == i] <- Watt75P
}

# Standardisierung von PWC75% auf das Körpergewicht
result$PWC75P <- result$Watt75P / result$f.21002.0.0

# Ausschluss unplausibler Werte (maximaler Herzfrequenz < 40 oder > 220 bpm,
negative Steigung oder PWC75P, <2 Datenpunkte während inkrementeller Phase)
flow2 <- subset(flow1, f.6033.0.0 > 40 & f.6033.0.0 < 220 & !is.na(f.6033.0.0) &
PWC75P >= 0 & lm_slope >= 0 & !is.na(PWC75P) & !is.na(PWC75P) & rampdatapoints >=
2 & !is.na(rampdatapoints))

# Winsorisierung der kardiorespiratorischen Fitness im Analyse-Datensatz (z.B.
nach Ausschluss von prävalenten malignen Karzinomen etc.):
ccds1$PWC75Psw <- ccds1$PWC75P
ccds1$PWC75Psw[ccds1$f.31.0.0==1 & ccds1$PWC75Psw < 0.5669172] <- 0.5669172
ccds1$PWC75Psw[ccds1$f.31.0.0==1 & ccds1$PWC75Psw > 3.3601982] <- 3.3601982
ccds1$PWC75Psw[ccds1$f.31.0.0==0 & ccds1$PWC75Psw < 0.4425088] <- 0.4425088
ccds1$PWC75Psw[ccds1$f.31.0.0==0 & ccds1$PWC75Psw > 2.5169007] <- 2.5169007

#####

###
# Aufbereitung der Kolorektalen Karzinom Diagnosen - Auszüge
###

# Transponieren in Langes Format am Beispiel des Diagnosedatums (Variablen
f.40005.0.0 - f.40005.31.0)
shape4 <- melt(test4, id=c("f.eid"))
shape4a <- shape4[!is.na(shape4$value) , ]
shape4a$no_occ <- substring(shape4a$variable, 9)
shape4a$no <- sub("\\.\\.*", "", shape4a$no_occ)
shape4a$no_numeric <- as.numeric(shape4a$no)
colnames(shape4a)[3] <- "date_diagnose"
shape_date <- shape4a[ , c(1,3,4,6)]
row.names(shape_date) <- NULL

# Durchzuführen für alle relevanten Krebsvariablen (Diagnosedatum, ICD9, ICD10,
Behaviour, Histologie, Datenquelle, Todeszeitpunkt) und Merge dieser Datensätze

# Ausschluss aller Teilnehmenden mit mindestens einer malignen Krebsdiagnose vor
der Baseline (Ausnahme nicht-melanozytärer Hautkrebs (C44))
merge5$frage1 <- 0; merge5$frage2 <- 0; merge5$frage3 <- 0; merge5$einschluss <- 0
merge5$frage1[merge5$date_diagnose > merge5$f.53.0.0] <- 1

```

```

merge5$frage2[merge5$behaviour == 0] <- 1
merge5$frage3[(merge5$ICD9_3 == 173 | (merge5$ICD10_letter == "C" &
merge5$ICD10_code == 44))
merge5$einschluss[merge5$frage1 == 1] <- 1
merge5$einschluss[merge5$frage1 == 0 & merge5$frage2 == 1] <- 1
merge5$einschluss[merge5$frage1 == 0 & merge5$frage3 == 1] <- 1
raus <- subset(merge5, einschluss == 0)
merge6 <- anti_join(merge5, raus, by="f.eid")

# nur primary malignant oder microinvasive cancer berücksichtigen
merge7 <- merge6[merge6$behaviour %in% c(3,5), ]

# Codierung Kolorektalkarzinom nach Morris 2016
# colon allg: C18, C18.8, C18.9      1
# colon proximal: C18.0 - C18.5    2
# colon distal: C18.6, C18.7      3
# rectum: C19, C20                4
merge7$site_morris <- 0
merge7$site_morris[merge7$ICD10 %in% c("C18", "C188", "C189")] <- 1
merge7$site_morris[merge7$ICD10 %in% c("C180", "C181", "C182", "C183", "C184", "C185")] <- 2
merge7$site_morris[merge7$ICD10 %in% c("C186", "C187")] <- 3
merge7$site_morris[merge7$ICD10 %in% c("C19", "C20")] <- 4

# Identifizierung 1. Datum mit primary malignant/ microinvasive CRC Diagnose;
multiple CRC Diagnosen nicht für Subsite Analysen verwenden
merge10a <- merge10[order(merge10$f.eid, merge10$datum), ]
firstdateds <- unique(merge10a, by=c("f.eid"))
colnames(firstdateds)[4] <- "datediagnosefirstCRC"
colnames(firstdateds)[19] <- "datumfirstCRC"
merge11 <- left_join(merge10a, firstdateds[ , c("f.eid", "datediagnosefirstCRC",
"datumfirstCRC")], by="f.eid")
merge12 <- merge11[merge11$datum == merge11$datumfirstCRC , ]
merge12$count <- ave(merge12$f.eid, merge12$f.eid, FUN=seq_along)
multipliediag <- subset(merge12, count != 1)
multipliediag$multiple <- 1
multipliediag1 <- multipliediag[!duplicated(multipliediag$f.eid), ]
merge13 <- left_join(merge12, multipliediag1[ , c("f.eid", "multiple")],
by="f.eid")

# Erstellung eines Datensatzes mit einer Zeile pro Teilnehmendem
merge13a <- merge13[ , c("f.eid", "date_diagnose", "ICD10", "behaviour", "einschluss",
"site_morris", "datediagnosefirstCRC", "multiple")]
merge13a$site_morris[merge13a$site_morris == 1] <- "allg"
merge13a$site_morris[merge13a$site_morris == 2] <- "prox"
merge13a$site_morris[merge13a$site_morris == 3] <- "dist"
merge13a$site_morris[merge13a$site_morris == 4] <- "rect"
merge14 <- spread(merge13a, site_morris, ICD10)
# Subsite-Analysen werden nur gezählt, wenn nicht an diesem Tage mehrere
Diagnosen!
merge14$colorect <- 0; merge14$colon <- 0; merge14$subrectum <- 0; merge14$suballg
<- 0; merge14$subdist <- 0; merge14$subprox <- 0

```

```

merge14$colorect[!is.na(merge14$allg) | !is.na(merge14$dist) | is.na(merge14$prox)
| !is.na(merge14$rect)] <- 1
merge14$colon[(!is.na(merge14$allg) | !is.na(merge14$dist) | !is.na(merge14$prox))
& is.na(merge14$rect)] <- 1
merge14$subrectum[!is.na(merge14$rect) & is.na(merge14$multiple)] <- 1
merge14$suballg[!is.na(merge14$allg) & is.na(merge14$multiple)] <- 1
merge14$subdist [!is.na(merge14$dist) & is.na(merge14$multiple)] <- 1
merge14$subprox[!is.na(merge14$prox) & is.na(merge14$multiple)] <- 1

# Zensierung: kompletter Follow up, Kolorektalkarzinomdiagnose oder Tod, je
nachdem was zuerst eintritt.
# Todeszeitpunkt (Variablen "f.40000.0.0", "f.40000.1.0", "f.40000.2.0")
merge17$death <- 1
merge17$death[is.na(merge17$DateCensorDeath)] <- 0

# Datum komplettes Follow-up je nach Studienregion
# Schottland -> 31 Dec 2014
# England/ Wales -> 30 Nov 2014
merge17$DateCompleteFU <- as.Date(NA)
merge17$DateCompleteFU[merge17$f.54.0.0 %in% c("11004", "11005")] <- "2014-12-31"
EngWal <- c("10003", "11001", "11002", "11003", "11006", "11007", "11008", "11009",
"11010", "11011", "11012", "11013", "11014", "11016", "11017", "11018", "11020", "11021",
"11022", "11023", "11024", "11025")
merge17$DateCompleteFU[merge17$f.54.0.0 %in% EngWal] <- "2014-11-30"

merge18 <- gather(merge17, "date_diagnose", "DateCensorDeath", "DateCompleteFU",
key="whicheverfirst", value="Datum")
merge19 <- merge18[!is.na(merge18$Datum), ]
merge19$Datumnum <- as.numeric(merge19$Datum)
merge20 <- merge19[order(merge19$f.eid, merge19$Datumnum), ]
merge21 <- merge20[!duplicated(merge20$f.eid),]

# wenn Diagnosedatum NACH komplettem Follow-up ist, zählt das Outcome nicht -> auf
0 codieren
merge21$behaviour[(merge21$colorect==1&merge21$whicheverfirst=="DateCompleteFU")]
<- NA
merge21$datediagnosefirstCRC[(merge21$colorect==1&merge21$whicheverfirst=="
DateCompleteFU")] <- NA
merge21$multiple[(merge21$colorect==1 & merge21$whicheverfirst=="DateCompleteFU")]
<- NA
merge21$colon[(merge21$colorect==1 & merge21$whicheverfirst=="DateCompleteFU")]<-0
merge21$subrectum[(merge21$colorect==1&merge21$whicheverfirst=="DateCompleteFU")]
<-0
merge21$suballg[(merge21$colorect==1&merge21$whicheverfirst=="DateCompleteFU")]<-0
merge21$subdist[(merge21$colorect==1&merge21$whicheverfirst=="DateCompleteFU")]<-0
merge21$subprox[(merge21$colorect==1&merge21$whicheverfirst=="DateCompleteFU")]<-0
merge21$colorect[(merge21$colorect==1&merge21$whicheverfirst=="DateCompleteFU")]
<-0

# Berechnung Follow-up Zeit und Alter zur Zensierung
merge22$Fuptime <- NA
merge22$Fuptime <- merge22$Datum - merge22$f.53.0.0

```

```
merge22$Ageout <- NA
merge22$Ageout <- round(merge22$f.21003.0.0 + merge22$FUttime/365.25, 2)

#####

###
# Multiple Imputation mit Chained Equations (MICE) mittels Parallelisierung auf
# Server-Rechner - Auszüge
###

# Spezifizierung der Parameter
dry <- mice(imputds8, maxit=0)
meth1 <- dry$method

# Imputationsmodelle basierend auf Variablenklasse
# numeric: pmm (default)
# binary: Logreg
# factor: polyreg
# ordinal: polr
meth1["educat4"] <- "polr"
meth1["income3"] <- "polr"
meth1["alc3"] <- "polr"
meth1["smoke3"] <- "polr"
meth1["redprocmeat"] <- "polr"
meth1["healthrate"] <- "polr"
meth1["f.2296.0.0"] <- "polr"
meth1["f.1548.0.0"] <- "polr"
meth1["ratio_group"] <- "polr"
meth1["f.924.0.0"] <- "polr"
meth1["f.943.0.0"] <- "polr"

# Prädiktormatrix: welche Variablen sollen nicht als Prädiktor verwendet werden:
pred <- dry$predictorMatrix
pred[ , "f.eid"] <- 0
pred[ , "lmslope"] <- 0
pred[ , "lmintercept"] <- 0
pred[ , "CRFmis"] <- 0
pred[ , "PWC75P"] <- 0

# Imputationsreihenfolge
dry$visitSequence

# Parallelisierung
cores_2_use <- 10

cl <- makeCluster(cores_2_use)
clusterSetRNGStream(cl, 9956)
clusterExport(cl, "imputds8")
clusterExport(cl, "meth1")
clusterExport(cl, "pred")
clusterEvalQ(cl, library(mice))
```



