

**Ludwig-Maximilians-Universität
Fakultät für Mathematik, Informatik und Statistik
Institut für Statistik**



**Auswirkungen des Rauchens auf die Lungenfunktion
von Kindern und Jugendlichen
– Eine Auswertung im Rahmen der SOLAR II-Studie –**

Bachelorarbeit

in Zusammenarbeit mit
dem Institut und der Poliklinik für Arbeits-, Sozial- und Umweltmedizin
des Klinikums der Universität München

vorgelegt von
Martin Bartke
2009

Betreuer/-in:

PD Dr. Christian Heumann, Ludwig-Maximilians-Universität München

Prof. Dr. Katja Radon, Dr. Sabine Heinrich, Institut und Poliklinik für Arbeits-, Sozial- und Umweltmedizin des Klinikums der Universität München

Danksagung

Die vorliegende Bachelorarbeit zur Thematik „*Auswirkungen des Rauchens auf die Lungenfunktion von Kindern und Jugendlichen – Eine Auswertung im Rahmen der SOLAR II-Studie*“ entstand in Zusammenarbeit mit dem Institut für Statistik sowie dem Institut und der Poliklinik für Arbeits-, Sozial- und Umweltmedizin des Klinikums München.

An dieser Stelle möchte ich die Gelegenheit nutzen und mich bei meinen Betreuern Herrn PD Dr. Christian Heumann, Frau Prof. Dr. Radon und Frau Dr. Sabine Heinrich für die freundliche sowie sehr hilfreiche Betreuung und Beratung recht herzlich bedanken.

Inhaltsverzeichnis

ABBILDUNGSVERZEICHNIS.....	V
TABELLENVERZEICHNIS	VII
EINLEITUNG.....	1
1. TABAKKONSUM BEI JUGENDLICHEN	3
1.1 Tabakkonsum in der Bundesrepublik Deutschland	3
1.2 Forschungshintergrund der ISAAC- und SOLAR-Studien.....	6
1.2.1 ISAAC-Studien	7
1.2.2 SOLAR-Studien	9
2. DATENMANAGEMENT	13
2.1 Variablenselektion aus der ISAAC II- und SOLAR II-Studie	13
2.1.1 Zielgrößen.....	13
2.1.2 Potentielle Einflussgrößen	14
2.1.3 Potentielle Confounder	15
2.2 Datenbereinigung.....	18
2.3 Berechnung und Umcodierung zusätzlicher Variablen	19
2.3.1 Zusätzliche Zielgrößen	19
2.3.2 Zusätzliche potentielle Einflussgrößen	20
2.3.3 Zusätzliche potentielle Confounder	22

3.	DESKRIPTIVE AUSWERTUNGEN	24
3.1	Deskriptive univariate Analyse	25
3.1.1	Zielgrößen - Lungenfunktionsparameter	25
3.1.2	Potentielle Einflussgrößen - Aktives Rauchverhalten.....	26
3.1.3	Potentielle Einflussgrößen - Passives Rauchverhalten.....	28
3.1.4	Potentielle Confounder	28
3.2	Zusammenhangsanalyse.....	30
3.2.1	Einsekundenkapazität und potentielle Einflussgrößen/ Confounder	30
3.2.2	Forcierte Vitalkapazität und potentielle Einflussgrößen/ Confounder.....	38
3.2.3	Tiffeneau-Wert und potentielle Einflussgrößen/ Confounder	45
4.	MULTIPLE LINEARE REGRESSION.....	52
4.1	Modellannahmen	53
4.2	Parameterschätzung	54
4.3	Generalisiertes additives Modell (GAM).....	55
4.4	Variablenselektion auf Basis des AIC-Kriteriums	55
4.5	Kreuzvalidierung	57
4.6	Potentielle Einflussgrößen/ Confounder und Datenbasis.....	58
4.7	Ergebnisse für die Zielgröße „Einsekundenkapazität“	62
4.7.1	Ergebnisse für die gesamte Datenbasis	62
4.7.2	Ergebnisse für die männlichen Teilnehmer	63
4.7.3	Ergebnisse für die weiblichen Teilnehmer	64
4.8	Ergebnisse für die Zielgröße „forcierte Vitalkapazität“	66
4.8.1	Ergebnisse für die gesamte Datenbasis	66
4.8.2	Ergebnisse für die männlichen Teilnehmer	67
4.8.3	Ergebnisse für die weiblichen Teilnehmer	68
4.9	Ergebnisse für die Zielgröße „Tiffeneau-Wert“	69
4.9.1	Ergebnisse für die gesamte Datenbasis	69
4.9.2	Ergebnisse für die männlichen Teilnehmer	71
4.9.3	Ergebnisse für die weiblichen Teilnehmer	72

4.10	Überprüfung der Modellannahmen.....	73
5.	QUANTILREGRESSION	75
5.1	Modell und Schätzverfahren	75
5.2	Ergebnisse der Quantilregression.....	77
6.	ZUSAMMENFASSUNG	83
	LITERATURVERZEICHNIS	87
	ANHANG I: CD-INHALT	89
	EIDESSTATTLICHE ERKLÄRUNG	90

Abbildungsverzeichnis

<i>Abbildung 1: Weibliche und männliche Raucher zwischen 12 und 17 Jahren</i>	<i>3</i>
<i>Abbildung 2: Weibliche und männliche Nichtraucher zwischen 12 und 17 Jahren</i>	<i>4</i>
<i>Abbildung 3: Lebenszeit- und 30 Tage-Prävalenz des Tabakkonsums</i>	<i>5</i>
<i>Abbildung 4: Aufbau der ISAAC II-Studie</i>	<i>8</i>
<i>Abbildung 5: Aufbau der SOLAR I-Studie</i>	<i>10</i>
<i>Abbildung 6: Aufbau der SOLAR II-Studie</i>	<i>12</i>
<i>Abbildung 7: Job-Exposure-Matrix</i>	<i>17</i>
<i>Abbildung 8: Beantwortungsbeispiel zur passiven Tagesrauchexposition</i>	<i>21</i>
<i>Abbildung 9: Verteilung der Zielgrößen (Lungenfunktionsparameter)</i>	<i>25</i>
<i>Abbildung 10: Verteilung der potentiellen Einflussgrößen (aktives Rauchverhalten)</i>	<i>27</i>
<i>Abbildung 11: Aktives Rauchverhalten</i>	<i>31</i>
<i>Abbildung 12: Passivrauchexposition</i>	<i>33</i>
<i>Abbildung 13: Passivrauchexposition (kumuliert)</i>	<i>34</i>
<i>Abbildung 14: Geschlecht, Bildung und Erhebungszentrum</i>	<i>35</i>
<i>Abbildung 15: Körpergröße, Gewicht und Lungenfunktionsparameter (ISAAC II-Studie)</i>	<i>36</i>
<i>Abbildung 16: Asthma</i>	<i>38</i>
<i>Abbildung 17: Aktives Rauchverhalten</i>	<i>39</i>
<i>Abbildung 18: Passivrauchexposition</i>	<i>40</i>
<i>Abbildung 19: Passivrauchexposition (kumuliert)</i>	<i>41</i>
<i>Abbildung 20: Geschlecht, Bildung und Erhebungszentrum</i>	<i>42</i>
<i>Abbildung 21: Körpergröße, Gewicht und Lungenfunktionsparameter (ISAAC II-Studie)</i>	<i>43</i>
<i>Abbildung 22: Asthma</i>	<i>44</i>
<i>Abbildung 23: Aktives Rauchverhalten</i>	<i>45</i>
<i>Abbildung 24: Passivrauchexposition</i>	<i>46</i>

Abbildungsverzeichnis

<i>Abbildung 25: Passivrauchexposition (kumuliert)</i>	47
<i>Abbildung 26: Geschlecht, Bildung und Erhebungszentrum</i>	48
<i>Abbildung 27: Körpergröße, Gewicht und Lungenfunktionsparameter (ISAAC II-Studie)</i>	49
<i>Abbildung 28: Asthma</i>	51
<i>Abbildung 29: Geschätzte Funktionen für die stetigen Kovariablen</i>	61
<i>Abbildung 30: Quantil-Quantil Diagramme</i>	74

Tabellenverzeichnis

<i>Tabelle 1: Zielgrößen (SOLAR II-Studie)</i>	<i>14</i>
<i>Tabelle 2: Variablen zur Aktivrauchexposition (SOLAR II-Studie)</i>	<i>14</i>
<i>Tabelle 3: Variablen zur Passivrauchexposition (SOLAR II-Studie)</i>	<i>15</i>
<i>Tabelle 4: Confounder (SOLAR II-Studie)</i>	<i>16</i>
<i>Tabelle 5: Confounder (ISAAC II-Studie)</i>	<i>18</i>
<i>Tabelle 6: (Zusätzliche) Zielgrößen</i>	<i>19</i>
<i>Tabelle 7: (Zusatz-)Variablen zur Aktivrauchexposition</i>	<i>20</i>
<i>Tabelle 8: (Zusatz-)Variablen zur Passivrauchexposition</i>	<i>21</i>
<i>Tabelle 9: (Zusätzliche) Confounder</i>	<i>23</i>
<i>Tabelle 10: Verteilung der Zielgrößen (Lungenfunktionsparameter)</i>	<i>26</i>
<i>Tabelle 11: Verteilung der potentiellen Einflussgrößen (aktives Rauchverhalten)</i>	<i>27</i>
<i>Tabelle 12: Verteilung der potentiellen Einflussgrößen (passives Rauchverhalten)</i>	<i>28</i>
<i>Tabelle 13: Verteilung der potentiellen Confounder (demographische Variablen)</i>	<i>29</i>
<i>Tabelle 14: Verteilung der potentiellen Confounder (metrische Variablen)</i>	<i>29</i>
<i>Tabelle 15: Verteilung der potentiellen Confounder (kategoriale Variablen)</i>	<i>30</i>
<i>Tabelle 16: Korrelationen unter den kategorialen Variablen zur Passivrauchexposition</i>	<i>59</i>
<i>Tabelle 17: Aufteilung der Datenbasis („complete cases“)</i>	<i>60</i>
<i>Tabelle 18: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell</i>	<i>62</i>
<i>Tabelle 19: Gegenüberstellung der Modellwahlkriterien</i>	<i>63</i>
<i>Tabelle 20: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell</i>	<i>64</i>
<i>Tabelle 21: Gegenüberstellung der Modellwahlkriterien</i>	<i>64</i>
<i>Tabelle 22: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell</i>	<i>65</i>
<i>Tabelle 23: Gegenüberstellung der Modellwahlkriterien</i>	<i>65</i>
<i>Tabelle 24: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell</i>	<i>66</i>

Tabellenverzeichnis

<i>Tabelle 25: Gegenüberstellung der Modellwahlkriterien</i>	67
<i>Tabelle 26: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell</i>	67
<i>Tabelle 27: Gegenüberstellung der Modellwahlkriterien</i>	68
<i>Tabelle 28: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell</i>	69
<i>Tabelle 29: Gegenüberstellung der Modellwahlkriterien</i>	69
<i>Tabelle 30: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell</i>	70
<i>Tabelle 31: Gegenüberstellung der Modellwahlkriterien</i>	70
<i>Tabelle 32: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell</i>	71
<i>Tabelle 33: Gegenüberstellung der Modellwahlkriterien</i>	72
<i>Tabelle 34: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell</i>	72
<i>Tabelle 35: Gegenüberstellung der Modellwahlkriterien</i>	73
<i>Tabelle 36: Ergebnisse simultaner Quantilregressionen und einer OLS-Regression</i>	80
<i>Tabelle 37: Ergebnisse simultaner Quantilregressionen und einer OLS-Regression</i>	81
<i>Tabelle 38: Ergebnisse simultaner Quantilregressionen und einer OLS-Regression</i>	82

Einleitung

Im Jahr 2006 verstarben in der Bundesrepublik Deutschland 42.348 Menschen an Krebserkrankungen, welche in Bezug zum Tabakkonsum gebracht werden können. Im Jahr 2007 ist die Anzahl an Sterbefällen bundesweit um ca. 1,6% auf 43.005 Personen gestiegen.¹ Dies entsprach einem Anteil von ca. 5,2% aller Todesfälle (vgl. Statistisches Bundesamt (2007)). Im Vergleich zum Jahr 2000 sind damit ca. 6% mehr Personen an Krebserkrankungen der Lunge bzw. der Bronchien, des Kehlkopfs oder der Luftröhre gestorben. Darüber hinaus ist davon auszugehen, dass bei einer Vielzahl von weiteren Todesfällen, z.B. bei Verstorbenen mit Herz-Kreislaufkrankungen oder chronisch obstruktiven Lungenerkrankungen (COPD) der Konsum von Tabakprodukten ebenfalls zur Todesursache beigetragen hat (vgl. Statistisches Bundesamt (2008)).²

Die Regierung der Bundesrepublik Deutschland hat aufgrund der Vielzahl an Erkrankungen, welche ursächlich in Zusammenhang mit dem Rauchen gebracht werden können, die Reduzierung des Tabakkonsums in der Gesellschaft zu einem vordringlichen gesundheitspolitischen Ziel erklärt. Bekräftigt wurde dieses Vorhaben durch die Erklärungen der Tabakrahenkonvention der Weltgesundheitsorganisation (FCTC), welche von Deutschland im Jahr 2004 ratifiziert und im März 2005 in Kraft getreten ist. Gemäß dieser Tabakrahenkonvention haben sich die jeweiligen Vertragsparteien dazu verpflichtet, nationale Strategien und Programme zur Eindämmung des Tabakkonsums zu erarbeiten, umzusetzen und kontinuierlich zu aktualisieren.

¹ Davon verstarben im Jahr 2006 nach Angaben des Statistischen Bundesamtes 30.259 Männer (ca. 71,5%) bzw. 12.099 Frauen (ca. 28,5%) und im darauffolgenden Jahr 30.406 Männer (ca. 70,7%) bzw. 12.599 Frauen (ca. 29,3%) an Krebserkrankungen, welche in Zusammenhang mit dem Tabakkonsum gebracht werden können.