```

imp_pars <-
  parLapply(cl = cl, X = 1:cores_2_use, fun = function(no){
    imput_neu <- mice(imputds8, meth=meth1, maxit=5, m=1, pred=pred, print=F)
  })

stopCluster(cl)

imp_merged <- imp_pars[[1]]
for (n in 2:length(imp_pars)){
  imp_merged <-
    ibind(imp_merged,
          imp_pars[[n]])
}

alllong <- complete(imp_merged, "long", include=T)
einzeln0 <- complete(imp_merged, 0)

#####

###
# Alters-Adjustierte Charakteristika der Studienteilnehmenden zur
# Basisuntersuchung (Automatisiert) - Auszüge
###

# Geschlechtsspezifische Quartile der kardiorespiratorischen Fitness
ccds1$PWCquantsw <- NA
ccds1$PWCquantsw[ccds1$f.31.0.0==1 & ccds1$PWC75Psw <= 1.2813445] <- 1
ccds1$PWCquantsw[ccds1$f.31.0.0==1 & ccds1$PWC75Psw > 1.2813445 & ccds1$PWC75Psw
<= 1.6536759] <- 2
ccds1$PWCquantsw[ccds1$f.31.0.0==1 & ccds1$PWC75Psw > 1.6536759 & ccds1$PWC75Psw
<= 2.0730383] <- 3
ccds1$PWCquantsw[ccds1$f.31.0.0==1 & ccds1$PWC75Psw > 2.0730383] <- 4
ccds1$PWCquantsw[ccds1$f.31.0.0==0 & ccds1$PWC75Psw <= 0.9005375] <- 1
ccds1$PWCquantsw[ccds1$f.31.0.0==0 & ccds1$PWC75Psw > 0.9005375 & ccds1$PWC75Psw
<= 1.1617587 ] <- 2
ccds1$PWCquantsw[ccds1$f.31.0.0==0 & ccds1$PWC75Psw > 1.1617587 &
ccds1$PWC75Psw <= 1.4725402] <- 3
ccds1$PWCquantsw[ccds1$f.31.0.0==0 & ccds1$PWC75Psw > 1.4725402 ] <- 4
ccds1$PWCquantsw <- as.factor(ccds1$PWCquantsw)

# Alterskategorien
ccds1$age_center_q <- NULL
ccds1$age_center_q[ccds1$f.21003.0.0 <= 50] <- 1
ccds1$age_center_q[ccds1$f.21003.0.0 > 50 & ccds1$f.21003.0.0 <= 60] <- 2
ccds1$age_center_q[ccds1$f.21003.0.0 > 60 & ccds1$f.21003.0.0 <= 70] <- 3
ccds1$age_center_q <- as.factor(ccds1$age_center_q)

# A, Kategoriale Kovariablen

# Position der relevanten Kovariablen im Datensatz (Spaltennummer)
dd <- c(13,14,15,16,17,19,20,21,22,23,25,29,34,35,64,66)
out_list1 <- vector(mode = "list", length =16)

```

```

table1df <- ccds1

# n gesamt: 59191
age1 <- 16414 / 59191
age2 <- 20463 / 59191
age3 <- 22314 / 59191
age1 + age2 + age3# 1

for (i in dd){
t1 <- as.data.frame(table(table1df[,i],table1df$age_center_q,
table1df$PWCquantsw))
t2 <- aggregate(t1[ , "Freq"], by=list(Var3=t1$Var3, Var2=t1$Var2), FUN=sum)

t3 <- merge(t1, t2, by=c("Var2", "Var3"), all=T)
t3$percent_age <- t3$Freq / t3$x * 100
t3$age_dist[t3$Var2 == 1] <- age1
t3$age_dist[t3$Var2 == 2] <- age2
t3$age_dist[t3$Var2 == 3] <- age3
t3$percent_stand <- t3$percent_age * t3$age_dist
t4 <- aggregate(t3[ , "percent_stand"], by=list(Var3=t3$Var3, Var1=t3$Var1),
FUN=sum)
t4$x <- round(t4$x, 2)
t4s <- spread(t4, Var3,x)
noNA <- sum(is.na(table1df[,i]))
row <- data.frame(Variable=as.factor(colnames(table1df)[i]), ergoQ1=NA, ergoQ2=NA,
ergoQ3=NA, ergoQ4=NA, numberNA=noNA)
colnames(t4s)[1] <- "Variable"
colnames(t4s)[2] <- "ergoQ1"
colnames(t4s)[3] <- "ergoQ2"
colnames(t4s)[4] <- "ergoQ3"
colnames(t4s)[5] <- "ergoQ4"
t4s$numberNA <- NA
t4sdf <- rbind(row, t4s)
out_list1[[i]] <- t4sdf
names(out_list1)[i] <- paste("df", i, sep = "_")
}

testab <- rbind(as.data.frame(out_list1[[13]]), as.data.frame(out_list1[[14]]),
as.data.frame(out_list1[[15]]), as.data.frame(out_list1[[16]]),
as.data.frame(out_list1[[17]]), as.data.frame(out_list1[[19]]),
as.data.frame(out_list1[[20]]),as.data.frame(out_list1[[21]]),
as.data.frame(out_list1[[22]]),as.data.frame(out_list1[[23]]),
as.data.frame(out_list1[[25]]),as.data.frame(out_list1[[29]]),
as.data.frame(out_list1[[34]]),as.data.frame(out_list1[[35]]),
as.data.frame(out_list1[[64]]),as.data.frame(out_list1[[66]])

# B, Kontinuierliche Kovariablen

# Position der relevanten Kovariablen im Datensatz (Spaltennummer)
dk <- c(9, 12, 18, 24, 26, 27, 28, 51, 65)
out_list2 <- vector(mode = "list", length =9)

```

```

for (i in dk){
kont1 <- aggregate(table1df[,i]~age_center_q*PWCquantsw, table1df, mean)
kont1$age_dist[kont1$age_center_q== 1] <- age1
kont1$age_dist[kont1$age_center_q== 2] <- age2
kont1$age_dist[kont1$age_center_q== 3] <- age3
kont1$mean_stand <- kont1$age_dist * kont1[, 3]
kont2 <- aggregate(kont1[, "mean_stand"], by=list(PWCquantsw=kont1$PWCquantsw),
FUN=sum)
kont2$x <- round(kont2$x, 2)

tmydf = setNames(data.frame(t(kont2[,-1])), kont2[,1])
noNA <- sum(is.na(table1df[,i]))
row1 <- data.frame(Variable=as.factor(colnames(table1df)[i]), ergoQ1=NA,
ergoQ2=NA, ergoQ3=NA, ergoQ4=NA, numberNA=noNA)
colnames(tmydf)[1] <- "ergoQ1"
colnames(tmydf)[2] <- "ergoQ2"
colnames(tmydf)[3] <- "ergoQ3"
colnames(tmydf)[4] <- "ergoQ4"
tmydf$numberNA <- NA
tmydf$Variable <- as.factor(NA)

t4skont <- rbind(row1, tmydf)
out_list2[[i]] <- t4skont
names(out_list2)[i] <- paste("df", i, sep = "_")
}
testabk <- rbind(as.data.frame(out_list2[[9]]), as.data.frame(out_list2[[12]]),
as.data.frame(out_list2[[65]]), as.data.frame(out_list2[[18]]),
as.data.frame(out_list2[[24]]), as.data.frame(out_list2[[51]]),
as.data.frame(out_list2[[26]]), as.data.frame(out_list2[[27]]),
as.data.frame(out_list2[[28]]))

#####

###
# Cox proportional Hazards Regression mit multipel imputiertem Datensatz - Auszüge
###

# Beispiel: Gesamte Kohorte, Imputierter Datensatz, Endpunkt Kolorektales
Karzinom, Voll adjustiertes Modell:

ddd <- datadist(alllong[einzelN0$CRFmis==1, ])
options(datadist='ddd')

S1 <- Surv(einzelN0[einzelN0$CRFmis==1, "f.21003.0.0"],
einzelN0[einzelN0$CRFmis==1, "Ageoutnum"],
einzelN0[einzelN0$CRFmis==1, "colorect"])

CRF1 <- fit.mult.impute(S1 ~ PWC75Psw + f.31.0.0 + f.50.0.0+ alc3+ redprocmeat+
educat4 + smoke3+income3 + screen_bowel + f.54.0.0 + aspiuse + familybowel+
IPAQtv, fitter=cph, xtrans=imp_merged, data = alllong,subset=(einzelN0$CRFmis==1),
y=T, x=T)

```

```

# Schoenfeld Residuen
cox.zph(CRF1)

# Hazard Ratio und 95 % Konfidenzintervall
print(paste0(round(summary(test)[2, 4],2), " (", round(summary(test)[2, 6],2),"-",
round(summary(test)[2, 7],2), ")"))

# Gesamt-p-Wert
round(anova(test)[1:2, 1:3],3)

# Effektmodifikation (am Beispiel Geschlecht)
CRF2 <- fit.mult.impute(S1 ~ PWC75Psw * f.31.0.0 + f.50.0.0+ alc3+ redprocmeat+
educat4 +
smoke3+income3 + screen_bowel + f.54.0.0 + aspibuse + familybowel+ IPAQtv,
fitter=cph, xtrans=imp_merged, data = alllong, subset=(einzelN0$CRFmis==1), y=T,
x=T)
anova(CRF2)
lrtest(CRF1, CRF2)

#####
#####

```

7.4.5 Auszüge R Code – Körperliche Aktivität, Greifkraft, sedentäres Verhalten und Risiko von malignem Melanom in der UK Biobank

```

#####

# Laden von R Paketen

library(reshape)
library(dplyr)
library(data.table)
library(tidyr)
library(ggplot2)
library(rms)
library(mice)
library(gridExtra)
library(tsModel)
library(parallel)

#####

###
# Aufbereitung der körperlichen Aktivität - Auszüge
###

# A, Ausschluss von Ausreißern (Summe von Gehen, moderater und intensiver
Körperlicher Aktivität > 960 Minuten pro Tag)
dsIPAQ$outliercheck <- rowSums(dsIPAQ[, c("f.874.0.0", "f.894.0.0",
"f.914.0.0")], na.rm=F)
dsIPAQ$f.874.0.0[dsIPAQ$outliercheck > 960 & !is.na(dsIPAQ$outliercheck)] <- NA

```

```
dsIPAQ$f.894.0.0[dsIPAQ$outliercheck > 960 & !is.na(dsIPAQ$outliercheck)] <- NA
dsIPAQ$f.914.0.0[dsIPAQ$outliercheck > 960 & !is.na(dsIPAQ$outliercheck)] <- NA
```

```
# B, Umcodierung von Angaben mit < 10 Minutes auf 0
```

```
# Gehen
```

```
dsIPAQ$f.864.0.0[dsIPAQ$f.874.0.0 < 10] <- 0
```

```
dsIPAQ$f.874.0.0[dsIPAQ$f.874.0.0 < 10] <- 0
```

```
# moderat
```

```
dsIPAQ$f.884.0.0[dsIPAQ$f.894.0.0 < 10] <- 0
```

```
dsIPAQ$f.894.0.0[dsIPAQ$f.894.0.0 < 10] <- 0
```

```
# intensiv
```

```
dsIPAQ$f.904.0.0[dsIPAQ$f.914.0.0 < 10] <- 0
```

```
dsIPAQ$f.914.0.0[dsIPAQ$f.914.0.0 < 10] <- 0
```

```
# C, Winsorisierung hoher Werte: Umcodierung auf 180 Minuten pro Tag
```

```
dsIPAQ$f.874.0.0[dsIPAQ$f.874.0.0 > 180] <- 180
```

```
dsIPAQ$f.894.0.0[dsIPAQ$f.894.0.0 > 180] <- 180
```

```
dsIPAQ$f.914.0.0[dsIPAQ$f.914.0.0 > 180] <- 180
```

```
# D, Bildung des kontinuierlichen Scores (Gewichtung der Dauer und Häufigkeit mit Intensitätsfaktoren)
```