² Aktuelle Daten zur Todesursachenstatistik (Berichtsjahr 2008) des Statistischen Bundesamtes werden voraussichtlich erst zum Jahresende 2009 zur Verfügung stehen.

Auf politischer Ebene wurden zwischenzeitlich viele Maßnahmen getroffen, um den Tabakkonsum in der deutschen Gesellschaft zu reduzieren. Zu nennen sind bspw. die Durchführung von zwei Tabaksteuererhöhungen in den Jahren 2002 und 2005 sowie der Verabschiedung von Nichtraucherschutzgesetzen auf Bundes- und Landesebene im Jahr 2007. Auch hinsichtlich der Prävention des Tabakkonsums bei Jugendlichen wurden verschiedene Regelungen getroffen, wie z.B. die gesetzliche Verankerung eines Rauchverbots an Schulen in vielen Bundesländern sowie die Umstellung der Tabakautomaten auf eine Chipkartenbezahlung Anfang 2009 (vgl. Bericht der Fachgruppe „Suchtprävention“ im Auftrag des Drogen- und Suchtrates (2008, S. 1 ff)).

Vor diesem Hintergrund ist es die Zielsetzung der vorliegenden Arbeit, den Einfluss des Rauchens auf die Gesundheit, insbesondere auf die Lungenfunktion heranwachsender Jugendlicher unter Anpassung von Regressionsmodellen, eingehender zu untersuchen.

Zu diesem Zweck ist es das Ziel des **ersten Kapitels**, das Rauchverhalten von Kindern und Jugendlichen in Deutschland zunächst kurz zu beschreiben. Da die empirische Untersuchung der Aufgabenstellung auf Basis des Datenmaterials der Studien „ISAAC II“ und „SOLAR II“ erfolgt, soll der Forschungshintergrund bzw. der Aufbau dieser Studien im ersten Kapitel ebenfalls komprimiert dargestellt werden.

Darauf aufbauend behandelt das **zweite und dritte Kapitel** das Datenmanagement sowie die deskriptive Veranschaulichung des zugrunde liegenden Datenmaterials.

Inwieweit das Rauchen und ggf. weitere Faktoren, wie z.B. zuvor diagnostizierte Asthmaerkrankungen, einen Einfluss auf die Lungenfunktionstätigkeit nehmen, soll mit Hilfe der multiplen linearen Regression und der Quantilregression im **vierten und fünften Kapitel** untersucht werden.

Das **sechste Kapitel** umfasst die Zusammenfassung der gewonnenen Ergebnisse ab.

1. *Tabakkonsum bei Jugendlichen*

1.1 *Tabakkonsum in der Bundesrepublik Deutschland*

Die Bundeszentrale für gesundheitliche Aufklärung (BZgA) führt seit den 70-er Jahren in regelmäßigen Abständen Repräsentativerhebungen zur Drogenaffinität³ Jugendlicher und junger Erwachsener im Alter zwischen 12 und 17 Jahren durch.⁴

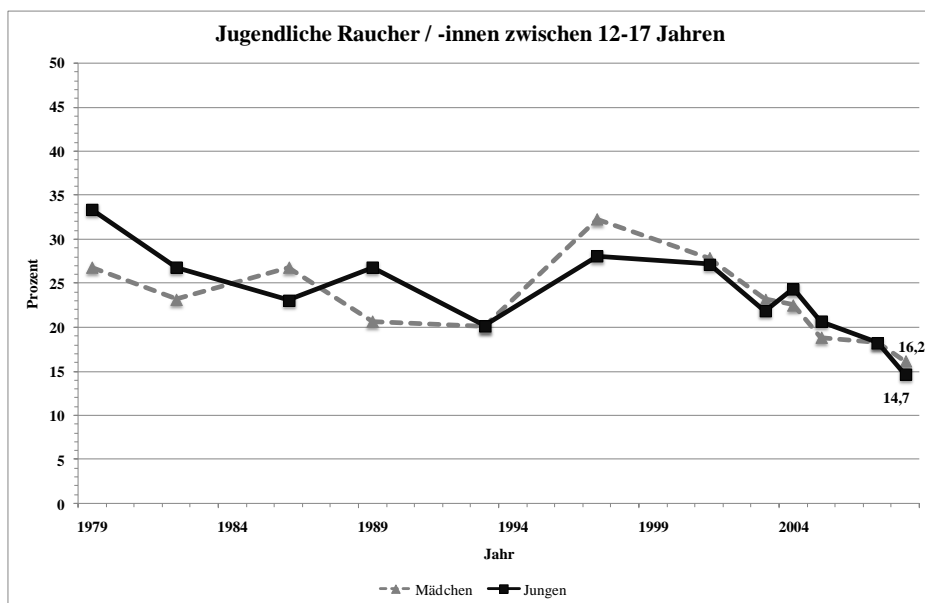
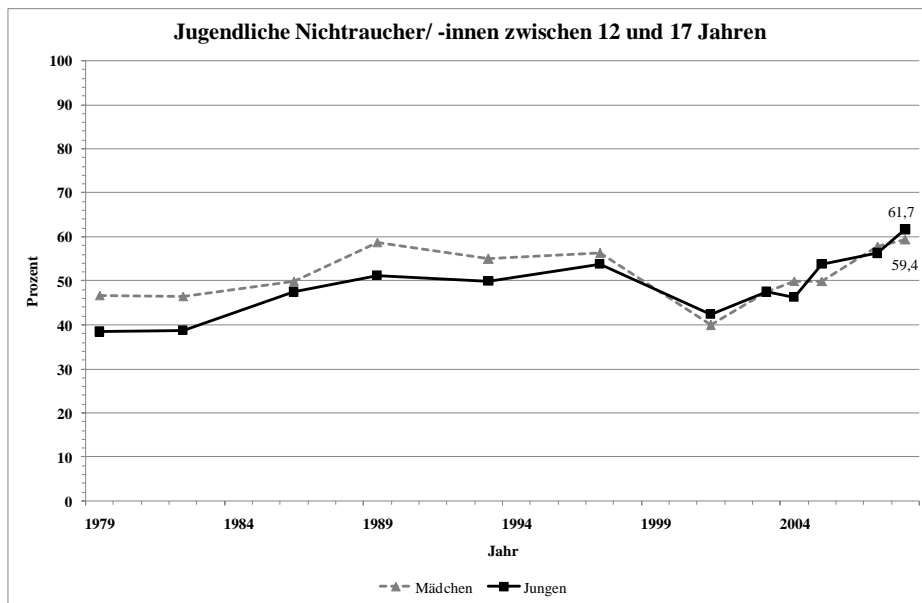


Abbildung 1: Weibliche und männliche Raucher zwischen 12 und 17 Jahren
(in Anlehnung an die Ergebnisse der Repräsentativerhebung des BZgA (2008, S. 8))

³ Unter dem Begriff der Drogenaffinität wird gemäß der Bundeszentrale für gesundheitliche Aufklärung nicht nur der Konsum illegaler Rauschmittel wie z.B. Cannabis verstanden, sondern auch der Alkohol- und Tabakkonsum.

⁴ Die Repräsentativerhebung erfolgt in Zusammenarbeit mit der Gesellschaft für Sozialforschung und statistische Analysen mbH (forsa), welche eine Random-Stichprobe (computerbasierte Zufalls telefonnummern, Zufallsauswahl von Haushalten, Stichprobengröße von 3.000 Probanden) durchführt (vgl. BZgA (2001, S. 4)).

Bis Anfang der 90-er Jahre ist der Anteil rauchender Jugendlichen gemäß *Abbildung 1* zunächst zurückgegangen. Ab dem Jahr 1993 ist die Quote männlicher und weiblicher Raucher wieder angestiegen und erst ab dem Jahr 2001 erneut gesunken. Wobei bei den männlichen Jugendlichen die Quote von 27,2% (2001) auf 14,7% im Jahr 2008 und bei den weiblichen Rauchern von 27,9% (2001) auf 16,2% (2008) zurückgegangen ist. Der Anteil männlicher Raucher liegt damit geringfügig (ca. 1,5%) unter dem der weiblichen Raucher (vgl. BZgA (2008)).



*Abbildung 2: Weibliche und männliche Nichtraucher zwischen 12 und 17 Jahren
(in Anlehnung an die Ergebnisse der Repräsentativerhebung des BZgA (2008, S. 8))*

Neben den Informationen zur Entwicklung des Rauchverhaltens stellt die Repräsentativerhebung des BZgA auch Angaben zu den Nichtrauchern, d.h. derjenigen Jugendlichen welche noch nie in ihrem Leben geraucht haben, zur Verfügung. Entsprechend der *Abbildung 2* ist die Quote der Nichtraucher bis Ende der 90-er Jahre zuerst deutlich gestiegen und dann zurückgegangen. Komplementär dazu hat sich der Anteil der jugendlichen Raucherinnen und Raucher zu diesem Zeitpunkt entwickelt. Ab dem Jahr 2001 ist die Quote der Nichtraucher erneut angewachsen. Im Jahr 2008 lag der Anteil der männlichen Nichtraucher bei ca. 61,7% und der der weiblichen Jugendlichen bei ca. 59,4% (vgl. BZgA (2008, S. 9)).

Die Ergebnisse zur Entwicklung des Rauchverhaltens und zur Geschlechterverteilung können durch die Ergebnisse der „Europäischen Schülerstudie zu Alkohol und anderen Drogen (ESPAD)“ bestätigt werden.⁵

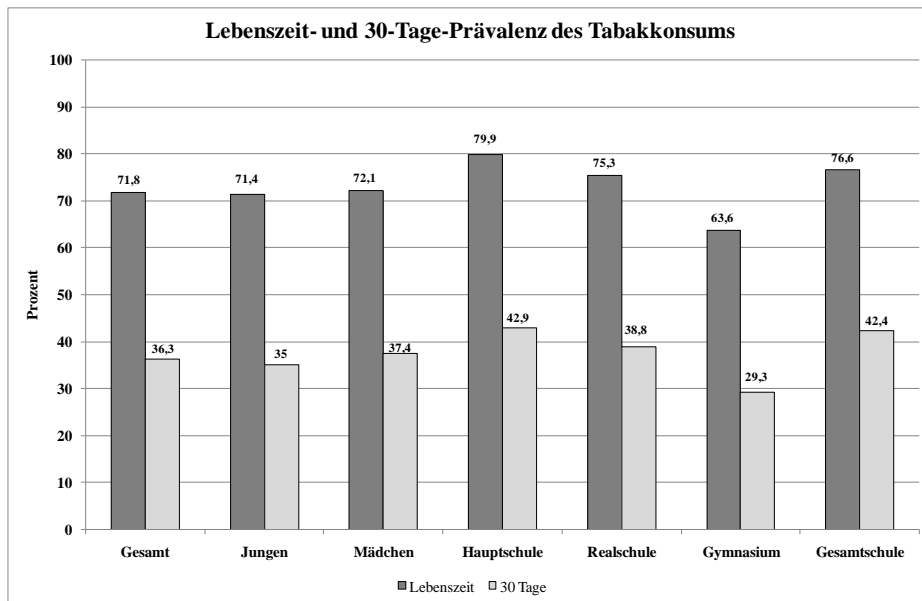


Abbildung 3: Lebenszeit- und 30 Tage-Prävalenz des Tabakkonsums
(in Anlehnung an die Ergebnisse der ESPAD-Studie (Kraus et al. (2008, S. 8))

Im Jahr 2007 berichteten 71,8% der befragten Schülerinnen und Schüler der neunten bzw. zehnten Klasse, jemals in ihrem Leben (Lebenszeitprävalenz) und 36,3% innerhalb der letzten 30 Tage vor der Erhebung (30 Tage-Prävalenz) geraucht zu haben. Im Vergleich zur Vorerhebung aus dem Jahr 2003 ist der Anteil der Raucher (hier bezogen auf die 30 Tage-Prävalenz) von 46,7% um ca. 10% zurückgegangen. Die *Abbildung 3* zeigt, dass der Tabakkonsum bei den Schülerinnen etwas stärker verbreitet ist, als bei den Schülern. Dieses Ergebnis gilt sowohl für die Lebenszeit- wie auch 30 Tage-Prävalenz. Hinsichtlich der Kategorisierung des Tabakkonsums nach den verschiedenen Schulformen ergeben sich diverse Unterschiede. An Gymnasien ist der Anteil an Jugendlichen, welche in ihrem Leben mindestens einmal geraucht haben, mit 63,6% am geringsten und an Hauptschulen mit 79,9% am höchsten. Ein vergleichbares Ergebnis ergibt sich auch für den Anteil der aktuellen Raucher (vgl. Kraus et al. (2008, S. 47 ff.)).

⁵ Die Bundesrepublik Deutschland nimmt seit dem Jahr 2003 an der europäischen Studie teil, in welcher Schülerinnen und Schüler der neunten und zehnten Klasse an Schulen verschiedener Bundesländer (u.a. Bayern) zum Tabak-, Alkohol- und Drogenkonsum befragt werden. Die ESPAD-Studie wurde im Jahr 2007 erneut durchgeführt. In die Auswertungen konnten Daten von 12.338 Schülerinnen und Schüler einbezogen werden (vgl. Kraus et al. (2008, S. 19 ff.)).

Zu den hier dargestellten Ergebnissen können in der Literatur eine Reihe weiterer nationaler (wie z.B. der Kinder- und Jugendgesundheitssurvey (KiGGS)) oder auch internationaler Studien (wie z.B. die Health Behavior in School-aged Children-Studie (HBSC)) gefunden werden, wobei z.B. im Rahmen der KiGGS-Studie aus dem Jahr 2004 vergleichbare Ergebnisse zum Rauchverhalten nach den oben genannten Schulformen, aber unterschiedliche Resultate für männliche und weibliche Raucher (51,3% der Jungen und 48,7% der Mädchen) beobachtet wurden (vgl. Lambert (2007)).

1.2 *Forschungshintergrund der ISAAC- und SOLAR-Studien*

Der Konsum von Tabak kann u.a. Erkrankungen der Lunge oder der Atemwege, wie z.B. chronisch obstruktive Bronchitis, Lungenemphysem (COPD), die Bildung von Lungenkarzinomen oder chronisch-entzündliche Erkrankungen wie Asthma bronchiale verursachen oder negativ beeinflussen (vgl. Andreas et al. (2009)). Dabei verantworten Raucher⁶ nicht nur ihr eigenes Gesundheitsrisiko, sondern fügen tabakrauchbelasteten Nichtraucher (so genannte Passivraucher) teilweise schwerwiegende gesundheitliche Schäden zu. Die Tabakrauchbelastung bei Passivrauchern kann eine Vielzahl von Erkrankungen wie Husten, Übelkeit, Kopfschmerzen, akuten und chronischen Herz-Kreislaufkrankungen, Erkrankungen der unteren Atemwege wie Lungenentzündungen oder Asthma sowie Krebserkrankungen hervorrufen. Demzufolge können Passivraucher – wenn auch in geringem Ausmaß und mit geringerer Häufigkeit – die gleichen Gesundheitsschäden erleiden wie Raucher (vgl. dkfz (2003, S. 5)).

Vor diesem Hintergrund befasst sich die vorliegende Arbeit eingehender mit dem Einfluss des aktiven und passiven Rauchens auf die Gesundheit, insbesondere auf die Lungenfunktion heranwachsender Jugendlicher. Die empirische Analyse erfolgt dabei auf Basis der Daten, welche im Rahmen der ISAAC II- und SOLAR II-Studie gewonnen wurden.

Der nachfolgenden Beschreibung ist vor auszuschicken, dass die ISAAC- wie auch die SOLAR-Studien hinsichtlich ihrer Forschungszielsetzungen wesentlich breiter angelegt waren, als auf die explizite Betrachtung der Tabakrauchexposition heranwachsender Jugendlicher. Zielsetzung beider Studien war im Wesentlichen die Untersuchung der welt-

⁶ Aus Vereinfachungsgründen und zugunsten einer besseren Lesbarkeit des nachfolgenden Textes werden die Begriffe „Raucher“, „Nichtraucher“, „Probanden“, „Schüler“, „Jugendliche“ sowie „Teilnehmer“ soweit nicht anders vermerkt für weibliche und männliche Personen synonym verwendet.

weiten Prävalenz von Symptomen allergischer und asthmatischer Erkrankungen bei Kindern und Jugendlichen.

1.2.1 ISAAC-Studien

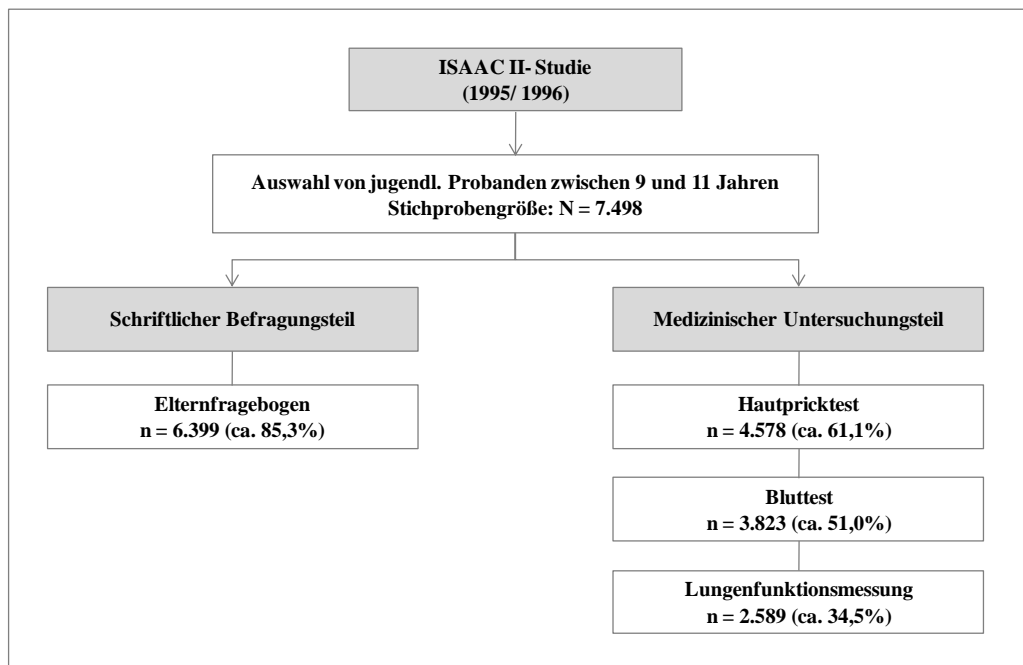
Die weltweit angelegte ISAAC-Studie (International Study on Asthma and Allergies in Childhood) wurde in den 90-er Jahren in zwei Phasen durchgeführt. In den Jahren 1994/1995 beteiligten sich im Rahmen der ersten Phase (**ISAAC I**) 56 Länder und 119 Studienzentren, darunter auch München und Dresden, an der Untersuchung der Prävalenz von Symptomen asthmatischer und allergischer Erkrankungen im Kindes- und Jugendalter. Mittels eines Elternfragebogens nahmen weltweit 250.000 Kinder im Alter zwischen sechs und sieben Jahren und 460.000 Jugendliche im Alter zwischen 13 und 14 Jahren an der Studie teil. Nach den damaligen Ergebnissen befand sich die BRD an dritthöchster Stelle der Asthmaprävalenz in Mittel- und Nordeuropa.

Die zweite Phase (**ISAAC II**), welche in den Jahren 1995/1996 durchgeführt wurde, diente der Gewinnung zusätzlicher objektiver Marker für Sensibilisierungen und allergische bzw. asthmatische Erkrankungen sowie zur Allergenexposition. Neben den bisherigen Studienorten München und Dresden, beteiligte sich an der zweiten Phase auch Leipzig an der Studie. Für die zweite deutsche Studienphase wurde erneut eine Stichprobe von Kindern und Jugendlichen im Alter zwischen fünf und sieben Jahren bzw. neun und 11 Jahren gezogen (vgl. Radon et al. (2005)). Die empirischen Auswertungen dieser Arbeit fokussieren sich lediglich auf die ISAAC II-Studie, resp. auf die Erhebungsdaten jugendlicher Probanden der Studienzentren München und Dresden. Alle nachfolgenden Beschreibungen, z.B. zur Probandenauswahl und zur Studiendurchführung beziehen sich auf diese getroffene Untersuchungseinschränkung.

Für die prospektive Kohortenstudie, welche vor Beginn der Pubertät durchgeführt werden sollte, wurden Schulen der Studienorte München und Dresden bzw. Schüler der vierten Klassenstufe zufällig ausgewählt.⁷ Insgesamt wurden 7.498 Jugendliche bzw. Eltern zur Teilnahme an der ISAAC II-Studie eingeladen. Dabei gliedert sich die zweite Phase der ISAAC-Studie in einen schriftlichen Befragungsteil und in diverse medizinische Untersu-

⁷ Nicht in die Studie einbezogen wurden Schulen mit körperlich oder geistig behinderten Kindern. Zudem wurden Schulen mit einem Ausländeranteil von über 80% von der Studie ausgeschlossen. Mit diesen Einschränkungen hinsichtlich der Studienpopulation sollten genetische oder Lebensstil-Faktoren als mögliche Einflussgrößen im Vorfeld minimiert werden.

chungen auf. Gemäß *Abbildung 4* nahmen 6.399 Eltern (ca. 85,3%) an der schriftlichen Befragung teil und beantworteten u.a. Fragen zum soziodemographischen Hintergrund (z.B. Geburtsdatum, Staatsangehörigkeit), Fragen zu früheren Erkrankungen, zur Gesundheit in der Familie oder auch erneut gestellte Fragen aus der ISAAC I-Studie (z.B. zur Prävalenz von Asthma, Atemwegkrankungen des Kindes). Die medizinischen Untersuchungen, wie der Hautprick- oder Bluttest sowie die Spirometrie und bronchiale Provokation wurden bei einem Teil der jugendlichen Probanden durchgeführt.



*Abbildung 4: Aufbau der ISAAC II-Studie
(in Anlehnung an Radon et al. (2005, S. 34))*

Die ISAAC-Studien waren ursprünglich auf eine reine Querschnittsbetrachtung ausgelegt. Allerdings lieferten die bisherigen Studien bereits wichtige Beiträge zur Beschreibung der weltweiten Prävalenz von Asthma und Allergien im Kindes- und Jugendalter. Unter longitudinalen Gesichtspunkten, d.h. unter Einbeziehung von Datenreihen zum Kindes- Jugend- und Erwachsenenalter bzw. zu zeitlichen Phasen der beruflichen Ausbildung bzw. Aufnahme der ersten Berufstätigkeit, sollten die bisherigen Studien in den nachfolgenden Jahren stetig ausgebaut werden. Damit sollten auf Basis der im weiteren Verlauf prospektiv ausgelegten Kohortenstudie valide Forschungsergebnisse über allergische Erkrankungen, bronchiale Hyperreaktivitäten und Atemwegsymptome vorliegen (vgl. Radon et al (2005)).

1.2.2 SOLAR-Studien

Die erste SOLAR-Studie (Studie in Ost- und Westdeutschland zu beruflichen Allergie Risiken) wurde in den Jahren 2002 und 2003 durchgeführt. Die longitudinale Fortführung der Studien sollte weltweit erstmalig die Möglichkeit eröffnen, Langzeitprognosen zu erlauben und wichtige Informationen über spezifische Risikofaktoren (z.B. zum Rauchverhalten) für das Auftreten von allergischen und asthmatischen Erkrankungen bereitstellen. Demzufolge war es die Zielsetzung der Follow-Up-Studie, den Einfluss von allergischen und asthmatischen Erkrankungen auf die Berufswahl bei Jugendlichen im Alter zwischen 16 und 18 Jahren sowie den Zusammenhang zwischen verschiedenen beruflichen Tätigkeiten und dem Auftreten derartiger Erkrankungen eingehend zu untersuchen.

Zur Bearbeitung der zentralen Fragestellungen der **SOLAR I-Studie** wurde als Untersuchungsinstrument wiederum ein Erhebungsbogen eingesetzt. Neben neuen Fragen wurden Befragungselemente aus der ISAAC II-Studie sowie dem ECRHS-Fragebogen (European Community Respiratory Health Survey) entnommen. Der Erhebungsbogen zur SOLAR I-Studie untergliedert sich u.a. in die Themenbereiche zur Demographie, Genetik, Atemwegsymptome und -erkrankungen, Wohnung, Haustierkontakt, Stressfaktoren sowie zum Rauchverhalten und Exposition gegenüber Passivrauch.

Die Studie wurde durch das Institut und die Poliklinik für Arbeits-, Sozial- und Umweltmedizin des Klinikums der Universität München in Kooperation mit dem Dr. von Haunerschen Kinderspital des Klinikums München, der Kinderklinik der Carl-Gustav-Carus Universität Dresden, dem Institut für Epidemiologie der Universität Ulm sowie der Justus-Liebig-Universität Gießen durchgeführt. Nachdem die Ethikkommission der Bayerischen Landesärztekammer sowie des Universitätsklinikums in Dresden keine Bedenken hinsichtlich der Umsetzung der SOLAR I-Studie geäußert haben, wurden die Studienunterlagen an den plausibilitätsgeprüften Kreis der Adressaten (N = 4.893) aus der ISAAC II-Studie versendet.⁸ Anzumerken ist in diesem Kontext, dass von den im Rahmen der ISAAC II-Studie teilnehmenden 6.399 Erziehungsberechtigten 5.438 (ca. 85%) im Vorfeld ihre Einverständniserklärung zur erneuten Kontaktaufnahme erteilt hatten (s.a. *Abbildung 5*).

⁸ Die Gültigkeit der Adressen aus der ISAAC II-Studie wurde im Vorfeld durch Abfragen bei den Einwohnermeldeämtern und Verwendung der örtlichen Telefonbücher überprüft. Für die SOLAR I-Studie konnten 4.893 Probanden aus der ISAAC II-Studie erneut angeschrieben werden (vgl. Radon et al. (2005, S. 32)).

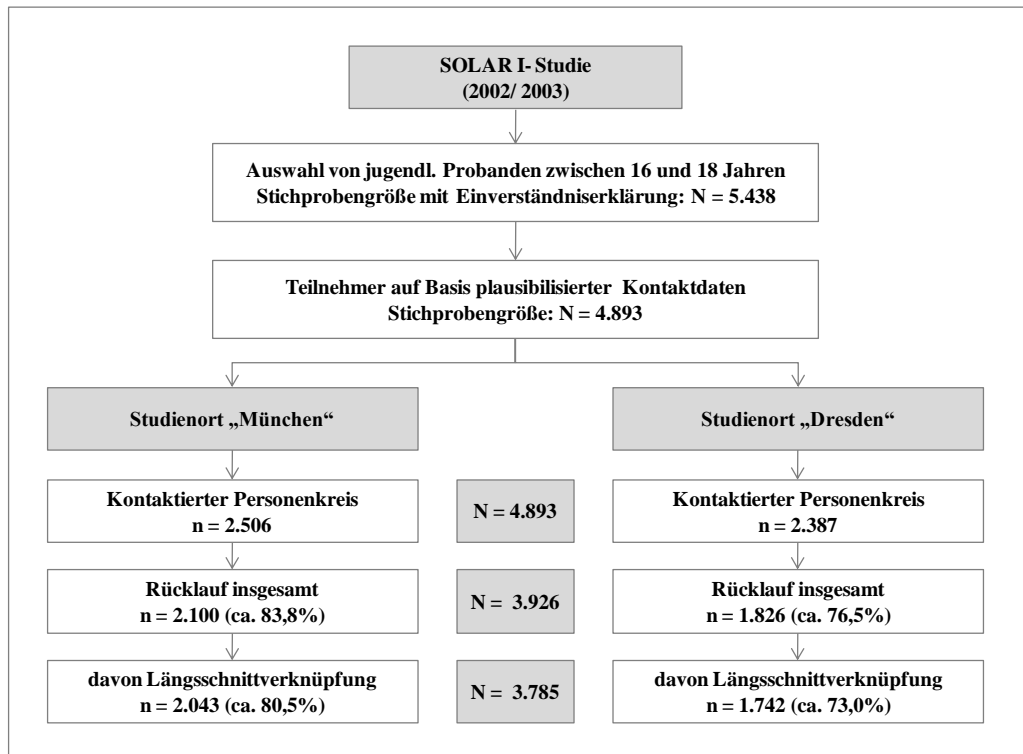


Abbildung 5: Aufbau der SOLAR I-Studie
(in Anlehnung an Radon et al. (2005, S. 34))

Nach mehreren Nachfassaktionen konnte für die beiden Studienorte München und Dresden letztlich eine Rücklaufquote von ca. 80,3% (N = 3.929) erzielt werden. Wobei diese Quote noch auf 77,4% zu reduzieren ist, weil die Einverständniserklärung zur Längsschnittverknüpfung der erhobenen Daten lediglich von 3.785 Probanden vorlag (vgl. Radon et al. (2005)).