```
dsIPAQ$METminweekWalk <- 3.3 * dsIPAQ$f.874.0.0 * dsIPAQ$f.864.0.0
```

```
dsIPAQ$METminweekMod <- 4 * dsIPAQ$f.894.0.0 * dsIPAQ$f.884.0.0
```

```
dsIPAQ$METminweekVig <- 8 * dsIPAQ$f.914.0.0 * dsIPAQ$f.904.0.0
```

```
dsIPAQ$METminweekTotal <- rowSums(dsIPAQ[, c("METminweekWalk", "METminweekMod", "METminweekVig")], na.rm=F)
```

```
dsIPAQ$METhrweekTotal <- dsIPAQ$METminweekTotal/60
```

```
###
```

```
# Aufbereitung der Greifkraft
```

```
###
```

```
dsIPAQ$gripmw <- (dsIPAQ$f.46.0.0+dsIPAQ$f.47.0.0)/2
```

```
###
```

```
# Aufbereitung des sedentären Verhaltens
```

```
###
```

```
# "Weiss nicht" und "keine Angabe" werden zu NA umcodiert, "weniger al seine Stunde am Tag" wird zu 0.5 umcodiert (Beispielhaft für Fernsehschauen)
```

```
dsIPAQ$IPAQtv <- dsIPAQ$f.1070.0.0
```

```
dsIPAQ$IPAQtv[dsIPAQ$f.1070.0.0 %in% c(-1, -3)] <- NA
```

```
dsIPAQ$IPAQtv[dsIPAQ$f.1070.0.0 == -10] <- 0.5
```

```
# Summe aus TV, PC und Fortbewegung bilden und auf 24 Stunden winsorisieren
```

```
dsIPAQ$SBTotal <- rowSums(dsIPAQ[, c("IPAQtv", "IPAQpc", "IPAQdrive")], na.rm=F)
```

```
dsIPAQ$SBTotal[dsIPAQ$SBTotal > 24] <- 24
```

```
###
```

```
# Aufbereitung der malignes Melanom Diagnosen - Auszüge
```

```
###
```

```
# Analog zu 7.4.4 (R-Code Auszüge - Kardiorespiratorische Fitness und
```

```
#Kolorektalkarzinom in der UK Biobank) müssen zunächst alle Variablen in das lange
```

```
#Format transponiert werden. Ebenso werden alle Teilnehmende mit mindestens einer
```

```

# malignen Krebsdiagnose(Ausnahme nichtmelanozytärer Hautkrebs (C44))
ausgeschlossen.

# Identifizierung malignes Melanom (primär malignant oder mikroinvasiv)
merge6$malig[merge6$behaviour == "Malignant, microinvasive" | merge6$behaviour ==
"Malignant, primary site"] <- 1
# Malignes Melanom an Lokalisation mit chronischer UV Expo
merge6$mmelanoma[merge6$ICD10 %in% c("C430", "C431", "C432", "C433", "C434",
"C436") & merge6$histology %in% c(8720:8790)] <- 1
# Malignes Melanom an Lokalisation mit sporadischer UV Expo
merge6$mmelanoma[merge6$ICD10 %in% c("C435","C437") & merge6$histology %in%
c(8720:8790)] <- 2
merge6$mm_prim <- 0
merge6$mm_prim[merge6$malig == 1 & merge6$mmelanoma %in% c(1,2) &
merge6$date_diagnose > merge6$f.53.0.0 & !is.na(merge6$date_diagnose)] <- 1

# Überprüfung, ob es am Tag der malignen Melanom-Diagnose mehrere Diagnosen an
chronisch und sporadisch UV-exponierten Hautstellen gibt
MMDS <- subset(merge6, mm_prim == 1)
MMDS1 <- MMDS[order(MMDS$f.eid, MMDS$date_diagnose),]
MMDS2 <- MMDS1[!duplicated(MMDS1$f.eid), ]
MMDS3 <- left_join(MMDS2[ , c("f.eid", "date_diagnose")], MMDS1, by=c("f.eid",
"date_diagnose" )
MMDS3UV <- MMDS3[MMDS3$mmelanoma == 1, c("f.eid", "mmelanoma", "date_diagnose")]
MMDS3nUV <- MMDS3[MMDS3$mmelanoma == 2, c("f.eid", "mmelanoma", "date_diagnose")]
MMDS3UVMERGE <- inner_join(MMDS3UV, MMDS3nUV, by=c("f.eid", "date_diagnose"))

# Zensierung: Je nachdem was zuerst eintritt, Komplette Nachbeobachtung, Malignes
Melanom oder Tod (Separat für gesamtes Malignes Melanom und für malignes Melanom
an chronisch und sporadisch UV-exponierten Lokalisationen

# A, Tod (ebenfalls ins Lange Format transponiert)
outcomeflags3 <- left_join(outcomeflags1, censordeath, by="f.eid")
outcomeflags3$death <- 0
outcomeflags3$death[!is.na(outcomeflags3$DateCensorDeath)] <- 1
#table(outcomeflags3$death)

# B, Komplette Nachbeobachtung (Update 2019)
# Schottland: 31 Oktober 2015
# England/ Wales: 31. März 2016
outcomeflags3$DateCompleteFU[outcomeflags3$f.54.0.0 %in% c("11004", "11005")] <-
"2015-10-31"
EngWal <- c("10003", "11001", "11002", "11003", "11006", "11007","11008","11009",
"11010","11011","11012","11013","11014","11016","11017","11018", "11020","11021",
"11022","11023","11024","11025")
outcomeflags3$DateCompleteFU[outcomeflags3$f.54.0.0 %in% EngWal] <- "2016-03-31"

# C, Inzidenze Diagnose malignes Melanom

# D, Identifizierung des Zensierungs-Events (beispielhaft für gesamtes malignes
Melanom)

```

```

whicheverfirst <- gather(outcomeflags3, "date_diagnose", "DateCensorDeath",
"DateCompleteFU", key="whicheverfirst", value="Datum")
whicheverfirst1 <- whicheverfirst[!is.na(whicheverfirst$Datum), ]
whicheverfirst1$Datumnum <- as.numeric(whicheverfirst1$Datum)
whicheverfirst2 <- whicheverfirst1[order(whicheverfirst1$f.eid,
whicheverfirst1$Datumnum), ]
whicheverfirst3 <- whicheverfirst2[!duplicated(whicheverfirst2$f.eid),]
summary(whicheverfirst3$Other_before_MM)
# Falls Diagnose erst nach komplettem Nachbeobachtungszeitraum: auf NA codieren
whicheverfirst3$ICD10[whicheverfirst3$whicheverfirst == "DateCompleteFU" &
!is.na(whicheverfirst3$ICD10)] <- NA
whicheverfirst3$behaviour[whicheverfirst3$whicheverfirst == "DateCompleteFU" &
!is.na(whicheverfirst3$behaviour)] <- NA
whicheverfirst3$histology[whicheverfirst3$whicheverfirst == "DateCompleteFU" &
!is.na(whicheverfirst3$histology)] <- NA
whicheverfirst3$source[whicheverfirst3$whicheverfirst == "DateCompleteFU" &
!is.na(whicheverfirst3$source)] <- NA
whicheverfirst3$mm_prim[whicheverfirst3$whicheverfirst == "DateCompleteFU" &
!is.na(whicheverfirst3$mm_prim)] <- NA
whicheverfirst3$flag_multMMday[whicheverfirst3$whicheverfirst == "DateCompleteFU"
& !is.na(whicheverfirst3$flag_multMMday)] <- NA
whicheverfirst3$Other_before_MM[whicheverfirst3$whicheverfirst == "DateCompleteFU"
& !is.na(whicheverfirst3$Other_before_MM)] <- NA

# E, Alter zur Zensierung, Nachbeobachtungszeit (beispielhaft für gesamtes
malignes Melanom)
whicheverfirst5$FUt看ime <- whicheverfirst5$Datum - whicheverfirst5$f.53.0.0
whicheverfirst5$Ageout <- round(whicheverfirst5$f.21003.0.0 +
whicheverfirst5$FUt看ime/365.25, 2)

###
# Multiple Imputation mit Chained Equations (MICE) mittels Parallelisierung -
Auszüge
###

# Spezifizierung der Parameter
dry <- mice(imputDS, maxit=0)
meth1 <- dry$method

# Imputationsmodelle basierend auf Variablenklasse
# numeric: pmm (default)
# binary: Logreg
# factor: polyreg
# ordinal: polr
meth1["educat4"] <- "polr"
meth1["healthrate"] <- "polr"
meth1["alc3"] <- "polr"
meth1["smoke3"] <- "polr"
meth1["facialage"] <- "polr"

# Prädiktormatrix: welche Variablen sollen nicht als Prädiktor verwendet werden
pred[ , "f.eid"] <- 0

```

```

pred[ , c("PAquant", "SBquant", "Gripquant", "PADummy")] <- 0
pred[ , c("mmelanoma_UVExpo", "mmelanoma_noUVExpo")] <- 0
pred[ , c("Ageoutnum_UVExpo", "Ageoutnum_noUVExpo", "Ageoutnum")] <- 0
pred[ , c("A1", "A2", "A3", "A4")] <- 0
pred[ , "UV_job"] <- 0
pred[ , "smoke3"] <- 0
pred[ , "alc3"] <- 0
# Prädiktormatrix: welche Variablen sollen nicht imputiert werden (z.B. wenn keine
NA vorliegt):
pred[c("f.eid", "METminweekTotal", "SBTotal", "gripmw", "PAquant", "SBquant",
"Gripquant", "PADummy", "mm_prim", "mmelanoma_UVExpo", "mmelanoma_noUVExpo", "Ageoutnum
_UVExpo", "Ageoutnum_noUVExpo", "Ageoutnum", "A1", "A2", "A3", "A4", "f.21003.0.0",
"sex", "region") , ] <- 0

# Parallelisierung
cores_2_use <- 2

cl <- makeCluster(cores_2_use)
clusterSetRNGStream(cl, 9956)
clusterExport(cl, "imputDS")
clusterExport(cl, "meth1")
clusterExport(cl, "pred")
clusterEvalQ(cl, library(mice))

imp_pars <-
  parLapply(cl = cl, X = 1:cores_2_use, fun = function(no){
    imput_neu <- mice(imputDS, meth=meth1, maxit=5, m=5, pred=pred, print=F)
  })

stopCluster(cl)

imp_merged <- imp_pars[[1]]
for (n in 2:length(imp_pars)){
  imp_merged <-
    ibind(imp_merged,
          imp_pars[[n]])
}

alllong <- complete(imp_merged, "long", include=T)
einzel_n0 <- mice::complete(imp_merged, 0)

###
# Alters-Adjustierte Charakteristika der Studienteilnehmenden zur
Basisuntersuchung (Automatisiert) - Analog zu Abschnitt 7.4.4
###