Die **SOLAR II-Studie** setzt inhaltlich auf der SOLAR I-Studie auf und befasst sich mit der Zielsetzung ein Instrumentarium zur Optimierung der individuellen Beratung atopischer⁹ Jugendlicher bei ihrer Berufswahl zu schaffen. Unter effektivem Instrumenteneinsatz soll zukünftig vermieden werden, dass Jugendliche aufgrund ihrer Neigung allergische Erkrankungen zu entwickeln, unnötiger Weise von einer Vielzahl von Berufen ausgeschlossen werden. Demzufolge soll aufgrund einer optimierten individuellen Beratung ein konkreter Beitrag zur Senkung der Jugendarbeitslosigkeit geleistet werden. Zusätzlich soll das Instrumentarium zur Früherkennung von Kennzeichen allergischer Atemwegs- und Hauterkrankungen bereits während der Ausbildung bzw. den ersten Berufsjahren einge-

⁹ Unter „Atopie“ sind anlagenbedingte allergische Erkrankungen mit Überempfindlichkeit gegen Umweltstoffe zu verstehen. Als klassische Beispiele können endogenes Ekzem oder Asthma bronchiale genannt werden.

setzt werden, so dass rechtzeitig die erforderlichen Arbeitsschutzmaßnahmen getroffen werden können. Auf diese Weise soll der Schutz der Beschäftigten vor berufsbedingten Erkrankungen optimiert und die Erhaltung der Arbeitsfähigkeit der Betroffenen gesichert werden (vgl. Radon et al. (2008)).

Wie in der ISAAC II-Studie kamen im Rahmen der SOLAR II-Studie als Untersuchungsinstrumente eine schriftliche Befragung sowie eine Kombination verschiedener medizinischer Untersuchungen in Betracht. Für den Erhebungsbogen wurden erneut Fragen aus der ISAAC II-Studie, dem ECRHS-Fragebogen (European Community Respiratory Health Survey) und der SOLAR I-Studie ausgewählt. Der finale Erhebungsbogen bestand aus 136 Fragen zu diversen Themenbereichen, wie z.B. zur Wohn- und Arbeitssituation, zu sportlichen Aktivitäten, zur körperlichen Entwicklung, zu Belastungssituationen und zum Rauchverhalten. Für einen Teil der Jugendlichen sollten wie in der ISAAC II-Studie verschiedene medizinische Untersuchungen wie z.B. ein Hautprick- oder Bluttest sowie eine Spirometrie und eine bronchiale Provokation durchgeführt werden.

Die Studie wurde nach Zustimmung der Ethikkommissionen aus München und Dresden in der Zeit von Juli 2007 bis Juni 2009 durch das Institut und die Poliklinik für Arbeits-, Sozial- und Umweltmedizin des Klinikums der Universität München in Kooperation mit dem Dr. von Haunerschen Kinderspital des Klinikums München, der Kinderklinik der Carl-Gustav-Carus Universität Dresden, dem Institut für Epidemiologie der Universität Ulm sowie der Justus-Liebig-Universität Gießen durchgeführt.

Gemäß der *Abbildung 6* haben 3.054 Probanden aus der SOLAR I-Studie ihre Einwilligung für eine erneute Kontaktaufnahme erteilt. Von diesem Teilnehmerkreis konnten im Rahmen der SOLAR II-Studie 2.905 Probanden (ca. 95,1%) erneut kontaktiert werden. Anzumerken ist in diesem Kontext, dass sich die befragten Jugendlichen zum Zeitpunkt der Studiendurchführung im Alter zwischen 21 und 23 Jahren befanden, d.h. zu Beginn bzw. bereits während des Berufslebens. Im Vergleich zur SOLAR I-Studie umfasst die Stichprobe neben Berufseinsteigern mit Haupt- und Realschulabschluss nun auch die Abiturienten. An der schriftlichen Befragung beteiligten sich letztlich 2.047 Probanden. Von diesem Personenkreis nahmen rd. 57,1% ($n = 1.168$) an den nachfolgenden medizinischen Untersuchungen teil. Davon haben die Lungenfunktionsuntersuchungen 1.144 Jugendliche in Anspruch genommen (vgl. Radon et al. (2008)).

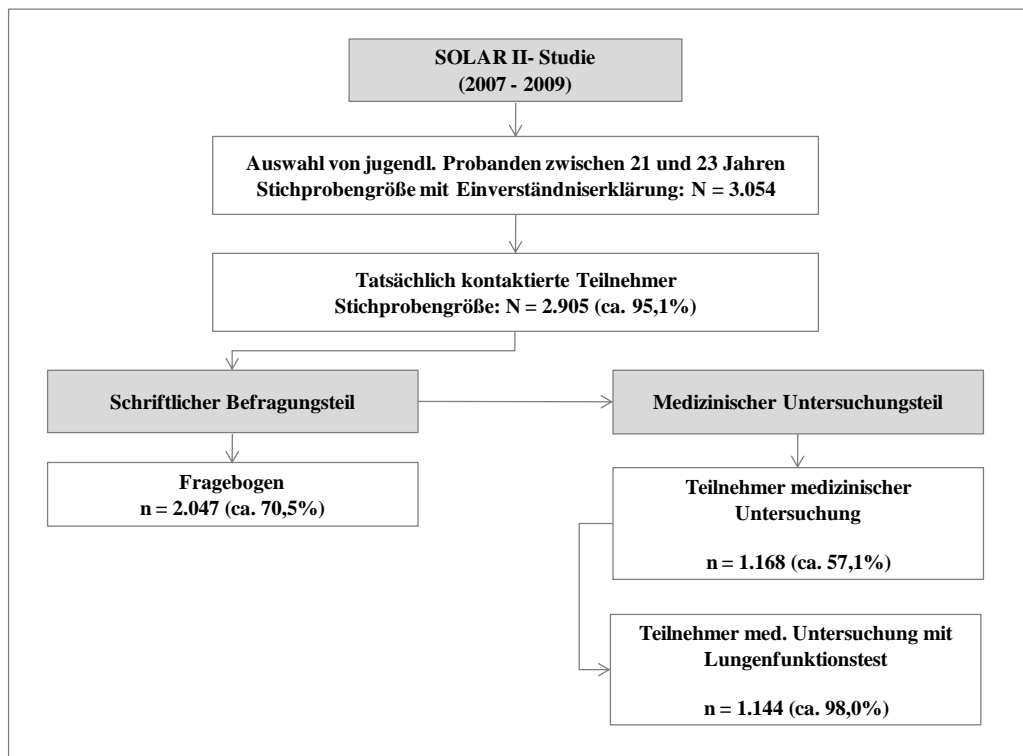


Abbildung 6: Aufbau der SOLAR II-Studie
(in Anlehnung an Radon et al. (2008))

Für die weitergehenden Auswertungen wurde lediglich ein Datenauszug aus der SOLAR II-Studie verwendet. Hinsichtlich der vorliegenden Themenstellung – Einfluss des aktiven und passiven Rauchverhaltens auf die Lungenfunktion – wurde der Datenauszug zudem auf den Probandenkreis mit Ergebnissen zur Lungenfunktionsuntersuchung ($n = 1.144$) beschränkt.

Die Zielsetzung des nachfolgenden Kapitels ist es, das verwendete Datenmaterial aus der ISAAC II- und SOLAR II-Studie detaillierter vorzustellen. Dabei sollen die Zielgrößen, potentiellen Einflussgrößen und die möglichen Confounder aus den jeweiligen Studien betrachtet sowie neu generierte Variablen kurz beschrieben werden. Des Weiteren sollen die Datenbereinigungsschritte bzw. die getroffenen Maßnahmen im Rahmen der Plausibilitätsprüfung veranschaulicht werden.

2. *Datenmanagement*

2.1 *Variablenselektion aus der ISAAC II- und SOLAR II-Studie*

Der Zuschnitt der Datengrundlage für die deskriptiven und multivariaten Analysen erfolgte im Wesentlichen auf Basis der SOLAR II-Studie. Bezogen auf die vorliegende Themenstellung umfasst der Erhebungsbogen der SOLAR II-Studie sieben Fragestellungen zur aktiven und passiven Tabakrauchexposition von Jugendlichen, von denen sechs in den Datenzuschnitt übernommen wurden.¹⁰ Zusätzlich wurden Fragen zur Demographie, zur Gesundheit und zur Arbeitssituation ausgewählt. Aus dem Erhebungsbogen zur ISAAC II-Studie wurden lediglich zwei Fragen aus dem Bereich zur Gesundheit der Familie in das zugeschnittene Datenmaterial integriert.¹¹ Darüber hinaus wurden vom Institut und der Poliklinik für Arbeits-, Sozial- und Umweltmedizin des Klinikums der Universität München ausgewählte Variablen aus den medizinischen Untersuchungen der ISAAC II- und SOLAR II-Studie zur Verfügung gestellt.

Nachfolgend erfolgt eine kurze inhaltliche Beschreibung der interessierenden Variablen aus den ISAAC II- und SOLAR II-Studien. Hinsichtlich der anschließenden Analysen werden die jeweiligen Variablen dabei in den Kategorien „Zielgrößen“, „Einflussgrößen“ sowie „Confounder“ vorgestellt.

2.1.1 *Zielgrößen*

Entsprechend der *Tabelle 1* sind als Zielgrößen für die späteren Auswertungen, insb. Regressionsanalysen, die stetigen Informationen zur Lungenfunktionstätigkeit, resp. zur Einse-

¹⁰ Die Fragebögen zur ISAAC II- und SOLAR II-Studie wurden auf dem beiliegenden Datenträger hinterlegt.

¹¹ Die Zusammenführung der Beobachtungen aus der ISAAC II- und SOLAR II-Studie wurde auf Basis der im Datenmaterial enthaltene Kohortennummer (Identifikationsvariable „knr“) durchgeführt.

kundenkapazität (FEV1) und zur forcierten Vitalkapazität (FVC) gemessen in Liter, von Interesse.

Die ausgewählten Datenwerte wurden im Rahmen einer spirometrischen Untersuchung gemessen. Dabei atmet der Proband zunächst langsam bis zur Totalkapazität (maximale inspiratorische Vitalkapazität) ein und anschließend bis zum Residualvolumen (maximale expiratorische Vitalkapazität) aus. Das maximale Volumen in Litern, welches mit maximaler Anstrengung nach maximaler Einatmung ausgeatmet werden kann, entspricht der forcierten Vitalkapazität („Forced Vital Capacity“). Die Einsekundenkapazität („Forced Expiratory Volume in 1 second“) entspricht dahingegen dem in einer Sekunde ausgeatmeten Volumen (vgl. Baur (2001)).

Variable	Bezeichnung	Typ
s2lufu07	FEV1 (Einsekundenkapazität gemessen in Liter)	stetig
s2lufu06	FVC (forcierte Vitalkapazität gemessen in Liter)	stetig

Tabelle 1: Zielgrößen (SOLAR II-Studie)

2.1.2 Potentielle Einflussgrößen

Als potentielle Einflussgrößen auf die Lungenfunktionstätigkeit wurden die Variablen zur aktiven und passiven Tabakrauchexposition betrachtet (s.a. *Tabelle 2 und 3*). Hinsichtlich der aktiven Rauchexposition wurden die Angaben zur Konsumkontinuität bezogen auf ein Jahr (dichotome Auswahlfrage) und zur Rauchintensität bzgl. des durchschnittlichen täglichen Zigarettenkonsums und zum Einstiegs- bzw. Ausstiegsalter aus dem zugrunde liegenden Fragenkatalog ausgewählt.

Variable	Bezeichnung	Ausprägung	Typ
s2f73	Haben sie schon einmal ein Jahr lang geraucht?	0 = nein	kategorial
		1 = ja	
s2f77	Wie viel Rauchen Sie durchschnittlich (Zigaretten/ Tag)?	-	stetig
s2f74	Wie alt waren sie als sie anfangen zu Rauchen (in Jahren)?	-	stetig
s2f76	Wie alt waren Sie, als Sie das Rauchen aufgegeben haben (in Jahren)?	-	stetig

Tabelle 2: Variablen zur Aktivrauchexposition (SOLAR II-Studie)

Zusätzlich zu den Angaben zum aktiven Rauchverhalten wurden in den Datenzuschnitt auch Informationen zur Passivrauchexposition einbezogen. Anzumerken ist in diesem

Kontext, dass die Fragen zur Passivrauchexposition gemäß Filterführung im Erhebungsbogen sowohl von den Rauchern wie auch Nichtrauchern beantwortet werden konnten.

Variable	Bezeichnung	Ausprägung	Typ
s2f78	Passivrauchexposition	0 = nein	kategorial
		1 = ja	
s2f79_1a	Passivrauchexposition zu Hause mehr als eine Stunde	-	stetig
s2f79_1b	Passivrauchexposition zu Hause weniger als eine Stunde	1 = ja	kategorial
s2f79_2a	Passivrauchexposition am Arbeitsplatz mehr als eine Stunde	-	stetig
s2f79_2b	Passivrauchexposition am Arbeitsplatz weniger als eine Stunde	1 = ja	kategorial
s2f79_3a	Passivrauchexposition in Bars, Restaurants,... mehr als eine Stunde	-	stetig
s2f79_3b	Passivrauchexposition zu Bars, Restaurants,... weniger als eine Stunde	1 = ja	kategorial
s2f79_4a	Passivrauchexposition an anderen Orten mehr als eine Stunde	-	stetig
s2f79_4b	Passivrauchexposition an anderen Orten weniger als eine Stunde	1 = ja	kategorial

Tabelle 3: Variablen zur Passivrauchexposition (SOLAR II-Studie)

Die Probanden wurden im Rahmen der Einstiegsfrage zu diesem Themenkomplex (s2f78) zunächst nach der regelmäßigen (d.h. mindestens einmal an den meisten Tagen oder Nächten) Passivrauchexposition befragt (dichotome Filterfrage). Mit Bejahung wurden weiterführende Fragen zum Ort (z.B. zu Hause, am Arbeitsplatz, in Bars oder Restaurants) und zur Intensität der passiven Tabakrauchbelastung (d.h. kategoriale Angabe bei Passivrauchbelastung von weniger als einer Stunde bzw. stetige Angabe bei Passivrauchexposition von mehr als einer Stunde) gestellt (weitere Details s.a. *Tabelle 3*).

2.1.3 Potentielle Confounder

Als mögliche weitere Einflussgrößen (Confounder) neben der aktiven und passiven Tabakrauchexposition auf die Lungenfunktion kommen die Variablen zur Körpergröße, Gewicht, Geschlecht, Erhebungszentrum, zum Bildungsstand, zur Asthma-Vordiagnose der Probanden sowie zum beruflichen Asthmarisiko in Betracht.¹²

Die Körpergröße in Meter und das Gewicht in Kilogramm wurden dabei während der medizinischen Untersuchung erfasst. Durch den Erhebungsbogen wurden die Informationen zum Bildungsstand bzw. zum Schulabschluss (sechs Antwortkategorien) abgefragt. Darüber hinaus umfasst der Fragebogen Informationen zur Asthmaerkrankung (s2f20_01x)

¹² Anzumerken ist in diesem Kontext, dass das Alter der Probanden nicht als potentielle Einflussgröße einbezogen wurde, da sich alle Befragten in etwa im gleichen Alter befanden (zwischen 21 bis 23 Jahren zum Zeitpunkt der SOLAR II-Studie).

bzw. zum beruflichen Asthmarisiko der Probanden (gemäß *Tabelle 4* untergliedert in NORISK_ges, LOWRISK_ges und HIGHRISK_ges).

Variable	Bezeichnung	Ausprägung	Typ
s2ou04	Größe in Meter	-	stetig
s2ou05	Gewicht in Kilogramm	-	stetig
sex	Geschlecht	0 = männlich	kategorial
		1 = weiblich	
s2centre	Erhebungszentrum	1 = München	kategorial
		2 = Dresden	
s2f80	Bildung	0 = Hauptschulabschluss	kategorial
		1 = Realschulabschluss	
		2 = Fachoberschulabschluss	
		3 = Abitur	
		4 = anderen Abschluss	
		5 = noch ohne Abschluss	
		6 = keinen Abschluss	
s2f20_01x	Vordiagnose - Asthma	0 = nie	kategorial
		1 = einmal	
		2 = mehrmals	
NORISK_ges	Kein berufliches Asthmarisiko	0 = nicht vorhanden	kategorial
		1 = vorhanden	
LOWRISK_ges	Niedriges berufliches Asthmarisiko	0 = nicht vorhanden	kategorial
		1 = vorhanden	
HIGHRISK_ges	Hohes berufliches Asthmarisiko	0 = nicht vorhanden	kategorial
		1 = vorhanden	

Tabelle 4: Confounder (SOLAR II-Studie)

Entsprechend den vorgegebenen Antwortkategorien sollten die Probanden bei der Frage zur Asthmaerkrankung angeben, ob diese durch einen Arzt nie, einmal oder mehrmals festgestellt wurde. Zur Einordnung der Fragen zum beruflichen Asthmarisiko mussten die Probanden im Vorfeld Angaben zur beruflichen Tätigkeit vornehmen. Um die für die Asthma- und Allergieentstehung relevanten beruflichen Expositionen einordnen zu können, wurde die so genannte Job-Exposure-Matrix (JEM) verwendet (s.a. *Abbildung 7*).

Hohes Asthmarisiko				Niedriges Asthmarisiko
Hochmolekulare Stoffe HMW	Niedrigmolekulare Stoffe LMW	Gemischte Stoffe MIXED	Irritative Spitzenexposition IRRPEAKS	Niedriges Asthmarisiko LOWRISK
Tierexposition	Reaktive Stoffexposition	Flüssigmetall-exposition	Irritative Spitzenexposition	Abgasexposition
Fischexposition	Isocyanatexposition	Textilienexposition		Passivrauch-exposition
Mehlexposition	Reinigungsmittel-exposition	Landwirtschafts-exposition		Irritanzen
Pflanzenexposition	Metallexposition			Geringe Antigenexposition
Milbenexposition	Holzstaubexposition			
Enzymexposition				
Latexexposition				
Bioaerosolexposition				
Pharmakaexposition				

Abbildung 7: Job-Exposure-Matrix
(in Anlehnung an Radon et al. (2005, S. 40))

In der JEM werden die spezifischen Expositionen betrachtet, die zu Berufsasthma führen können. Vier Kategorien der JEM bedeuten hohes berufliches und die fünfte steht für ein niedriges Asthmarisiko. Auf Basis der JEM wurden die Angaben zur beruflichen Tätigkeit in die drei Variablen kein, niedriges und hohes berufliches Asthmarisiko überführt.

Aus der ISAAC II-Studie wurden weitere Zusatzinformationen in das Datenmaterial eingefügt. Hier wurden aus der Erhebung 1995/ 1996 als mögliche Confounder sowohl die Lungenfunktionswerte „Einsekundenkapazität“ und die „forcierte Vitalkapazität“ sowie die Variablen zur Asthmadiagnose der Eltern betrachtet. Die Items zur Asthmadiagnose umfassen jeweils drei Kategorien mit den Antwortmöglichkeiten „ja“, „nein“ und „nicht bekannt“ (s.a. Tabelle 5).

Variable	Bezeichnung	Ausprägung	Typ
s0FEV1_B	FEV1(Einsekundenkapazität gemessen in Liter)	-	stetig
s0FVC_B	FVC (forcierte Vitalkapazität gemessen in Liter)	-	stetig
s0MUAS	Mutter Asthma	0 = ja	kategorial
		1 = nein	
		2 = nicht bekannt	
s0VAAS	Vater Asthma	0 = ja	kategorial
		1 = nein	
		2 = nicht bekannt	

Tabelle 5: Confounder (ISAAC II-Studie)

2.2 Datenbereinigung

Für die weitere Datenanalyse wurden die relevanten Variablen hinsichtlich ihrer Plausibilität geprüft. Diese Plausibilitätsprüfungen bezogen sich insb. auf die Fragen zum aktiven und passiven Rauchverhalten.

Der Proband hatte nach entsprechender Verneinung der dichotomen Einstiegsfilterfrage („Haben Sie schon einmal geraucht?“) die nachfolgenden Fragen zum aktiven Rauchen nicht mehr zu beantworten. Wurden trotzdem Angaben zum aktiven Rauchverhalten (Fragen 74 bis 77) geäußert, wurden diese auf den Wert „missing“ bzw. „not available“ (im Datensatz als „NA“ codiert) gesetzt.

Im Themenkomplex zum aktiven Rauchverhalten umfasst die Frage nach der durchschnittlichen Anzahl gerauchter Zigaretten pro Tag (Nummer 77) 19 Beobachtungen mit der Angabe „Null“. Diese 19 Teilnehmer hatten im Erhebungsbogen explizit angegeben, weniger als drei Zigaretten pro Woche zu rauchen. Um diese Angaben in die Analysen einbeziehen zu können, wurden die Angaben auf „Eins“ gesetzt.¹³

Auch der Themenkomplex zum Passivrauchen beginnt mit einer dichotomen Filterfrage (Frage 78 – „Sind Sie regelmäßig in den letzten 12 Monaten Tabakrauch anderer ausgesetzt gewesen?“). Wurde die Filterfrage vom Probanden verneint, war der Themenkomplex zur aktiven bzw. passiven Tabakrauchexposition beendet (Fortsetzung mit dem Themenkomplex zur Arbeitssituation (Frage 80 ff.)). Das Antwortverhalten zeigte, dass diese Filterführung von den Probanden entsprechend berücksichtigt wurde. Falls die Filterfrage

¹³ Diese und alle nachfolgenden inhaltlichen Datenbereinigungen erfolgten in Rücksprache mit den Projektpartnern.

bejaht wurde, folgt die Frage (Nummer 79) zum Ort und zur zeitlichen Intensität der Passivrauchexposition. Hinsichtlich der zeitlichen Intensität erfolgt im Erhebungsbogen die Unterscheidung zwischen weniger als eine Stunde (kategorial) und mehr als eine Stunde pro Tag (stetig). Hat ein Proband zu beiden Unterfragen eine Beantwortung vorgenommen, so wurde die stetige Antwort höher gewichtet als die kategoriale. D.h. die explizite Stundenangabe wurde beibehalten und die kategoriale Frage wurde im Datensatz auf „missing“ gesetzt. Des Weiteren wurden Angaben (gemeint sind dabei explizite Stundeneintragen bei einer der vier Unterfragen), die über 24 Stunden hinausgingen, ebenfalls als „not available“ eingestuft.

2.3 Berechnung und Umcodierung zusätzlicher Variablen

Neben den in Abschnitt 2.1 beschriebenen Variablen wurden für die nachfolgenden deskriptiven und multivariaten Analysen verschiedene Variablen neu berechnet bzw. bestehende Variablen geeignet umcodiert.

2.3.1 Zusätzliche Zielgrößen

Als weitere Zielgröße zur Einsekundenkapazität (FEV1) und zur forcierten Vitalkapazität (FCV) wurde entsprechend der *Tabelle 6* die relative Einsekundenkapazität (Tiffeneau-Wert)¹⁴ definiert.

Variable	Bezeichnung	Anmerkung	Typ
s2lufu07	FEV1 (Einsekundenkapazität gemessen in Liter)	[s.a. Tabelle 1]	stetig
s2lufu06	FVC (forcierte Vitalkapazität gemessen in Liter)	[s.a. Tabelle 1]	stetig
s2fev1.fvc	Relative Einsekundenkapazität (Tiffeneau – Wert)	[neu berechnet]	stetig

Tabelle 6: (Zusätzliche) Zielgrößen

Diese neue Variable wurde durch Division von FEV1 und FVC erzeugt.

$$\text{Tiffeneau – Wert} = \frac{FEV1}{FVC}$$

Mit Hilfe der drei ausgewählten Zielgrößen FEV1, FVC und deren Verhältnis zueinander ist eine Unterscheidung der obstruktiven (versperrende) und restriktiven (einschränkende)

¹⁴ Dieser Wert beträgt im Normalfall $\geq 75\%$ (vgl. Baur (2001)).

Ventilationsstörung sowie die Angabe ihres Schweregrades in den meisten Fällen möglich (vgl. Baur (2001)).

2.3.2 *Zusätzliche potentielle Einflussgrößen*

Die Dauer des aktiven Tabakkonsums wurde aus den Variablen zum Einstiegs- und Ausstiegsalters (s2f74 bzw. s2f76) berechnet. Bei Probanden, welche kein Ausstiegsalter (Annahme: Zum Zeitpunkt der Erhebungsdurchführung wurde immer noch geraucht!) angegeben haben, wurde das Erhebungsdatum bzw. das Datum des Fragebogeneinganges als Informationswert zur Bestimmung der Rauchdauer herangezogen. Für die Beschreibung der Zusammenhänge im deskriptiven Teil zwischen den Zielgrößen in Abhängigkeit von den verschiedenen Einflussgrößen sowie zur Berechnung der Regressionsmodelle wurde den Nichtrauchern im Datensatz zur Rauchdauer der Wert „Null“ zugewiesen. Auch die Variable zum durchschnittlichen täglichen Tabakkonsum (s2zig) wurde um den Datenwert „Null“ für die Nichtraucher erweitert (s.a. *Tabelle 7*).

Variable	Bezeichnung	Anmerkung	Ausprägung	Typ
s2dau	Dauer aktiven Rauchens (in Jahren) incl. Nichtraucher mit Null	[neu berechnet/ Datenänderung]	-	stetig
s2zig	Anzahl Zigaretten pro Tag incl. Nichtraucher mit Null	[Datenänderung]	-	stetig
s2packy	(Anzahl Zigaretten pro Tag / 20) * Dauer aktiven Rauchens	[neu berechnet]	-	stetig
s2f73	Haben sie schon einmal ein Jahr lang geraucht?	[s.a. Tabelle 2]	0 = nein	kategorial
			1 = ja	

Tabelle 7: (Zusatz-)Variablen zur Aktivrauchexposition

Als zusätzliche neue Variable wurde - basierend auf den Informationen zur Rauchdauer und zur durchschnittlichen täglichen Zigarettenanzahl – das so genannte Packungsjahr („packyears“) berechnet. Dabei bedeutet ein Packungsjahr, dass eine Person ein Jahr lang jeden Tag eine Packung Zigaretten (= 20 Zigaretten) geraucht hat.

$$\text{Packungsjahr} = \frac{\text{durchschnittlicher Zigarettenkonsum pro Tag}}{20} \times \text{Rauchdauer}$$

Datenmanagement

Variable	Bezeichnung	Anmerkung	Ausprägung	Typ
s2passrau	Passivrauchexposition in Stunden kumuliert	[neu berechnet]	-	stetig
s2f78	Passivrauchexposition	[s.a. Tabelle 3]	0 = nein 1 = ja	kategorial
s2zuH	Passivrauchexposition zu Hause	[umcodiert]	0 = nein 1 = weniger als eine Stunde 2 = mehr als eine Stunde	kategorial
s2amA	Passivrauchexposition am Arbeitsplatz	[umcodiert]	0 = nein 1 = weniger als eine Stunde 2 = mehr als eine Stunde	kategorial
s2inB	Passivrauchexposition in Bars, Restaurants,...	[umcodiert]	0 = nein 1 = weniger als eine Stunde 2 = mehr als eine Stunde	kategorial
s2andO	Passivrauchexposition an anderen Orten	[umcodiert]	0 = nein 1 = weniger als eine Stunde 2 = mehr als eine Stunde	kategorial

Tabelle 8: (Zusatz-)Variablen zur Passivrauchexposition

Die vier Unterfragen zum Ort und zur Dauer der Passivrauchexposition (Frage 79) wurden für eine weitere Verwendung in den Auswertungen entsprechend umgeformt. Dabei wurde die kategoriale Angabe „weniger als 1 Stunde“ mit dem Wert 0,5 belegt und die stetigen Stundenangaben unverändert übernommen. Darauf aufbauend wurden die Datenwerte zur Passivrauchexposition kumuliert und auf diese Art und Weise die Tagesexposition (Passivrauch insgesamt in Stunden) berechnet. Anzumerken ist in diesem Kontext, dass Summenwerte aus Plausibilitätsgründen den Wert „24“ nicht übersteigen dürfen. D.h. in solchen Fällen wurden die ermittelten Überschreitungswerte nachträglich bereinigt.