###
# Cox proportional Hazards Regression mit multipel imputiertem Datensatz - Auszüge
###

# Beispiel: Gesamte Kohorte, imputierter Datensatz, Endpunkt malignes Melanom,
voll adjustiertes Modell:

```



```
ddd <- datadist(alllong)
options(datadist='ddd')

CRFimp <- fit.mult.impute(Surv(f.21003.0.0, Ageoutnum, mm_prim) ~ gripmw +
METminweekTotal + SBTtotal + sex + strat(region) + educat4 + packyears + alk_gd +
coffeday + f.50.0.0 + outdoorsummer + sunburn_child + solarium + Score_Sens + A1
+ A2 + A3 + A4, fitter=cph, xtrans=imp_merged, data = alllong, y=T, x=T)

# Schoenfeld Residuen
PH_total <- cox.zph(CRFimp )

# Hazard Ratio und 95 % Konfidenzintervall
summary_total <- summary(CRFimp )
print(paste0(round(summary(CRFimp )[2, 4],2), " (", round(summary(CRFimp )[2,
6],2), "-" , round(summary(CRFimp )[2, 7],2), ")"))

# Gesamt-p-Wert
round(anova(CRFimp )[1, 3],3)

#####
#####
```

8 Literaturverzeichnis

1. Morris JN, Heady JA, Raffle PA, Roberts CG, Parks JW. Coronary heart-disease and physical activity of work. *Lancet* (London, England). 1953;262(6795):1053-7.
2. McKinney J, Lithwick DJ, Morrison BN, Nazzari H, Isserow S, Heilbron B, et al. The health benefits of physical activity and cardiorespiratory fitness. *BCMJ*. 2016;58(3):131-7.
3. WHO guidelines on physical activity and sedentary behaviour. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO.
4. 2018 Physical Activity Guidelines Advisory Committee. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. Washington, DC: U.S. Department of Health and Human Services, 2018.
5. Dempsey PC, Biddle SJH, Buman MP, Chastin S, Ekelund U, Friedenreich CM, et al. New global guidelines on sedentary behaviour and health for adults: broadening the behavioural targets. *International Journal of Behavioral Nutrition and Physical Activity*. 2020;17(1):151.
6. Segar ML, Marques MM, Palmeira AL, Okely AD. Everything counts in sending the right message: science-based messaging implications from the 2020 WHO guidelines on physical activity and sedentary behaviour. *The international journal of behavioral nutrition and physical activity*. 2020;17(1):135.
7. DiPietro L, Al-Ansari SS, Biddle SJH, Borodulin K, Bull FC, Buman MP, et al. Advancing the global physical activity agenda: recommendations for future research by the 2020 WHO physical activity and sedentary behavior guidelines development group. *The international journal of behavioral nutrition and physical activity*. 2020;17(1):143.
8. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Reports*. 1985;100(2):126-31.
9. American College of Sports Medicine. *ACSM's Guidelines for Exercise Testing and Prescription*. 8th ed. Baltimore, MD: Lippincott Williams & Wilkins; 2009.
10. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, et al. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. *The international journal of behavioral nutrition and physical activity*. 2017;14(1):75.
11. Leitzmann M, Powers H, Anderson AS, Scoccianti C, Berrino F, Boutron-Ruault MC, et al. European Code against Cancer 4th Edition: Physical activity and cancer. *Cancer epidemiology*. 2015;39 Suppl 1:S46-55.
12. Haskell WL. Physical activity by self-report: a brief history and future issues. *Journal of physical activity & health*. 2012;9 Suppl 1:S5-10.
13. Lagerros YT, Lagiou P. Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases. *Eur J Epidemiol*. 2007;22(6):353-62.
14. Keating XD, Zhou K, Liu X, Hodges M, Liu J, Guan J, et al. Reliability and Concurrent Validity of Global Physical Activity Questionnaire (GPAQ): A Systematic Review. *Int J Environ Res Public Health*. 2019;16(21).
15. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International Physical Activity Questionnaire: 12-Country Reliability and Validity. *Medicine & Science in Sports & Exercise*. 2003;35(8).
16. 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. *The international journal of behavioral nutrition and physical activity*. 2008;5:56.
17. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *British journal of sports medicine*. 2014;48(13):1019-23.
18. DeFina LF, Haskell WL, Willis BL, Barlow CE, Finley CE, Levine BD, et al. Physical activity versus cardiorespiratory fitness: two (partly) distinct components of cardiovascular health? *Progress in cardiovascular diseases*. 2015;57(4):324-9.
19. Bouchard C, Rankinen T, Timmons JA. Genomics and genetics in the biology of adaptation to exercise. *Comprehensive Physiology*. 2011;1(3):1603-48.

20. Zempo H, Miyamoto-Mikami E, Kikuchi N, Fuku N, Miyachi M, Murakami H. Heritability estimates of muscle strength-related phenotypes: A systematic review and meta-analysis. *Scandinavian journal of medicine & science in sports*. 2017;27(12):1537-46.
21. Blair SN, Cheng Y, Holder JS. Is physical activity or physical fitness more important in defining health benefits? *Medicine and science in sports and exercise*. 2001;33(6 Suppl):S379-99; discussion S419-20.
22. Pollan M, Casla-Barrio S, Alfaro J, Esteban C, Segui-Palmer MA, Lucia A, et al. Exercise and cancer: a position statement from the Spanish Society of Medical Oncology. *Clin Transl Oncol*. 2020.
23. Hojman P, Gehl J, Christensen JF, Pedersen BK. Molecular Mechanisms Linking Exercise to Cancer Prevention and Treatment. *Cell Metab*. 2018;27(1):10-21.
24. Lee S, Bacha F, Gungor N, Arslanian SA. Cardiorespiratory fitness in youth: relationship to insulin sensitivity and beta-cell function. *Obesity (Silver Spring, Md)*. 2006;14(9):1579-85.
25. Hojman P. Exercise protects from cancer through regulation of immune function and inflammation. *Biochemical Society transactions*. 2017;45(4):905-11.
26. Traustadottir T, Davies SS, Su Y, Choi L, Brown-Borg HM, Roberts LJ, 2nd, et al. Oxidative stress in older adults: effects of physical fitness. *Age (Dordrecht, Netherlands)*. 2012;34(4):969-82.
27. de Sousa CV, Sales MM, Rosa TS, Lewis JE, de Andrade RV, Simões HG. The Antioxidant Effect of Exercise: A Systematic Review and Meta-Analysis. *Sports medicine (Auckland, NZ)*. 2017;47(2):277-93.
28. Santos RV, Viana VA, Boscolo RA, Marques VG, Santana MG, Lira FS, et al. Moderate exercise training modulates cytokine profile and sleep in elderly people. *Cytokine*. 2012;60(3):731-5.
29. Lavie CJ, Church TS, Milani RV, Earnest CP. Impact of physical activity, cardiorespiratory fitness, and exercise training on markers of inflammation. *J Cardiopulm Rehabil Prev*. 2011;31(3):137-45.
30. Cash SW, Beresford SA, Vaughan TL, Heagerty PJ, Bernstein L, White E, et al. Recent physical activity in relation to DNA damage and repair using the comet assay. *Journal of physical activity & health*. 2014;11(4):770-6.
31. Schmid D, Leitzmann MF. Cardiorespiratory fitness as predictor of cancer mortality: a systematic review and meta-analysis. *Annals of oncology : official journal of the European Society for Medical Oncology*. 2015;26(2):272-8.
32. Robsahm TE, Falk RS, Heir T, Sandvik L, Vos L, Erikssen JE, et al. Measured cardiorespiratory fitness and self-reported physical activity: associations with cancer risk and death in a long-term prospective cohort study. *Cancer medicine*. 2016;5(8):2136-44.
33. Vainshelboim B, Muller J, Lima RM, Nead KT, Chester C, Chan K, et al. Cardiorespiratory fitness and cancer incidence in men. *Annals of epidemiology*. 2017;27(7):442-7.
34. Robsahm TE, Falk RS, Heir T, Sandvik L, Vos L, Erikssen J, et al. Cardiorespiratory fitness and risk of site-specific cancers: a long-term prospective cohort study. *Cancer medicine*. 2017;6(4):865-73.
35. Lakoski SG, Willis BL, Barlow CE, Leonard D, Gao A, Radford NB, et al. Midlife Cardiorespiratory Fitness, Incident Cancer, and Survival After Cancer in Men: The Cooper Center Longitudinal Study. *JAMA oncology*. 2015;1(2):231-7.
36. Marshall CH, Al-Mallah MH, Dardari Z, Brawner CA, Lamerato LE, Keteyian SJ, et al. Cardiorespiratory fitness and incident lung and colorectal cancer in men and women: Results from the Henry Ford Exercise Testing (FIT) cohort. *Cancer*. 2019;125(15):2594-601.
37. Pletnikoff PP, Tuomainen TP, Laukkanen JA, Kauhanen J, Rauramaa R, Ronkainen K, et al. Cardiorespiratory fitness and lung cancer risk: A prospective population-based cohort study. *Journal of science and medicine in sport*. 2016;19(2):98-102.
38. Pozuelo-Carrascosa DP, Alvarez-Bueno C, Cervero-Redondo I, Morais S, Lee IM, Martínez-Vizcaíno V. Cardiorespiratory fitness and site-specific risk of cancer in men: A systematic review and meta-analysis. *European Journal of Cancer*. 2019;113:58-68.
39. Celis-Morales CA, Welsh P, Lyall DM, Steell L, Petermann F, Anderson J, et al. Associations of grip strength with cardiovascular, respiratory, and cancer outcomes and all cause mortality: prospective cohort study of half a million UK Biobank participants. *BMJ*. 2018;361:k1651.

40. Jochem C, Wallmann-Sperlich B, Leitzmann MF. The Influence of Sedentary Behavior on Cancer Risk: Epidemiologic Evidence and Potential Molecular Mechanisms. *Current Nutrition Reports*. 2019;8(3):167-74.
41. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. *The Lancet Global health*. 2018.
42. Finger JD, Mensink G, Lange C, Manz K. Gesundheitsfördernde körperliche Aktivität in der Freizeit bei Erwachsenen in Deutschland. Robert Koch-Institut, Epidemiologie und Gesundheitsberichterstattung; 2017.
43. Strain T, Brage S, Sharp SJ, Richards J, Tainio M, Ding D, et al. Use of the prevented fraction for the population to determine deaths averted by existing prevalence of physical activity: a descriptive study. *The Lancet Global health*. 2020;8(7):e920-e30.
44. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018;68(6):394-424.
45. Vineis P, Wild CP. Global cancer patterns: causes and prevention. *Lancet (London, England)*. 2014;383(9916):549-57.
46. Global action plan on physical activity 2018–2030: more active people for a healthier world. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO.
47. Weiß C. Basiswissen Medizinische Statistik. 6. Auflage. 2013.: Springer-Verlag Berlin Heidelberg.
48. Skender S, Ose J, Chang-Claude J, Paskow M, Bruhmann B, Siegel EM, et al. Accelerometry and physical activity questionnaires - a systematic review. *BMC public health*. 2016;16:515.
49. Pettee Gabriel KK, Morrow JR, Jr., Woolsey AL. Framework for physical activity as a complex and multidimensional behavior. *Journal of physical activity & health*. 2012;9 Suppl 1:S11-8.
50. Broderick JM, Ryan J, O'Donnell DM, Hussey J. A guide to assessing physical activity using accelerometry in cancer patients. *Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer*. 2014;22(4):1121-30.