79. Bitte beschreiben Sie dies genauer: Wie viele Stunden sind Sie pro Tag Tabakrauch von anderen an den folgenden Orten ausgesetzt?

a) Zu Hause Stunden
Weniger als 1 Stunde ☐

b) Am Arbeitsplatz..... Stunden
Weniger als 1 Stunde ☒

c) In Bars, Restaurants, Kinos oder ähnlichem..... Stunden
Weniger als 1 Stunde ☐

d) Andere Orte Stunden
Weniger als 1 Stunde ☒

Abbildung 8: Beantwortungsbeispiel zur passiven Tagesrauchexposition

Auf Basis des nachfolgenden Beantwortungsbeispiels (s.a. *Abbildung 8*) wurde z.B. der Datenwert „8“ als Tagesexposition ermittelt.

Zusätzlich wurden die vier Unterfragen zum Ort und zur Dauer der Passivrauchexposition (Frage 79) wie folgt kategorisiert:

- 0 = keine Exposition (gemäß Frage 78 - Probanden ohne Passivrauchexposition)
- 1 = Exposition < 1 Stunde am Tag (gemäß Frage 79)
- 2 = Exposition \geq 1 Stunde am Tag (gemäß Frage 79)

Die zusätzlichen bzw. umcodierten Variablen zur Passivrauchexposition können der *Tabelle 8* entnommen werden.

2.3.3 *Zusätzliche potentielle Confounder*

Als weiterer möglicher Confounder zur Einsekundenkapazität (FEV1) und zur forcierten Vitalkapazität (FCV) aus der ISAAC II-Studie wurde, wie bei den Zielgrößen die relative Einsekundenkapazität (Tiffeneau-Wert), berechnet.

Daneben wurde die Variable „Bildung“ aus der SOLAR II-Studie dichotomisiert. Die Kategorien „Hauptschulabschluss“, „Realschulabschluss“, „anderer Abschluss“, „noch ohne Abschluss“ und „keinen Abschluss“ wurden zur Ausprägung „eher niedrig“ und die Kategorien „Abitur“ und „Fachhochschulabschluss“ zur Ausprägung „eher hoch“ zusammengefasst (Details s.a. *Tabelle 9*).

Zusätzlich wurden die drei Variablen zum beruflichen Asthmarisiko („NORISK_ges“, „LOWRISK_ges“ und „HIGHRISK_ges“) zu einer Variablen mit den Antwortkategorien „0 = nicht vorhanden“, „1 = gering“ und „2 = hoch“ vereint.

Datenmanagement

Variable	Bezeichnung	Anmerkung	Ausprägung	Typ
s0FEV1_B	FEV1 (Einsekundenkapazität gemessen in Liter)	[s.a. Tabelle 5]	-	stetig
s0FVC_B	FVC (forcierte Vitalkapazität gemessen in Liter)	[s.a. Tabelle 5]	-	stetig
s0fev1.fvc	relative Einsekundenkapazität (Tiffeneau – Wert)	[neu berechnet]	-	stetig
s2ou04	Größe in Meter	[s.a. Tabelle 4]	-	stetig
s2ou05	Gewicht in Kilogramm	[s.a. Tabelle 4]	-	stetig
sex	Geschlecht	[s.a. Tabelle 4]	0 = männlich	kategorial
			1 = weiblich	
s2centre	Erhebungszentrum	[s.a. Tabelle 4]	1 = München	kategorial
			2 = Dresden	
s2f80x	Bildung	[umcodiert]	0 = eher niedrig	kategorial
			1 = eher hoch	
s2f20_01x	Vordiagnose - Asthma	[s.a. Tabelle 4]	0 = nie	kategorial
			1 = einmal	
			2 = mehrmals	
s0MUAS	Mutter Asthma	[s.a. Tabelle 4]	1 = ja	kategorial
			2 = nein	
			3 = nicht bekannt	
s0VAAS	Vater Asthma	[s.a. Tabelle 4]	1 = ja	kategorial
			2 = nein	
			3 = nicht bekannt	
RISK	Berufliches Asthmarisiko	[neu berechnet]	0 = nicht vorhanden	kategorial
			1 = gering	
			2 = hoch	

Tabelle 9: (Zusätzliche) Confounder

Im nachfolgenden dritten Kapitel werden die hier vorgestellten Variablen nach verschiedenen Verfahren zunächst deskriptiv analysiert.

3. *Deskriptive Auswertungen*

Als Einstieg in die Datenanalyse wurden die ausgewählten Variablen aus der ISAAC II- und SOLAR II-Studie zunächst deskriptiv ausgewertet. Als Maße der zentralen Tendenz wurden hierfür Mittelwerte und Mediane berechnet. Als Streumaße kamen die Standardabweichung (SD) sowie Quantile und Spannweiten zur Anwendung. Die ordinalen und nominalen Daten wurden als absolute und relative Häufigkeiten dargestellt. Für die univariate graphische Darstellung wurden Histogramme und Boxplots eingesetzt.

Die Visualisierung des Zusammenhangs zwischen metrischen Zielgrößen und kategorialen erklärenden Variablen erfolgte durch Verwendung von Boxplots. Als Zusammenhangsmaß wurde hier die Korrelation nach Spearman berechnet bzw. im Fall von echt binären erklärenden Variablen (z.B. Geschlecht) kam der Bravais-Pearson-Korrelationskoeffizient zum Einsatz.

Die Zusammenhänge zwischen stetigen Zielgrößen und stetigen erklärenden Variablen wurden mit Streudiagrammen graphisch untersucht. Um einen möglichst genauen Eindruck von der Art des Zusammenhangs (linear oder nicht-linear) zu bekommen, wurde in den Streudiagrammen neben der „Kleinste-Quadrate-Gerade“ (K-Q-Gerade) der Lowess-Glätter („locally weighted scatterplot smoother“) abgebildet.¹⁵

¹⁵ Weitere Details zum Lowess-Glätter s.a. Fahrmeir et al. (2007, S. 339)).

3.1 Deskriptive univariate Analyse

3.1.1 Zielgrößen - Lungenfunktionsparameter

Die *Abbildung 9* und *Tabelle 10* geben einen Überblick über die Verteilung der Lungenfunktionsparameter (Zielgrößen) bei den Teilnehmern aus der SOLAR II-Studie. Die Werte zur Einsekundenkapazität (FEV1) schwanken zwischen 1,740 und 7,030 - wobei 50% im Bereich von 3,317 und 4,510 (Interquartilsabstand 1,193) Litern liegen. Die durchschnittliche Einsekundenkapazität liegt bei 3,926 bzw. der Median bei 3,780 Litern.

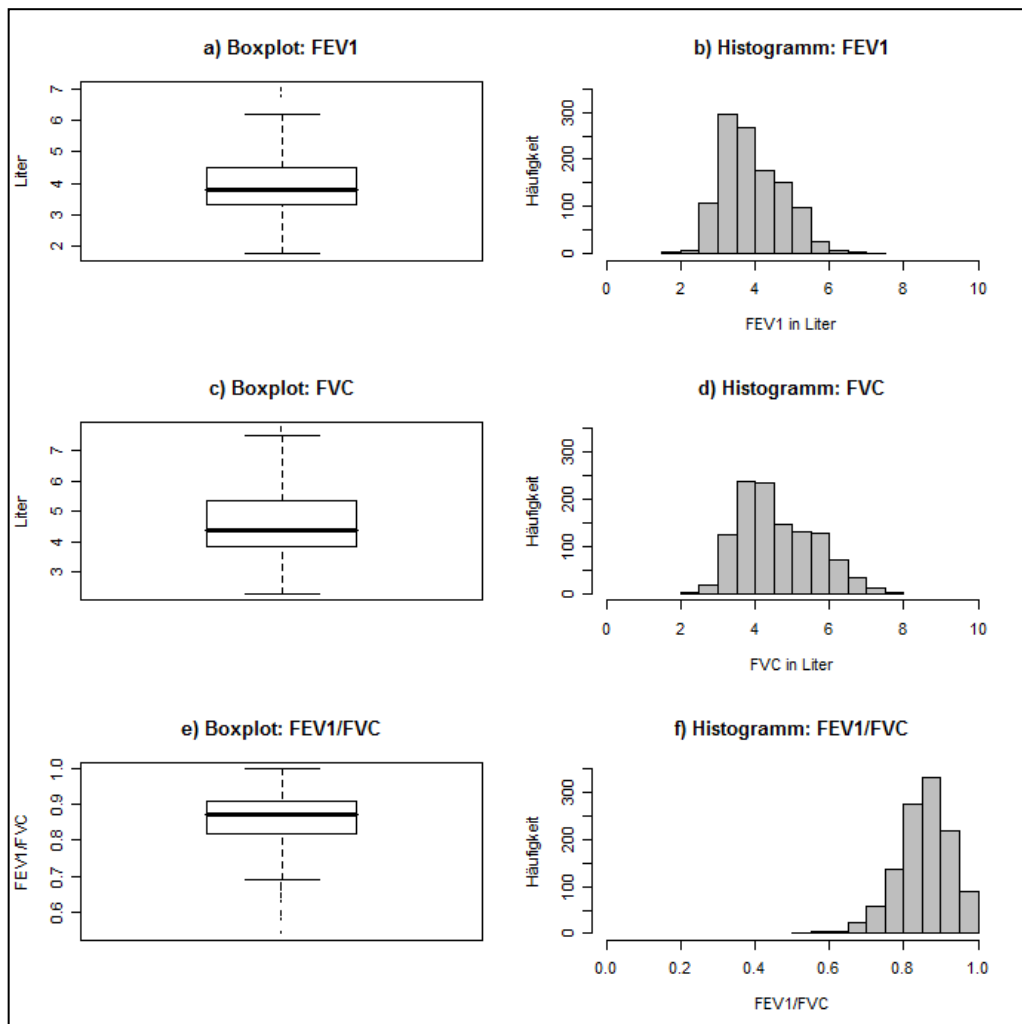


Abbildung 9: Verteilung der Zielgrößen (Lungenfunktionsparameter)

Boxplots und Histogramme für die metrischen Zielgrößen:

FEV1 ($n = 1.144$), FVC ($n = 1.141$) und FEV1/ FVC ($n = 1.141$)

Die Standardabweichung der FEV1-Variable beträgt 0,810. Das Histogramm sowie der Boxplot (s.a. *Abbildung 9* a) und b)) deuten auf eine linkssteile Verteilung hin.

Die forcierte Vitalkapazität (FVC) liegt zwischen 2,290 und 7,730 Litern und verfügt über einen Mittelwert von 4,596 bzw. einen Median von 4,400 Litern. Die Standardabweichung beträgt 1,010. Der Interquartilsabstand (IQR) wurde mit 1,520 Litern berechnet. Auch der FVC ist eher linkssteil verteilt (s.a. *Abbildung 9 c)* und d)).

Der aus den beiden Variablen gebildete Tiffeneau-Wert schwankt zwischen 0,540 und 1. Der mittlere Wert beträgt 0,860 bzw. der Median 0,870. Für die Standardabweichung ist ein Wert von 0,070 berechnet worden. Von den ermittelten Tiffeneau-Daten liegen 50% der Werte zwischen 0,820 und 0,910 (IQR 0,090). Das Histogramm und der Boxplot (s.a. *Abbildung 9 e)* und f)) weisen auf eine deutlich rechtssteile Verteilung hin.

Variable	n	Minimum	25% Quantil	Median	Mean	75% Quantil	Maximum	SD	Missing Values
FEV1	1.144	1,740	3,317	3,780	3,925	4,510	7,030	0,810	-
FVC	1.141	2,290	3,820	4,400	4,596	5,340	7,730	1,010	3
FEV1/ FVC	1.141	0,540	0,820	0,870	0,860	0,910	1,000	0,070	3

Tabelle 10: Verteilung der Zielgrößen (Lungenfunktionsparameter)

3.1.2 Potentielle Einflussgrößen - Aktives Rauchverhalten

Von den einbezogenen Studienteilnehmern haben 465 angegeben, schon einmal aktiv geraucht zu haben bzw. 459 Probanden machten Angaben zur Rauchdauer. Im Durchschnitt rauchten die Teilnehmer ca. sechs Jahre lang. Die Standardabweichung beträgt 2,370. Die Angaben schwankten dabei zwischen einem und 16 Jahren. Dabei rauchten 37,2% bis zu fünf Jahre, 35,5% zwischen sechs und sieben Jahren und 26,6% mehr als sieben Jahre. Die *Abbildung 10 a)* zeigt die Verteilung der Variablen „Dauer des aktiven Rauchens“. Dabei fällt ein extremer Wert von 16 Jahren besonders auf.

Zur Anzahl der Zigaretten pro Tag machten 454 Probanden Angaben im Erhebungsbogen (s.a. *Tabelle 11*). Bis zu fünf Zigaretten pro Tag rauchten 41,2% der befragten Teilnehmer, zwischen sechs und 10 Zigaretten konsumierten ca. 28,8% und mehr als 10 Zigaretten ca. 30% der Probanden. Das Minimum lag dabei bei einer Zigarette und das Maximum bei 30 Zigaretten pro Tag. Der Median lag bei 8 Zigaretten (Mittelwert 8,683 und SD 6,30). Auffallend ist ein relativ breiter Interquartilsabstand (10) und ein extremer Wert von 30 Zigaretten pro Tag.