51. Lee IM, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. *British journal of sports medicine*. 2014;48(3):197-201.
52. Leitzmann M, Gastell S, Hillreiner A, Herbolsheimer F, Baumeister SE, Bohn B, et al. Körperliche Aktivität in der NAKO Gesundheitsstudie: erste Ergebnisse des multimodalen Erhebungskonzepts. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*. 2020.
53. Schipf S, Schöne G, Schmidt B, Günther K, Stübs G, Greiser KH, et al. [The baseline assessment of the German National Cohort (NAKO Gesundheitsstudie): participation in the examination modules, quality assurance, and the use of secondary data]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*. 2020;63(3):254-66.
54. Ahrens W, Greiser H, Linseisen J, Kluttig A, Schipf S, Schmidt B, et al. [The design of a nationwide cohort study in Germany : the pretest studies of the German National Cohort (GNC)]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*. 2014;57(11):1246-54.
55. NAKO Gesundheitsstudie (2015). Ethik-Kodex der NAKO Gesundheitsstudie. https://nako.de/wp-content/uploads/2015/10/NAKO_Ethik-Kodex_v2.0_2015-10-20.pdf. Zitiert am 05.01.2021.
56. Deutsche Gesellschaft für Epidemiologie, Deutsche Gesellschaft für Medizinische Informatik Biometrie und Epidemiologie, Deutsche Gesellschaft für Sozialmedizin und Prävention, Deutsche Region der Internationalen Biometrischen Gesellschaft, Technologie- und Methodenplattform für die Vernetzte Medizinische Forschung E. V., Deutsches Netzwerk Versorgungsforschung E. V. (2018) Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP). https://www.dgepi.de/assets/Leitlinien-und-Empfehlungen/Leitlinien_fuer_Gute_Epidemiologische_Praxis_GEP_vom_September_2018.pdf. Zitiert am 05.01.2021.
57. Schipf S, Ostrzinski S, Melzer J. Vorgaben für die Datenbereinigung im Rahmen des Data Freeze 100.000. Version 01 (NAKO SOP QM03-GE). 05.10.2017.

58. World Health Organization (WHO). Noncommunicable diseases and their risk factors. Global Physical Activity Surveillance. <https://www.who.int/ncds/surveillance/steps/GPAQ/en/>. Zitiert am 05.01.2021.
59. The German National Cohort: aims, study design and organization. *Eur J Epidemiol*. 2014;29(5):371-82.
60. Rudolf K, Lammer F, Stassen G, Froböse I, Schaller A. Show cards of the Global Physical Activity Questionnaire (GPAQ) – do they impact validity? A crossover study. *BMC public health*. 2020;20(1):223.
61. Herrmann SD, Heumann KJ, Der Ananian CA, Ainsworth BE. Validity and Reliability of the Global Physical Activity Questionnaire (GPAQ) Measurement in Physical Education and Exercise Science. 2013;17(3):221-35.
62. Bull FC, Maslin TS, Armstrong T. Global physical activity questionnaire (GPAQ): nine country reliability and validity study. *Journal of physical activity & health*. 2009;6(6):790-804.
63. Cleland CL, Hunter RF, Kee F, Cupples ME, Sallis JF, Tully MA. Validity of the global physical activity questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour. *BMC public health*. 2014;14:1255.
64. Wanner M, Hartmann C, Pestoni G, Martin BW, Siegrist M, Martin-Diener E. Validation of the Global Physical Activity Questionnaire for self-administration in a European context. *BMJ open sport & exercise medicine*. 2017;3(1):e000206-e.
65. Chu AH, Ng SH, Koh D, Muller-Riemenschneider F. Reliability and Validity of the Self- and Interviewer-Administered Versions of the Global Physical Activity Questionnaire (GPAQ). *PLoS one*. 2015;10(9):e0136944.
66. World Health Organization (WHO). Global Physical Activity Questionnaire (GPAQ) Analysis Guide. Surveillance and Population-Based Prevention. Prevention of Noncommunicable Diseases Department. 20 Avenue Appia. 1211 Geneva 27. Switzerland. (www.who.int/chp/steps).
67. World Health Organization (WHO). Global Recommendations on Physical Activity for Health. 2010.
68. Kohler S, Behrens G, Olden M, Baumeister SE, Horsch A, Fischer B, et al. Design and Evaluation of a Computer-Based 24-Hour Physical Activity Recall (cpar24) Instrument. *Journal of medical Internet research*. 2017;19(5):e186.
69. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Jr., Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Medicine and science in sports and exercise*. 2011;43(8):1575-81.
70. Kitterød RH. Does the recording of parallel activities in Time Use Diaries affect the way people report their main activities? *Social Indicators Research*. 2001;56(2):145-78.
71. van der Loo MPJ. The stringdist Package for Approximate String Matching. *The R Journal* Vol 6/1. 2014. (<https://journal.r-project.org/archive/2014-1/loo.pdf>).
72. Van der Loo MPJ, van der Laan J, R Core Team, Logan N, Muir C. stringdist: Approximate String Matching and String Distance Functions. R package version 0.9.5.1. (<https://github.com/markvanderloo/stringdist>). 2018.
73. Matthews CE, Berrigan D, Fischer B, Gomersall SR, Hillreiner A, Kim Y, et al. Use of previous-day recalls of physical activity and sedentary behavior in epidemiologic studies: results from four instruments. *BMC public health*. 2019;19(2):478.
74. Vigl M. Untersuchungsdurchführung 7-Tage-Akzelerometrie. Version V01 (NAKO SOP UN02_SD_V01). 13.02.2015.
75. Migueles JH, Rowlands AV, Huber F, Sabia S, van Hees VT. GGIR: A Research Community-Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data. *Journal for the Measurement of Physical Behaviour*. 2019;2(3):188-96.
76. van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva IC, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *Journal of applied physiology (Bethesda, Md : 1985)*. 2014;117(7):738-44.

77. van Hees VT, Gorzelniak L, Dean Leon EC, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PloS one*. 2013;8(4):e61691.
78. Aittasalo M, Vähä-Ypyä H, Vasankari T, Husu P, Jussila A-M, Sievänen H. Mean amplitude deviation calculated from raw acceleration data: a novel method for classifying the intensity of adolescents' physical activity irrespective of accelerometer brand. *BMC sports science, medicine and rehabilitation*. 2015;7:18.
79. Hildebrand M, VT VANH, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Medicine and science in sports and exercise*. 2014;46(9):1816-24.
80. Doherty A, Jackson D, Hammerla N, Plotz T, Olivier P, Granat MH, et al. Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. *PloS one*. 2017;12(2):e0169649.
81. Trost SG, Pate RR, Freedson PS, Sallis JF, Taylor WC. Using objective physical activity measures with youth: how many days of monitoring are needed? *Medicine and science in sports and exercise*. 2000;32(2):426-31.
82. Luzak A, Heier M, Thorand B, Laxy M, Nowak D, Peters A, et al. Physical activity levels, duration pattern and adherence to WHO recommendations in German adults. *PloS one*. 2017;12(2):e0172503.
83. Hamrik Z, Sigmundova D, Kalman M, Pavelka J, Sigmund E. Physical activity and sedentary behaviour in Czech adults: results from the GPAQ study. *European journal of sport science*. 2014;14(2):193-8.
84. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: an overview of objective measures. *Front Nutr*. 2014;1:5-.
85. Dunton GF, Berrigan D, Ballard-Barbash R, Graubard B, Atienza AA. Joint associations of physical activity and sedentary behaviors with body mass index: results from a time use survey of US adults. *International journal of obesity (2005)*. 2009;33(12):1427-36.
86. Dunton GF, Berrigan D, Ballard-Barbash R, Graubard BI, Atienza AA. Social and physical environments of sports and exercise reported among adults in the American Time Use Survey. *Preventive medicine*. 2008;47(5):519-24.
87. Tudor-Locke C, Bittman M, Merom D, Bauman A. Patterns of walking for transport and exercise: a novel application of time use data. *The international journal of behavioral nutrition and physical activity*. 2005;2:5.
88. Brenner H, Kloor M, Pox CP. Colorectal cancer. *The Lancet*. 2014;383(9927):1490-502.
89. Luengo-Fernandez R, Leal J, Gray A, Sullivan R. Economic burden of cancer across the European Union: a population-based cost analysis. *The Lancet Oncology*. 2013;14(12):1165-74.
90. Gravitz L. Prevention: Tending the gut. *Nature*. 2015;521:S6.
91. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG, et al. Colorectal cancer. *Nature reviews Disease primers*. 2015;1:15065.
92. Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M. Protection from colorectal cancer after colonoscopy: a population-based, case-control study. *Annals of internal medicine*. 2011;154(1):22-30.
93. Mahmood S, MacInnis RJ, English DR, Karahalios A, Lynch BM. Domain-specific physical activity and sedentary behaviour in relation to colon and rectal cancer risk: a systematic review and meta-analysis. *Int J Epidemiol*. 2017;46(6):1797-813.
94. Moore SC, Lee IM, Weiderpass E, Campbell PT, Sampson JN, Kitahara CM, et al. Association of Leisure-Time Physical Activity With Risk of 26 Types of Cancer in 1.44 Million Adults. *JAMA internal medicine*. 2016;176(6):816-25.
95. Rezende LFM, Sa TH, Markozannes G, Rey-Lopez JP, Lee IM, Tsilidis KK, et al. Physical activity and cancer: an umbrella review of the literature including 22 major anatomical sites and 770 000 cancer cases. *British journal of sports medicine*. 2018;52(13):826-33.
96. Laukkanen JA, Pukkala E, Rauramaa R, Makikallio TH, Toriola AT, Kurl S. Cardiorespiratory fitness, lifestyle factors and cancer risk and mortality in Finnish men. *European journal of cancer (Oxford, England : 1990)*. 2010;46(2):355-63.

97. Steell L, Ho FK, Sillars A, Petermann-Rocha F, Li H, Lyall DM, et al. Dose-response associations of cardiorespiratory fitness with all-cause mortality and incidence and mortality of cancer and cardiovascular and respiratory diseases: the UK Biobank cohort study. *British journal of sports medicine*. 2019.
98. Nunez C, Clausen J, Jensen MT, Holtermann A, Gyntelberg F, Bauman A. Main and interactive effects of physical activity, fitness and body mass in the prevention of cancer from the Copenhagen Male Study. *Scientific reports*. 2018;8(1):11780.
99. McGuire KA, Ross R. Incidental physical activity is positively associated with cardiorespiratory fitness. *Medicine and science in sports and exercise*. 2011;43(11):2189-94.
100. Ross R, Blair SN, Arena R, Church TS, Despres JP, Franklin BA, et al. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. *Circulation*. 2016;134(24):e653-e99.
101. Trehearne A. Genetics, lifestyle and environment. UK Biobank is an open access resource following the lives of 500,000 participants to improve the health of future generations. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*. 2016;59(3):361-7.
102. UK Biobank. Cardio Assessment. Version 1.0. 2011. <https://biobank.ctsu.ox.ac.uk/crystal/docs/Cardio.pdf>.
103. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. 9th Edition. edn. Baltimore: Wolters Kluwer Health/Lippincott, Williams & Wilkins; 2014.
104. Noonan V, Dean E. Submaximal exercise testing: clinical application and interpretation. *Physical therapy*. 2000;80(8):782-807.
105. Eng JJ, Dawson AS, Chu KS. Submaximal exercise in persons with stroke: test-retest reliability and concurrent validity with maximal oxygen consumption. *Archives of physical medicine and rehabilitation*. 2004;85(1):113-8.
106. Gore CJ, Booth ML, Bauman A, Owen N. Utility of pwc75% as an estimate of aerobic power in epidemiological and population-based studies. *Medicine and science in sports and exercise*. 1999;31(2):348-51.
107. Hollmann W, Strüder H, Predel H-G, Tagarakis C. Spiroergometrie. *Kardiopulmonale Leistungsdiagnostik des Gesunden und Kranken*: Schattauer; 2006.
108. Batcho CS, Thonnard JL, Nielens H. PWC 75%/kg, a fitness index not linked to resting heart rate: testing procedure and reference values. *Archives of physical medicine and rehabilitation*. 2012;93(7):1196-200.
109. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*. 2001;37(1):153-6.
110. Reifman A, Keyton K. Winsorize. In N. J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 1636-1638). Thousand Oaks, CA: Sage; 2010.
111. Finger J, Gößwald A, Härtel S, Müters S, Krug S, Hölling H, et al. Measurement of cardiorespiratory fitness in the German Health Interview and Examination Survey for Adults (DEGS1). *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*. 2013;56(5-6):885-93.
112. Finger J, Krug S, Gößwald A, Härtel S, Bös K. Cardiorespiratory fitness among adults in Germany. *Robert Koch-Institut, Epidemiologie und Gesundheitsberichterstattung*; 2013.
113. Finger JD, Banzer W, Baumeister SE, Brandes M, Bös K, Gabrys L, et al. [Reference Values for Cardiorespiratory Fitness of the General Population: The German National Health Interview and Examination Survey for Adults (DEGS1) 2008-2011]. *Gesundheitswesen*. 2019.
114. Baumeister SE, Ricci C, Kohler S, Fischer B, Topfer C, Finger JD, et al. Physical activity surveillance in the European Union: reliability and validity of the European Health Interview Survey-Physical Activity Questionnaire (EHIS-PAQ). *The international journal of behavioral nutrition and physical activity*. 2016;13:61.
115. Niedermann K, Sidelnikov E, Muggli C, Dagfinrud H, Hermann M, Tamborrini G, et al. Effect of cardiovascular training on fitness and perceived disease activity in people with ankylosing spondylitis. *Arthritis care & research*. 2013;65(11):1844-52.
116. Nagashima J, Musha H, Takada H, Takagi K, Mita T, Mochida T, et al. Three-month exercise and weight loss program improves heart rate recovery in obese persons along with cardiopulmonary function. *Journal of cardiology*. 2010;56(1):79-84.

117. UK Biobank. Cancer data: linkage from national cancer registries. Version 1.4. 2013.
118. UK Biobank. Mortality data: linkage from national death registries. Version 1.3. 2013.
119. Møller H, Richards S, Hanchett N, Riaz SP, Lüchtenborg M, Holmberg L, et al. Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. *British journal of cancer*. 2011;105(1):170-6.
120. World Health Organization (WHO). ICD-10 : international statistical classification of diseases and related health problems / World Health Organization. Geneva: World Health Organization; 2004.
121. World Health Organization (WHO). ICD-9 : international statistical classification of diseases and related health problems / World Health Organization. Geneva: World Health Organization; 1978.
122. Fitz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, et al. International Classification of Diseases for Oncology, Third edition (ICD-O-3): World Health Organization 2000.
123. Hernán MA, Robins JM. (2019). *Causal Inference*: Boca Raton: Chapman & Hall/CRC, forthcoming.
124. Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GTH. Robust causal inference using directed acyclic graphs: the R package ‘dagitty’. *International Journal of Epidemiology*. 2016;45(6):1887-94.
125. VanderWeele TJ. Principles of confounder selection. *European journal of epidemiology*. 2019;34(3):211-9.
126. VanderWeele TJ, Shpitser I. A New Criterion for Confounder Selection. *Biometrics*. 2011;67(4):1406-13.
127. Pearl J, Mackenzie D. *The Book of Why: The New Science of Cause and Effect*: Basic Books, Inc.; 2018.
128. Taylor DP, Burt RW, Williams MS, Haug PJ, Cannon-Albright LA. Population-based family history-specific risks for colorectal cancer: a constellation approach. *Gastroenterology*. 2010;138(3):877-85.
129. Liang PS, Chen TY, Giovannucci E. Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis. *International journal of cancer*. 2009;124(10):2406-15.
130. Fedirko V, Tramacere I, Bagnardi V, Rota M, Scotti L, Islami F, et al. Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. *Annals of oncology : official journal of the European Society for Medical Oncology*. 2011;22(9):1958-72.
131. Song M, Garrett WS, Chan AT. Nutrients, foods, and colorectal cancer prevention. *Gastroenterology*. 2015;148(6):1244-60 e16.
132. Chan DS, Lau R, Aune D, Vieira R, Greenwood DC, Kampman E, et al. Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. *PloS one*. 2011;6(6):e20456.
133. Ma Y, Yang Y, Wang F, Zhang P, Shi C, Zou Y, et al. Obesity and risk of colorectal cancer: a systematic review of prospective studies. *PloS one*. 2013;8(1):e53916.
134. Khankari NK, Shu XO, Wen W, Kraft P, Lindstrom S, Peters U, et al. Association between Adult Height and Risk of Colorectal, Lung, and Prostate Cancer: Results from Meta-analyses of Prospective Studies and Mendelian Randomization Analyses. *PLoS medicine*. 2016;13(9):e1002118.
135. Kim SE, Paik HY, Yoon H, Lee JE, Kim N, Sung MK. Sex- and gender-specific disparities in colorectal cancer risk. *World journal of gastroenterology*. 2015;21(17):5167-75.
136. Guraya SY. Association of type 2 diabetes mellitus and the risk of colorectal cancer: A meta-analysis and systematic review. *World journal of gastroenterology*. 2015;21(19):6026-31.
137. Jiang Y, Ben Q, Shen H, Lu W, Zhang Y, Zhu J. Diabetes mellitus and incidence and mortality of colorectal cancer: a systematic review and meta-analysis of cohort studies. *Eur J Epidemiol*. 2011;26(11):863-76.
138. Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, et al. Diabetes and cancer: a consensus report. *Diabetes care*. 2010;33(7):1674-85.

139. Morris JS, Bradbury KE, Cross AJ, Gunter MJ, Murphy N. Physical activity, sedentary behaviour and colorectal cancer risk in the UK Biobank. *British journal of cancer*. 2018;118(6):920-9.
140. Algra AM, Rothwell PM. Effects of regular aspirin on long-term cancer incidence and metastasis: a systematic comparison of evidence from observational studies versus randomised trials. *The Lancet Oncology*. 2012;13(5):518-27.
141. Bosetti C, Rosato V, Gallus S, Cuzick J, La Vecchia C. Aspirin and cancer risk: a quantitative review to 2011. *Annals of oncology : official journal of the European Society for Medical Oncology*. 2012;23(6):1403-15.
142. Rothwell PM, Fowkes FG, Belch JF, Ogawa H, Warlow CP, Meade TW. Effect of daily aspirin on long-term risk of death due to cancer: analysis of individual patient data from randomised trials. *Lancet (London, England)*. 2011;377(9759):31-41.
143. Doubeni CA, Laiyemo AO, Reed G, Field TS, Fletcher RH. Socioeconomic and racial patterns of colorectal cancer screening among Medicare enrollees in 2000 to 2005. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2009;18(8):2170-5.
144. Limsui D, Vierkant RA, Tillmans LS, Wang AH, Weisenberger DJ, Laird PW, et al. Postmenopausal hormone therapy and colorectal cancer risk by molecularly defined subtypes among older women. *Gut*. 2012;61(9):1299-305.
145. Lin KJ, Cheung WY, Lai JY, Giovannucci EL. The effect of estrogen vs. combined estrogen-progestogen therapy on the risk of colorectal cancer. *International journal of cancer*. 2012;130(2):419-30.
146. de Borja AT, Jost RT, Gass R, Nedel FB, Cardoso DM, Pohl HH, et al. The influence of active and passive smoking on the cardiorespiratory fitness of adults. *Multidisciplinary respiratory medicine*. 2014;9(1):34.
147. Mondal H, Mishra SP. Effect of BMI, Body Fat Percentage and Fat Free Mass on Maximal Oxygen Consumption in Healthy Young Adults. *Journal of clinical and diagnostic research : JCDR*. 2017;11(6):CC17-CC20.
148. Loprinzi PD, Pariser G. Cardiorespiratory fitness levels and its correlates among adults with diabetes. *Cardiopulmonary physical therapy journal*. 2013;24(2):27-34.
149. Ombrellaro KJ, Perumal N, Zeiher J, Hoebel J, Itermann T, Ewert R, et al. Socioeconomic Correlates and Determinants of Cardiorespiratory Fitness in the General Adult Population: a Systematic Review and Meta-Analysis. *Sports medicine - open*. 2018;4(1):25.
150. Tarp J, Stole AP, Blond K, Grontved A. Cardiorespiratory fitness, muscular strength and risk of type 2 diabetes: a systematic review and meta-analysis. *Diabetologia*. 2019;62(7):1129-42.
151. Laryea JA, Siegel E, Klimberg S. Racial disparity in colorectal cancer: the role of equal treatment. *Diseases of the colon and rectum*. 2014;57(3):295-302.
152. Polite BN, Dignam JJ, Olopade OI. Colorectal cancer and race: understanding the differences in outcomes between African Americans and whites. *The Medical clinics of North America*. 2005;89(4):771-93.
153. Sabounchi S, Keihanian S, Anand BS. Impact of race on colorectal cancer. *Clinical colorectal cancer*. 2012;11(1):66-70.
154. Lee JY, Hong SN, Kim JH, Choe WH, Lee SY, Sung IK, et al. Risk for coronary heart disease increases risk for colorectal neoplasm. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*. 2013;11(6):695-702.
155. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011;45(3):1-67.
156. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research*. 2011;20(1):40-9.
157. White IR, Royston P. Imputing missing covariate values for the Cox model. *Statistics in Medicine*. 2009;28(15):1982-98.
158. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *American journal of epidemiology*. 1997;145(1):72-80.
159. Cologne J, Hsu WL, Abbott RD, Ohishi W, Grant EJ, Fujiwara S, et al. Proportional hazards regression in epidemiologic follow-up studies: an intuitive consideration of primary time scale. *Epidemiology (Cambridge, Mass)*. 2012;23(4):565-73.

160. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*: Springer International Publishing; 2015.
161. Wang M, Spiegelman D, Kuchiba A, Lochhead P, Kim S, Chan AT, et al. Statistical methods for studying disease subtype heterogeneity. *Stat Med*. 2016;35(5):782-800.