Die Betrachtung der Packungsjahre ergab, dass 32,7% der 449 Beobachtungen unter einem und 67,3% über einem Packungsjahr lagen. Der Median betrug zwei Packungsjahre (Mit-

telwert 2,825 und SD 2,600). Dabei schwankten die Werte zwischen 0,050 und 15 Packungsjahre.

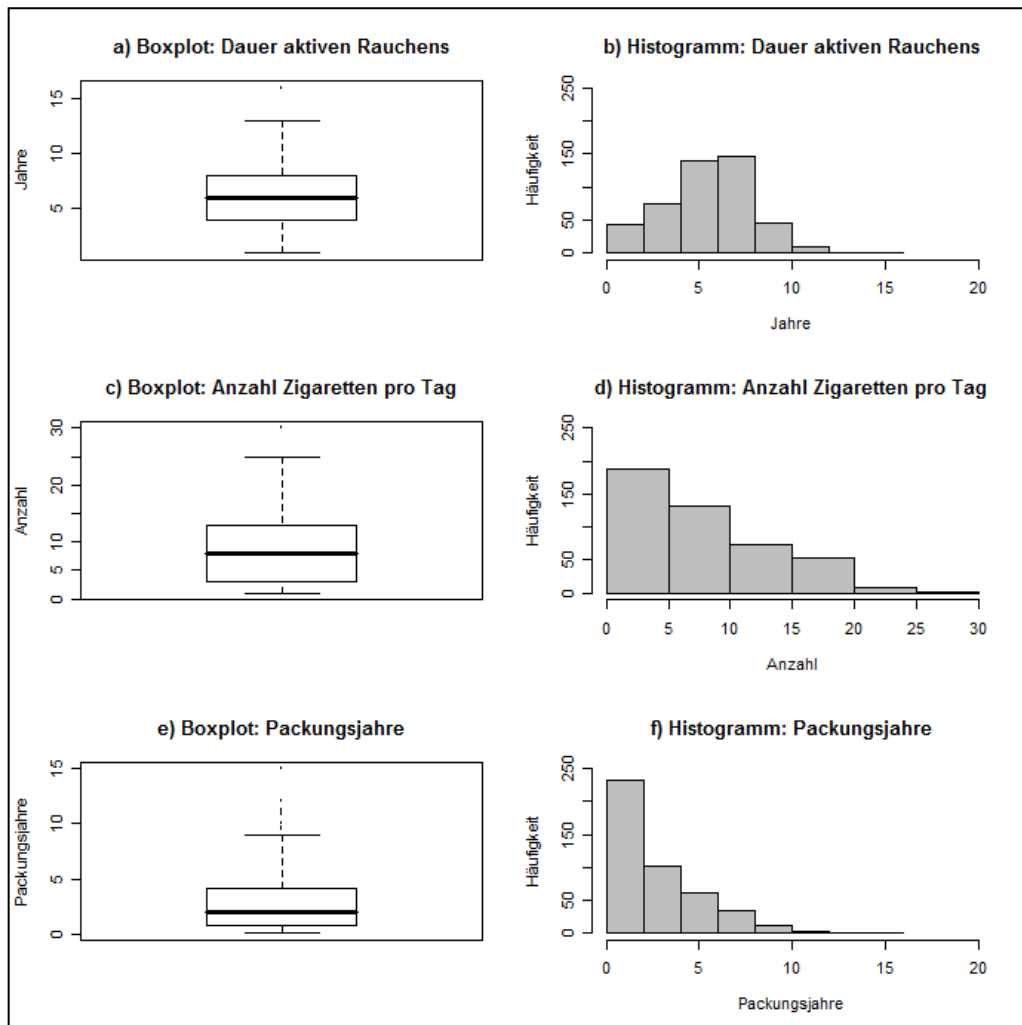


Abbildung 10: Verteilung der potentiellen Einflussgrößen (aktives Rauchverhalten)

Boxplots und Histogramm für die metrischen Einflussgrößen:

Dauer des aktives Rauchen ($n = 459$), Anzahl Zigaretten pro Tag ($n = 454$) und Packungsjahre ($n = 449$)

Variable	n	Minimum	25% Quantil	Median	Mean	75% Quantil	Maximum	SD	Missing Values
Dauer des aktiven Rauchens	459	1,000	4,000	6,000	6,024	8,000	16,000	2,370	6
Anzahl Zigaretten pro Tag	454	1,000	3,000	8,000	8,683	13,000	30,000	6,300	11
Packungsjahre	449	0,050	0,750	2,000	2,825	4,200	15,000	2,600	16

Tabelle 11: Verteilung der potentiellen Einflussgrößen (aktives Rauchverhalten)

3.1.3 Potentielle Einflussgrößen - Passives Rauchverhalten

Zu den Fragen zur Passivrauchexposition machten 585 der Teilnehmer entsprechende Angaben. Von den 395 Probanden, die zu Hause passiv rauchten, waren etwa 70% weniger als eine Stunde und ca. 30% mehr als eine Stunde Tabakrauch anderer ausgesetzt. Bei den 435 Teilnehmern, die am Arbeitsplatz und den 450 Teilnehmern, die an anderen Orten exponiert waren, war das Verhältnis ähnlich. Entsprechend der *Tabelle 12* gaben 524 Probanden eine Exposition in Bars etc. an. Hier war das Verhältnis ausgeglichener. Nahezu 54% waren weniger als eine Stunde Tabakrauch anderer ausgesetzt und ca. 46% mehr als eine Stunde am Tag.

Variable	Ausprägung	Häufigkeit (%)	Missing Values
Passivrauchexposition (n = 1.137)	nein	552 (48,6)	7
	ja	585 (51,5)	
zu Hause (n = 395)	< 1 Stunde pro Tag	274 (69,4)	190
	≥ 1 Stunde pro Tag	121 (30,6)	
am Arbeitsplatz (n = 435)	< 1 Stunde pro Tag	289 (66,4)	150
	≥ 1 Stunde pro Tag	146 (33,6)	
in Bars, Restaurants und ähnlichem (n = 524)	< 1 Stunde pro Tag	282 (53,8)	61
	≥ 1 Stunde pro Tag	242 (46,2)	
an anderen Orten (n = 450)	< 1 Stunde pro Tag	313 (69,6)	135
	≥ 1 Stunde pro Tag	137 (30,4)	

Tabelle 12: Verteilung der potentiellen Einflussgrößen (passives Rauchverhalten)

3.1.4 Potentielle Confounder

Die *Tabelle 13* gibt einen Überblick über die Verteilung der demographischen Merkmale. Die weibliche Anzahl an Teilnehmern überwog mit 59,6%. Auffällig ist der Anteil an Teilnehmern mit hohem Bildungsstand, von denen 70% angegeben haben, einen Fachoberschulabschluss bzw. ein Abitur zu besitzen. Ausgeglichen war dagegen das Verhältnis zwischen Probanden aus München und Dresden. Die Körpergröße und das Gewicht der Teilnehmer wurden in den medizinischen Untersuchungen erfasst. Die Werte sind in *Tabelle 14* zusammengefasst.

Von 565 Teilnehmern lagen ebenfalls die Lungenfunktionswerte aus der ISAAC II-Studie vor. Als die Vorgänger-Studie durchgeführt wurde, waren die Probanden zwischen sieben und neun Jahre alt. Der Median für die Einsekundenkapazität lag bei 2,090 Litern und für die forcierte Vitalkapazität bei 2,340 Litern. Die minimalen und maximalen Werte schwankten zwischen 1,320 und 3,220 Litern bzw. 3,500 Litern.

Variable	Ausprägung	Häufigkeit (%)	Missing Values
Geschlecht (n = 1.144)	männlich	462 (40,4)	0
	weiblich	682 (59,6)	
Bildungsstand (n = 1.142)	niedrig	343 (30,0)	2
	hoch	799 (70,0)	
Erhebungszentrum (n = 1.144)	München	559 (48,9)	0
	Dresden	585 (51,1)	

Tabelle 13: Verteilung der potentiellen Confounder (demographische Variablen)

Variable	n	Minimum	25% Quantil	Median	Mean	75% Quantil	Maximum	SD	Missing Values
Körpergröße	1.144	1,520	1,650	1,715	1,722	1,789	2,034	0,095	0
Gewicht	1.144	40,200	58,950	66,600	68,470	75,820	141,200	13,348	0
FEV1 (ISAAC II)	565	1,320	1,880	2,090	2,103	2,310	3,220	0,308	579
FVC (ISAAC II)	565	1,320	2,090	2,340	2,335	2,570	3,500	0,351	579
FEV1/ FVC (ISAAC II)	565	0,720	0,860	0,910	0,903	0,940	1,000	0,006	579

Tabelle 14: Verteilung der potentiellen Confounder (metrische Variablen)

In Tabelle 15 sind die für die vorliegenden Analysen berücksichtigten Prävalenzen von Asthma in den einzelnen Kategorien dargestellt. Hier wurden sowohl die Vordiagnosen der Probanden, die Asthmadiagnosen der Eltern sowie das Risiko Berufsasthma zu bekommen betrachtet. Dabei zeigt sich, dass bei etwa 90,4% der Probanden, die eine Angabe über ihre Asthmadiagnose machten, niemals Asthma festgestellt wurde. Eine einmalige Asthmadiagnose hatten 5,5% der Probanden erhalten und bei etwa 4,1% wurde mehrmals Asthma diagnostiziert.

Wie unter Punkt 2.1.3 erläutert, wurde die Job-Exposure Matrix eingesetzt, um das berufliche Asthmarisiko der Probanden quantifizieren zu können. Insgesamt sind fast 58% der Teilnehmer keinem beruflichen Asthmarisiko ausgesetzt. Dagegen unterlagen ca. 16%

einem geringen beruflichen Risiko und 26% sind einem erhöhten Asthmarisiko am Arbeitsplatz ausgesetzt. Die Familienanamnese ergab, dass ca. 5% der Mütter und Väter jemals an Asthma erkrankt waren.

Variable	Ausprägung	Häufigkeit (%)	Missing Values
Vordiagnose Asthma (n = 1.125)	nie	1.017 (90,4)	19
	einmal	62 (5,5)	
	mehrmals	46 (4,1)	
Mutter Asthma (n = 1.116)	ja	50 (4,5)	28
	nein	1.044 (93,6)	
	nicht bekannt	22 (2,0)	
Vater Asthma (n = 1.116)	ja	56 (5,0)	28
	nein	1.008 (90,3)	
	nicht bekannt	52 (4,7)	
Berufliches Asthmarisiko (n = 1.143)	kein Risiko	659 (57,7)	1
	geringes Risiko	187 (16,4)	
	hohes Risiko	297 (26,0)	

Tabelle 15: Verteilung der potentiellen Confounder (kategoriale Variablen)

3.2 Zusammenhangsanalyse

3.2.1 Einsekundenkapazität und potentielle Einflussgrößen/ Confounder

Aktives Rauchverhalten

Für den Lungenfunktionsparameter „FEV1“ sind Streudiagramme mit den metrischen Variablen „Dauer des aktiven Rauchens“, „Zigaretten pro Tag“ und „Packungsjahre“ in *Abbildung 11* dargestellt.

Ein Zusammenhang zwischen der Einsekundenkapazität und dem aktiven Rauchverhalten ist nicht zu erkennen. Die durchschnittlichen FEV1-Werte sowie die Streubreite scheinen annähernd gleich zu bleiben. So schwanken die FEV1-Werte bzgl. der Dauer des aktiven Rauchens bei den Nichtraucher (Null Jahre Dauer aktiven Rauchens) zwischen 1,74 und 6,96 Litern und bei den Rauchern, die bis zu fünf Jahre Tabak konsumierten, zwischen 2,41 und 7,03 Litern. Darüber hinaus, also bei einer aktiven Rauchdauer von über fünf Jah-

ren, liegen die FEV1-Werte bei 1,75 Litern im Minimum und 6,02 Litern im Maximum. Die im Streudiagramm dargestellte Kleinste-Quadrate-(KQ)-Gerade verläuft flach. Der Streudiagramm-Glätter nimmt ab einer Rauchdauer von ca. fünf Jahren leicht linear ab. Dieser wird evtl. von extremen Werten im Bereich über 10 Jahren beeinflusst.

Die Verteilung der FEV1-Werte in Abhängigkeit von der durchschnittlichen Zigarettenanzahl pro Tag ist in *Abbildung 11 b)* dargestellt.