162. Haneuse S, VanderWeele TJ, Arterburn D. Using the E-Value to Assess the Potential Effect of Unmeasured Confounding in Observational Studies. *JAMA*. 2019;321(6):602-3.
163. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: Introducing the e-value. *Annals of internal medicine*. 2017;167(4):268-74.
164. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. 2016.
165. Celis-Morales CA, Lyall DM, Anderson J, Iliodromiti S, Fan Y, Ntuki UE, et al. The association between physical activity and risk of mortality is modulated by grip strength and cardiorespiratory fitness: evidence from 498 135 UK-Biobank participants. *European Heart Journal*. 2017;38(2):116-22.
166. Vella CA, Van Guilder GP, Dalleck LC. Low Cardiorespiratory Fitness Is Associated with Markers of Insulin Resistance in Young, Normal Weight, Hispanic Women. *Metabolic syndrome and related disorders*. 2016;14(5):272-8.
167. Estaki M, Pither J, Baumeister P, Little JP, Gill SK, Ghosh S, et al. Cardiorespiratory fitness as a predictor of intestinal microbial diversity and distinct metagenomic functions. *Microbiome*. 2016;4(1):42.
168. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. *Journal of the National Cancer Institute*. 2013;105(24):1907-11.
169. Campbell KL, McTiernan A, Li SS, Sorensen BE, Yasui Y, Lampe JW, et al. Effect of a 12-Month Exercise Intervention on the Apoptotic Regulating Proteins Bax and Bcl-2 in Colon Crypts: A Randomized Controlled Trial. *Cancer Epidemiology Biomarkers Prevention*. 2007;16(9):1767-74.
170. Moreno-Villanueva M, Kramer A, Hammes T, Venegas-Carro M, Thumm P, Bürkle A, et al. Influence of Acute Exercise on DNA Repair and PARP Activity before and after Irradiation in Lymphocytes from Trained and Untrained Individuals. *Int J Mol Sci*. 2019;20(12):2999.
171. Missiaglia E, Jacobs B, D'Ario G, Di Narzo AF, Sonesson C, Budinska E, et al. Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features. *Annals of oncology : official journal of the European Society for Medical Oncology*. 2014;25(10):1995-2001.
172. Murphy N, Strickler HD, Stanczyk FZ, Xue X, Wassertheil-Smoller S, Rohan TE, et al. A Prospective Evaluation of Endogenous Sex Hormone Levels and Colorectal Cancer Risk in Postmenopausal Women. *Journal of the National Cancer Institute*. 2015;107(10):djv210.
173. Song M, Hu FB, Spiegelman D, Chan AT, Wu K, Ogino S, et al. Long-term status and change of body fat distribution, and risk of colorectal cancer: a prospective cohort study. *International journal of epidemiology*. 2016;45(3):871-83.
174. Sarri G, Davies M, Lumsden MA. Diagnosis and management of menopause: summary of NICE guidance. *BMJ : British Medical Journal*. 2015;351:h5746.
175. Conn VS, Hafdahl A, Phillips LJ, Ruppert TM, Chase J-AD. Impact of physical activity interventions on anthropometric outcomes: systematic review and meta-analysis. *The journal of primary prevention*. 2014;35(4):203-15.
176. Jackson AS, Sui X, Hebert JR, Church TS, Blair SN. Role of lifestyle and aging on the longitudinal change in cardiorespiratory fitness. *Archives of internal medicine*. 2009;169(19):1781-7.
177. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, 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. *Medicine and science in sports and exercise*. 2011;43(7):1334-59.
178. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*. 2013;14:365.

179. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* (Cambridge, Mass). 2008;19(5):640-8.
180. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *BMJ*. 2020;368:m131.
181. Wild CP, Weiderpass E, Stewart BW, editors (2020). *World Cancer Report: Cancer Research for Cancer Prevention*. Lyon, France: International Agency for Research on Cancer. Available from: <http://publications.iarc.fr/586>. Licence: CC BY-NC-ND3.0 IGO.
182. Rastrelli M, Tropea S, Rossi CR, Alaibac M. Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification. *In Vivo*. 2014;28(6):1005-11.
183. Ali Z, Yousaf N, Larkin J. Melanoma epidemiology, biology and prognosis. *EJC Suppl*. 2013;11(2):81-91.
184. Behrens G, Niedermaier T, Berneburg M, Schmid D, Leitzmann MF. Physical activity, cardiorespiratory fitness and risk of cutaneous malignant melanoma: Systematic review and meta-analysis. *PloS one*. 2018;13(10):e0206087.
185. Jardine A, Bright M, Knight L, Perina H, Vardon P, Harper C. Does physical activity increase the risk of unsafe sun exposure? *Health Promot J Austr*. 2012;23(1):52-7.
186. Lauretani F, Russo CR, Bandinelli S, Bartali B, Cavazzini C, Di Iorio A, et al. Age-associated changes in skeletal muscles and their effect on mobility: an operational diagnosis of sarcopenia. *Journal of applied physiology* (Bethesda, Md : 1985). 2003;95(5):1851-60.
187. Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ) – Short and Long Forms. Revised November 2005. <https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnx0aGVpcGFxfGd4OjE0NDgxMDk3NDU1YWRIZTM>. cited 2020/05/20.
188. UK Biobank. Grip Strength Measurement. Version 1.0. 2011. <http://biobank.ndph.ox.ac.uk/showcase/showcase/docs/Gripstrength.pdf>.
189. Prince SA, LeBlanc AG, Colley RC, Saunders TJ. Measurement of sedentary behaviour in population health surveys: a review and recommendations. *PeerJ*. 2017;5:e4130.
190. Ghiasvand R, Robsahm TE, Green AC, Rueegg CS, Weiderpass E, Lund E, et al. Association of Phenotypic Characteristics and UV Radiation Exposure With Risk of Melanoma on Different Body Sites. *JAMA Dermatol*. 2019;155(1):39-49.
191. Bradbury KE, Murphy N, Key TJ. Diet and colorectal cancer in UK Biobank: a prospective study. *Int J Epidemiol*. 2020;49(1):246-58.
192. Surdu S, Fitzgerald EF, Bloom MS, Boscoe FP, Carpenter DO, Haase RF, et al. Occupational exposure to ultraviolet radiation and risk of non-melanoma skin cancer in a multinational European study. *PloS one*. 2013;8(4):e62359.
193. Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung. GENESIS-UV: Available Results. <https://www.dguv.de/ifa/fachinfos/strahlung/genesis-uv/aktuelle-ergebnisse/index-2.jsp>. cited 2020/08/12.
194. Micek A, Godos J, Lafranconi A, Marranzano M, Pajak A. Caffeinated and decaffeinated coffee consumption and melanoma risk: a dose-response meta-analysis of prospective cohort studies. *Int J Food Sci Nutr*. 2018;69(4):417-26.
195. Lee PY, Silverman MK, Rigel DS, Vossaert KA, Kopf AW, Bart RS, et al. Level of education and the risk of malignant melanoma. *J Am Acad Dermatol*. 1992;26(1):59-63.
196. World Cancer Research Fund/ American Institute for Cancer Research. Continuous Update Project Expert Report 2018. Diet, nutrition, physical activity and skin cancer. Available at dietandcancerreport.org.
197. Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Publications and Reports of the Surgeon General. How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Atlanta (GA): Centers for Disease Control and Prevention (US); 2010.
198. Oliveria SA, Saraiya M, Geller AC, Heneghan MK, Jorgensen C. Sun exposure and risk of melanoma. *Arch Dis Child*. 2006;91(2):131-8.
199. Suppa M, Gandini S. Sunbeds and melanoma risk: time to close the debate. *Current Opinion in Oncology*. 2019;31(2).

200. German Guideline Program in Oncology (GGPO). Evidence-based Guideline on Prevention of Skin Cancer. Version 1.1 – April 2014. AWMF registration number: 032/052GGPO. Guideline (Long Version).
201. Gandini S, Sera F, Cattaruzza MS, Pasquini P, Picconi O, Boyle P, et al. Meta-analysis of risk factors for cutaneous melanoma: II. Sun exposure. *European journal of cancer (Oxford, England : 1990)*. 2005;41(1):45-60.
202. Dusingize JC, Olsen CM, An J, Pandeya N, Law MH, Thompson BS, et al. Body mass index and height and risk of cutaneous melanoma: Mendelian randomization analyses. *Int J Epidemiol*. 2020.
203. Chandrasekaran B, Ghosh A, Prasad C, Krishnan K, Chandrasha B. Age and anthropometric traits predict handgrip strength in healthy normals. *J Hand Microsurg*. 2010;2(2):58-61.
204. Ploegmakers JJ, Hepping AM, Geertzen JH, Bulstra SK, Stevens M. Grip strength is strongly associated with height, weight and gender in childhood: a cross sectional study of 2241 children and adolescents providing reference values. *J Physiother*. 2013;59(4):255-61.
205. Wang Y-C, Bohannon RW, Li X, Sindhu B, Kapellusch J. Hand-Grip Strength: Normative Reference Values and Equations for Individuals 18 to 85 Years of Age Residing in the United States. *Journal of Orthopaedic & Sports Physical Therapy*. 2018;48(9):685-93.
206. Schaefer L, Plotnikoff RC, Majumdar SR, Mollard R, Woo M, Sadman R, et al. Outdoor time is associated with physical activity, sedentary time, and cardiorespiratory fitness in youth. *J Pediatr*. 2014;165(3):516-21.
207. Lawler S, Sugiyama T, Owen N. Sun exposure concern, sun protection behaviors and physical activity among Australian adults. *Cancer Causes & Control*. 2007;18(9):1009-14.
208. Autier P. Sunscreen abuse for intentional sun exposure. *British Journal of Dermatology*. 2009;161(s3):40-5.
209. Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M, Becher H, Binder H, et al. State of the art in selection of variables and functional forms in multivariable analysis-outstanding issues. *Diagn Progn Res*. 2020;4:3.
210. Lorenz E, Jenkner C, Sauerbrei W, Becher H. Modeling Variables With a Spike at Zero: Examples and Practical Recommendations. *American journal of epidemiology*. 2017;185(8):650-60.
211. Pearce M, Strain T, Kim Y, Sharp SJ, Westgate K, Wijndaele K, et al. Estimating physical activity from self-reported behaviours in large-scale population studies using network harmonisation: findings from UK Biobank and associations with disease outcomes. *International Journal of Behavioral Nutrition and Physical Activity*. 2020;17(1):40.
212. McKenzie R, Bodeker G, Scott G, Slusser J, Lantz K. Geographical differences in erythemally-weighted UV measured at mid-latitude USDA sites. *Photochem Photobiol Sci*. 2006;5(3):343-52.
213. Met Office. UV index forecast. <https://www.metoffice.gov.uk/public/weather/uv-index-forecast/>. cited 2020/09/24.
214. Wolff D, Fitzhugh EC. The relationships between weather-related factors and daily outdoor physical activity counts on an urban greenway. *International journal of environmental research and public health*. 2011;8(2):579-89.
215. Linos E, Li WQ, Han J, Li T, Cho E, Qureshi AA. Lifetime ultraviolet radiation exposure and lentigo maligna melanoma. *The British journal of dermatology*. 2017;176(6):1666-8.
216. Walter SD, King WD, Marrett LD. Association of cutaneous malignant melanoma with intermittent exposure to ultraviolet radiation: results of a case-control study in Ontario, Canada. *Int J Epidemiol*. 1999;28(3):418-27.
217. Yang TO, Reeves GK, Green J, Beral V, Cairns BJ. Birth weight and adult cancer incidence: large prospective study and meta-analysis. *Annals of oncology : official journal of the European Society for Medical Oncology*. 2014;25(9):1836-43.
218. Dodds R, Denison HJ, Ntani G, Cooper R, Cooper C, Sayer AA, et al. Birth weight and muscle strength: a systematic review and meta-analysis. *J Nutr Health Aging*. 2012;16(7):609-15.
219. Strain T, Wijndaele K, Sharp SJ, Dempsey PC, Wareham N, Brage S. Impact of follow-up time and analytical approaches to account for reverse causality on the association between physical activity and health outcomes in UK Biobank. *International Journal of Epidemiology*. 2019;49(1):162-72.

9 Publikationsliste

Publikation im direkten Zusammenhang mit der vorliegenden Dissertation:

- **Hillreiner A**, Baumeister SE, Sedlmeier AM, Finger JD, Schlitt HJ, Leitzmann MF. Association between cardiorespiratory fitness and colorectal cancer in the UK Biobank. *Eur J Epidemiol.* 2020;35(10):961-73. doi: 10.1007/s10654-019-00575-6. Epub 2019 Nov 9.
- **Weber A**, Leitzmann MF, Sedlmeier AM, Baurecht H, Jochem C, Haferkamp S, Baumeister SE. Association between physical activity, grip strength, and sedentary behaviour with incidence of malignant melanoma: results from the UK Biobank. *British journal of cancer.* 2021;125(4):593-600.

Co-Autorenschaften bei Publikationen zum Thema Messung von körperlicher Aktivität, sedentärem Verhalten oder körperlicher Fitness in epidemiologischen Studien:

- Matthews CE, Berrigan D, Fischer B, Gomersall SR, **Hillreiner A**, Kim Y, et al. Use of previous-day recalls of physical activity and sedentary behavior in epidemiologic studies: results from four instruments. *BMC public health.* 2019;19(2):478.
- Leitzmann M, Gastell S, **Hillreiner A**, Herbolsheimer F, Baumeister SE, Bohn B, et al. Körperliche Aktivität in der NAKO Gesundheitsstudie: erste Ergebnisse des multimodalen Erhebungskonzepts. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz.* 2020.
- Kluttig A, Zschocke J, Haerting J, Schmermund A, Gastell S, Steindorf K, ..., **Hillreiner A**, ..., et al. Messung der körperlichen Fitness in der NAKO Gesundheitsstudie – Methoden, Qualitätssicherung und erste deskriptive Ergebnisse. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz.* 2020.

Co-Autorenschaft bei Publikation zum Thema sedentäres Verhalten und Krebs:

- Berger FF, Leitzmann MF, **Hillreiner A**, Sedlmeier AM, Prokopidi-Danisch ME, Burger M, et al. Sedentary Behavior and Prostate Cancer: A Systematic Review and

Meta-Analysis of Prospective Cohort Studies. Cancer prevention research (Philadelphia, Pa). 2019;12(10):675-88.

Weitere Co-Autorenschaft:

- Sedlmeier AM, Baumeister SE, **Weber A**, Fischer B, Thorand B, Ittermann T, et al. Relation of body fat mass and fat-free mass to total mortality: results from 7 prospective cohort studies. The American journal of clinical nutrition. 2021.

Die vorliegende Arbeit wurde mit Daten der NAKO Gesundheitsstudie durchgeführt (www.nako.de). Die NAKO Gesundheitsstudie wird durch das Bundesministerium für Bildung und Forschung (BMBF) [Förderkennzeichen 01ER1301A/B/C und 01ER1511D], die Bundesländer und die Helmholtz Gemeinschaft gefördert sowie durch die beteiligten Universitäten und Institute der Leibniz-Gemeinschaft finanziell unterstützt.

Ich danke allen Teilnehmerinnen und Teilnehmern der NAKO Gesundheitsstudie sowie den Mitarbeiterinnen und Mitarbeitern der NAKO Gesundheitsstudie.

Des Weiteren wurden Daten der UK Biobank verwendet. Die UK Biobank wurde vom Wellcome Trust, dem Medical Research Council, dem Department of Health, der schottischen Regierung und der Northwest Regional Development Agency unterstützt. Außerdem wurde sie finanziert von der walisischen Regierung und der British Heart Foundation.

Ich danke allen UK Biobank Teilnehmenden und dem Studienpersonal der UK Biobank.

10 Danksagung

Ich möchte allen ganz herzlich danken, die mich während meiner Promotionszeit in besonderem Maße unterstützt, gefördert und motiviert haben!

Mein besonderer Dank gilt meinem Betreuer Herrn Prof. Michael Leitzmann, der es mir ermöglicht hat, diese spannende Dissertation am Institut für Epidemiologie und Präventivmedizin anzufertigen und mir immer fachkundig zur Seite stand. Durch die sehr konstruktive, motivierende und inspirierende Betreuung sowie die Möglichkeit an vielen Konferenzen, Workshops und Fortbildungen teilzunehmen konnte ich sehr viel lernen!

Bei meinem Mentor Herrn Prof. Sebastian Baumeister bedanke ich mich sehr herzlich für die große Hilfe bei der Manuskripterstellung und die fortwährende Versorgung mit spannenden fortgeschrittenen statistischen Methoden! Meinem Mentor Herrn Prof. Michael Nerlich möchte ich für die wertvollen und hilfreichen Diskussionen in den Promotionskolloquien danken. Zu dem möchte ich mich ganz herzlich bei den Gutachtern Herrn Prof. Freerk Baumann und Herrn Prof. Hajo Zeeb, sowie den weiteren Mitgliedern der Prüfungskommission Frau Prof. Barbara Dobler und Frau Prof. Anne Herrmann-Johns bedanken.

Weiterhin danke ich dem gesamten NAKO Team Regensburg unter der erfolgreichen Leitung von Frau Dr. Beate Fischer und Frau Christine Töpfer, ohne deren unermüdlichen Einsatz die NAKO Gesundheitsstudie und damit diese Arbeit nicht möglich gewesen wären.

Sylvia Pietsch möchte ich dafür danken, dass sie immer alles möglich macht und immer ein offenes Ohr hat!

Vielen Dank auch Herrn Dr. Hansjörg Baurecht und Herrn Dr. Vincent van Hees, für die sehr angenehme, spannende und lehrreiche Zusammenarbeit, sowie Frau Dr. Carmen Jochem, für ihre ansteckende Motivation und Begeisterung bei der Lehre und Betreuung.

Ein großes Dankeschön an Alexandra Sauter und Magdalena Rohr, für die vielen schönen und aufbauenden Young-Researcher Treffen, und an Anja Sedlmeier, die vom ersten Tag an absolut unverzichtbar war! Ihr habt die Promotionszeit zu einer sehr schönen Zeit gemacht!

Tausend Dank an meine Eltern Inge und Herbert, die mich immer unterstützt, motiviert und geliebt, mir aber auch alle Freiheiten gelassen haben. Ohne Euch wäre ich heute nicht hier! Meinem Mann Martin möchte ich danken, dass er auch in anstrengenden Zeiten immer für mich da ist, mich immer aufbaut, zum Lachen bringt und sein Leben mit mir teilt. Danke für alles!

11 Lebenslauf

Persönliche Daten

Name	Andrea Weber (geb. Hillreiner)
Geburtstag und -ort	26. August 1991 in Dachau
Nationalität	Deutsch
Familienstand	Verheiratet
Email	andrea.weber@ukr.de

Ausbildungsdaten

Seit 01/2018	Promotion zum Dr. sc. hum. (Humanwissenschaften) Universität Regensburg, Fakultät für Medizin Dissertationsschrift: „Messung von körperlicher Aktivität in epidemiologischen Studien und Assoziationen mit Krebsinzidenz“
10/2014 - 09/2016	Master of Public Health (MPH) Ludwig-Maximilians-Universität München (Abschlussnote: 1,2) Thema Masterarbeit: "Is dietary quality associated with utilisation, direct and indirect healthcare cost? Results from the MONICA/KORA Augsburg cohort study"
10/2011 - 09/2014	B.Sc. Ernährungswissenschaft Technische Universität München (Abschlussnote: 1,8) Thema Bachelorarbeit: "Comparison of Food Composition Tables on the basis of Food4Me FFQs"
09/2002 - 05/2011	Abitur Ignaz-Taschner-Gymnasium Dachau (Abschlussnote: 1,8)

Berufserfahrung

Seit 01/2017	Wissenschaftliche Mitarbeiterin Universität Regensburg, Fakultät für Medizin Institut für Epidemiologie und Präventivmedizin
09/2015 - 01/2016	Praktikum Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH) Institut für Gesundheitsökonomie und Management im Gesundheitswesen
10/2014 - 07/2016	Studentische Hilfskraft Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH) Institut für Gesundheitsökonomie und Management im Gesundheitswesen
09/2014 01/2014 - 02/2014 08/2013 - 10/2013	Studentische Hilfskraft Technische Universität München, Wissenschaftszentrum Weihenstephan Zentralinstitut für Ernährungs- und Lebensmittelforschung

Vorträge

01.10.2020	15. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi), Onlineveranstaltung (Präsentation)
------------	---

27.06.2019	6. International Conference on Ambulatory Monitoring of Physical Activity and Movement (ICAMPAM) in Maastricht, Niederlande (Poster)
06.06.2019	18. Konferenz der International Society of Behavioral Nutrition and Physical Activity (ISBNPA) in Prag, Tschechische Republik (Präsentation)
02.10.2018	2. Tag der digitalen Lehre in Regensburg (Präsentation)
27.09.2018	13. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi) in Bremen (Poster)

Fortbildungen

02.10.2018 – 02.03.2020	Zertifikat Hochschullehre der Bayerischen Universitäten 120 Arbeitseinheiten aus den Bereichen Lehr-Lern-Konzepte, Präsentation und Kommunikation, Prüfen, Reflexion und Evaluation, Beraten und Begleiten
20.08.2018 – 24.08.2018	Intensive Short Course Advanced Epidemiologic Methods: Causal research and prediction modeling. Berlin School of Public Health. Charité Berlin. (Prof. Rolf Groenwold, Dr. Maarten van Smeden)
18.06.2018 – 22.06.2018	Blockkurs Fortgeschrittene Methoden in der Biostatistik in R. Universität Regensburg (Prof. Florian Hartig)
Seit 11/2017	HarvardX (online learning initiative of Harvard University mit edX) Causal Diagrams: Draw your assumptions before your conclusions (Prof. Miguel Hernán) Data Science: Visualization (Prof. Rafael Irizarry) Data Science: Wrangling (Prof. Rafael Irizarry) Data Science: Machine Learning (Prof. Rafael Irizarry)
07.08.2017 – 11.08.2017	Intensive Short Course Advanced Epidemiologic Methods. Berlin School of Public Health. Charité Berlin. (Prof. Matthew Fox)
WiSe 2015/2016	Seminar Betriebliche Gesundheitsförderung. Otto-Friedrich-Universität Bamberg für die Virtuelle Hochschule Bayern. (Prof. Jörg Wolstein)

Sonstiges

09/2019	Rapid Scoping Review für die 2020 WHO Guidelines on physical activity and sedentary behaviour
Seit 04/2018	Mitgliedschaft bei der Deutschen Gesellschaft für Epidemiologie
Seit 02/2018	Mitglied in der NAKO Expertengruppe körperliche Aktivität und Fitness

Zusätzliche Kenntnisse

Sprachen	Deutsch (Muttersprache), Englisch (C2), Französisch (Grundkenntnisse)
EDV	Sehr gute Kenntnisse in R, Microsoft Word, Microsoft Excel, Microsoft Powerpoint, Endnote Gute Kenntnisse in SAS, LaTeX Grundkenntnisse in Visual Basic, Python

12 Selbstständigkeitserklärung

Ich, Andrea Weber, Geburtsname Hillreiner, geboren am 26.08.1991 in Dachau erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Insbesondere habe ich nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ort, Datum

Andrea Weber