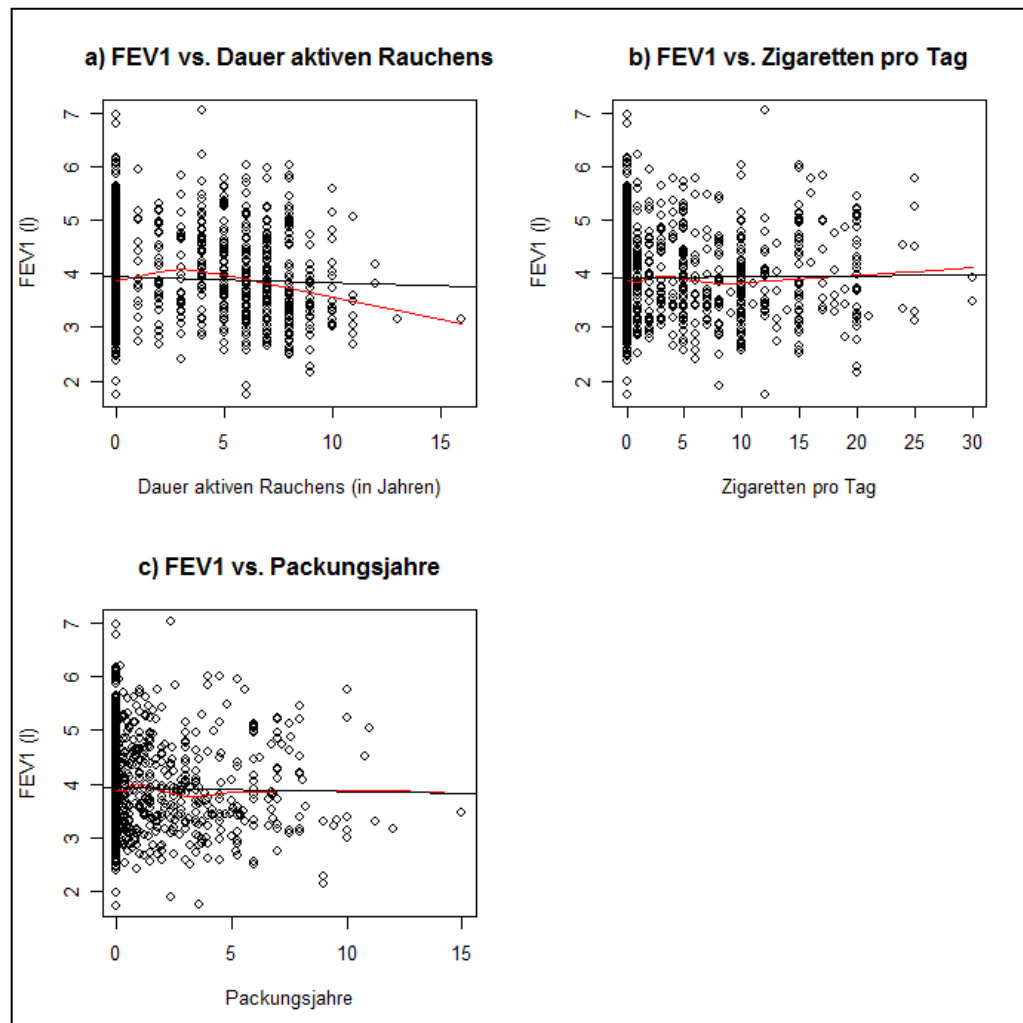


Abbildung 11: Aktives Rauchverhalten

Streudiagramme zwischen Einsekundenkapazität (FEV1) und der Dauer des aktiven Rauchens ($n = 1.132$), der Anzahl Zigaretten pro Tag ($n = 1.127$) und den Packungsjahren ($n = 1.122$) mit Regressionsgerade (schwarz) und Streudiagramm-Glätter („lowess - locally weighted scatterplot smoothing“)

Die Nichtraucher weisen FEV1-Werte von 1,74 Litern bis 6,96 Litern auf, wohingegen bei Teilnehmern, die z.B. mehr als 10 Zigaretten pro Tag rauchten, die Einsekundenkapazität

zwischen 1,75 und 7,03 Litern gemessen wurde. Die durchschnittlichen Lungenfunktionswerte (dargestellt durch die KQ-Gerade und der Lowess-Kurve) bleiben konstant.

Der Vergleich der Abhängigkeit der Einsekundenkapazität von der Anzahl der Packungsjahre zeigt ein ähnliches Muster (s.a. *Abbildung 11 c*). Während Nichtraucher FEV1-Werte zwischen 1,74 und 6,96 Litern aufwiesen, lagen Minimum und Maximum bei den Rauchern mit mehr als einem Packungsjahr bei 1,75 und 7,03 Litern. Auch die eingezeichnete KQ-Gerade und der Streudiagramm-Glätter deuten darauf hin, dass zudem kein Effekt von der Anzahl der Packungsjahre ausgeht.

Passivrauchexposition

Ein Zusammenhang zwischen der Passivrauchexposition und der Einsekundenkapazität ist ebenfalls nicht erkennbar. Wird die Passivrauchexposition zu Hause betrachtet, so beträgt der Median bei den nichtexponierten Teilnehmern 3,78 Liter. Die FEV1-Werte schwanken zwischen 1,74 und 6,78 Litern.

Bei Teilnehmern, die zu Hause weniger als eine Stunde Tabakrauch anderer ausgesetzt waren, liegt der Median bei 3,865 Litern. Das Minimum bzw. das Maximum bewegen sich zwischen 2,41 und 6,96 Litern. Bei längerer Exposition ergibt sich ein Median von 3,62 Litern mit Werten zwischen 1,90 und 6,02 Litern. Die Korrelation nach Spearman ist mit -0,02 verschwindend gering.

Bei der Passivrauchexposition am Arbeitsplatz bleiben die Mediane je nach Intensität annähernd gleich. Bei keiner Exposition betrug der Median 3,78 Liter, bei weniger als eine Stunde 3,82 Liter und bei mehr als eine Stunde 3,82 Liter. Auch die Streubreite bleibt annähernd gleich. Die Korrelation nach Spearman, die mit 0,02 berechnet wurde, untermauert den Eindruck der Unabhängigkeit von Einsekundenkapazität und Passivrauchexposition am Arbeitsplatz.

Bei der Passivrauchexposition in Bars etc. steigt der Median (3,86 Liter) gegenüber den Nichtexponierten (3,78 Liter) zunächst leicht an. Bei längerer Expositionsdauer geht der Median (3,70) wieder geringfügig zurück. Dieses Muster ist zudem bei einer Exposition an anderen Orten (s.a. *Abbildung 12 d*) zu erkennen, wo der mediane FEV1-Wert von 3,78 (keine Exposition) auf 3,82 Liter (weniger als eine Stunde) steigt und dann wieder auf 3,73 Liter (mehr als eine Stunde) zurückgeht. Allerdings sind die Differenzen sehr gering. Die Korrelationskoeffizienten nach Spearman von jeweils -0,01 deuten ebenfalls darauf hin,

dass zwischen der Einsekundenkapazität und den Variablen zur Passivrauchexposition in Bars und an anderen Orten kein Zusammenhang besteht.

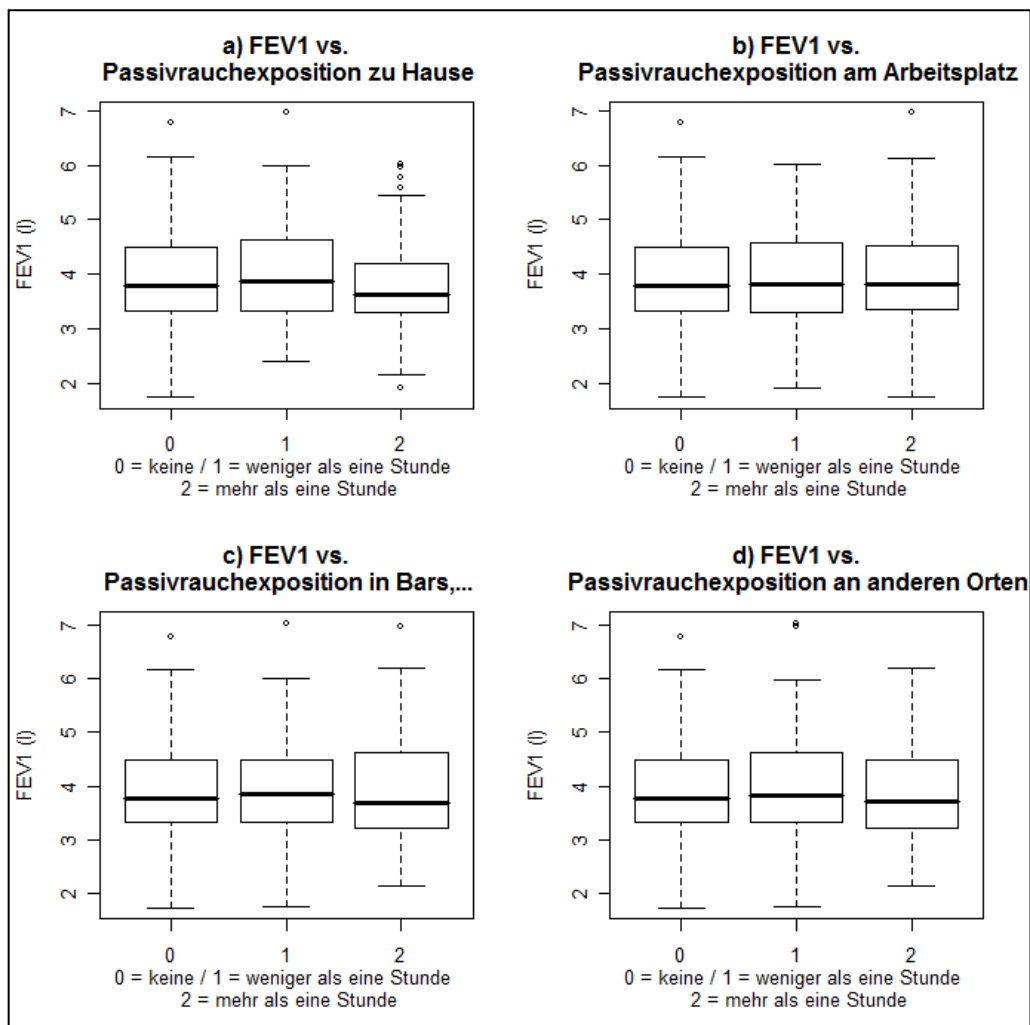


Abbildung 12: Passivrauchexposition

Verteilung der Einsekundenkapazität (FEV1) in Abhängigkeit

von der Passivrauchexposition zu Hause ($n = 947$),

am Arbeitsplatz ($n = 987$), in Bars etc. ($n = 1.076$) und an anderen Orten ($n = 1.002$)

Die FEV1-Werte der Studienteilnehmer in Abhängigkeit der kumulierten Passivrauchexposition scheinen dagegen gemäß *Abbildung 13* mit zunehmender Expositionsdauer leicht abzunehmen. Die KQ-Gerade und Lowess-Glätter weisen auf einen schwach negativen Zusammenhang hin. Allerdings kann eine Beeinflussung durch extreme Werte im Bereich von über 20 Stunden nicht ausgeschlossen werden.

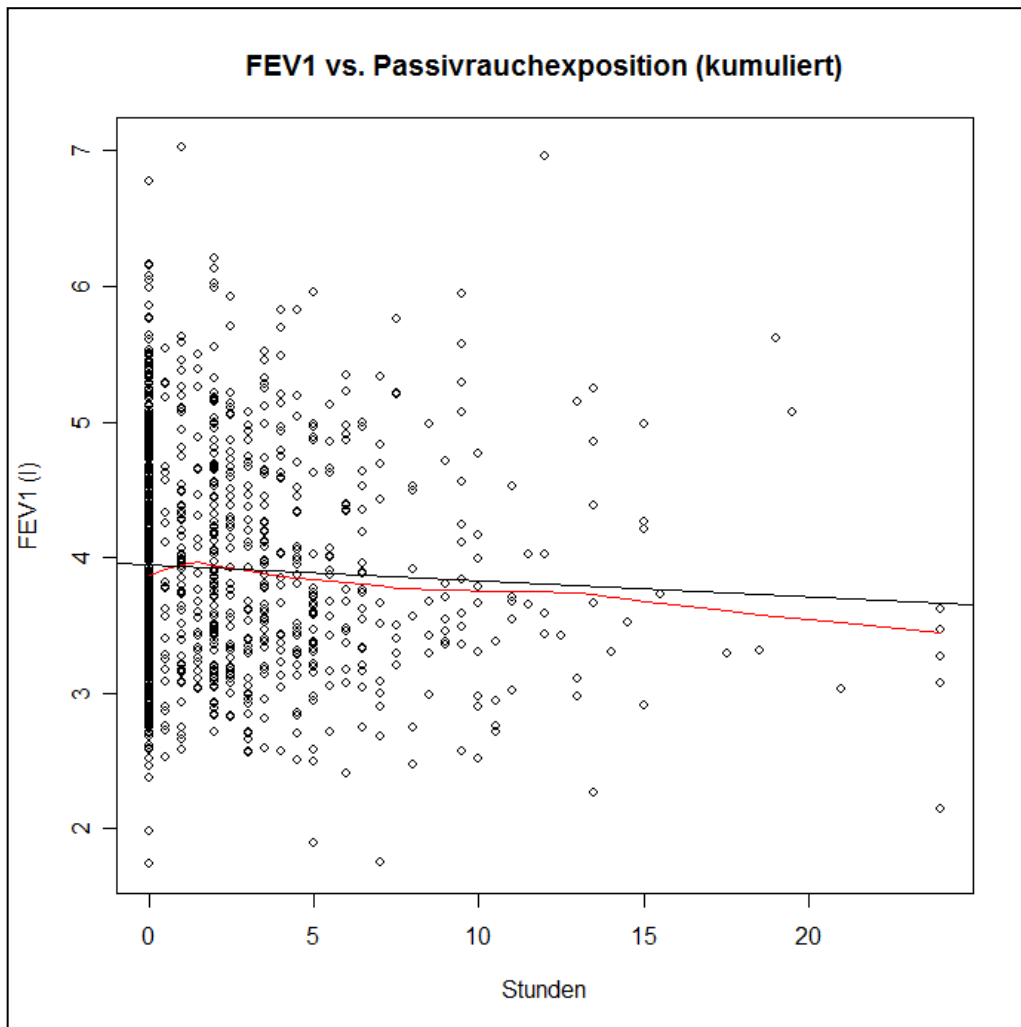


Abbildung 13: Passivrauchexposition (kumuliert)
 Streudiagramme zwischen Einsekundenkapazität (FEV1) und
 der kumulierten Passivrauchexposition ($n = 1.134$)
 mit Regressionsgerade (schwarz) und Streudiagramm-Glätter
 („lowess - locally weighted scatterplot smoothing“)

Geschlecht, Bildung und Erhebungszentrum

Ein geschlechtsspezifischer Unterschied bzgl. der Einsekundenkapazität ist deutlich zu erkennen (Abbildung 14).

Die männlichen Probanden weisen im Mittel höhere FEV1-Werte auf. Der Median bei den Männern liegt bei 4,645 Litern, während die weiblichen Teilnehmer einen Median von 3,4 Litern aufweisen. Bei den männlichen Jugendlichen schwanken die Werte zwischen 2,71 und 7,03 Litern und bei den weiblichen Probanden zwischen 1,74 und 5,83 Litern. Die Korrelation nach Pearson deutet zudem auf einen starken Zusammenhang hin ($r_{\text{pearson}} = -0,743$).

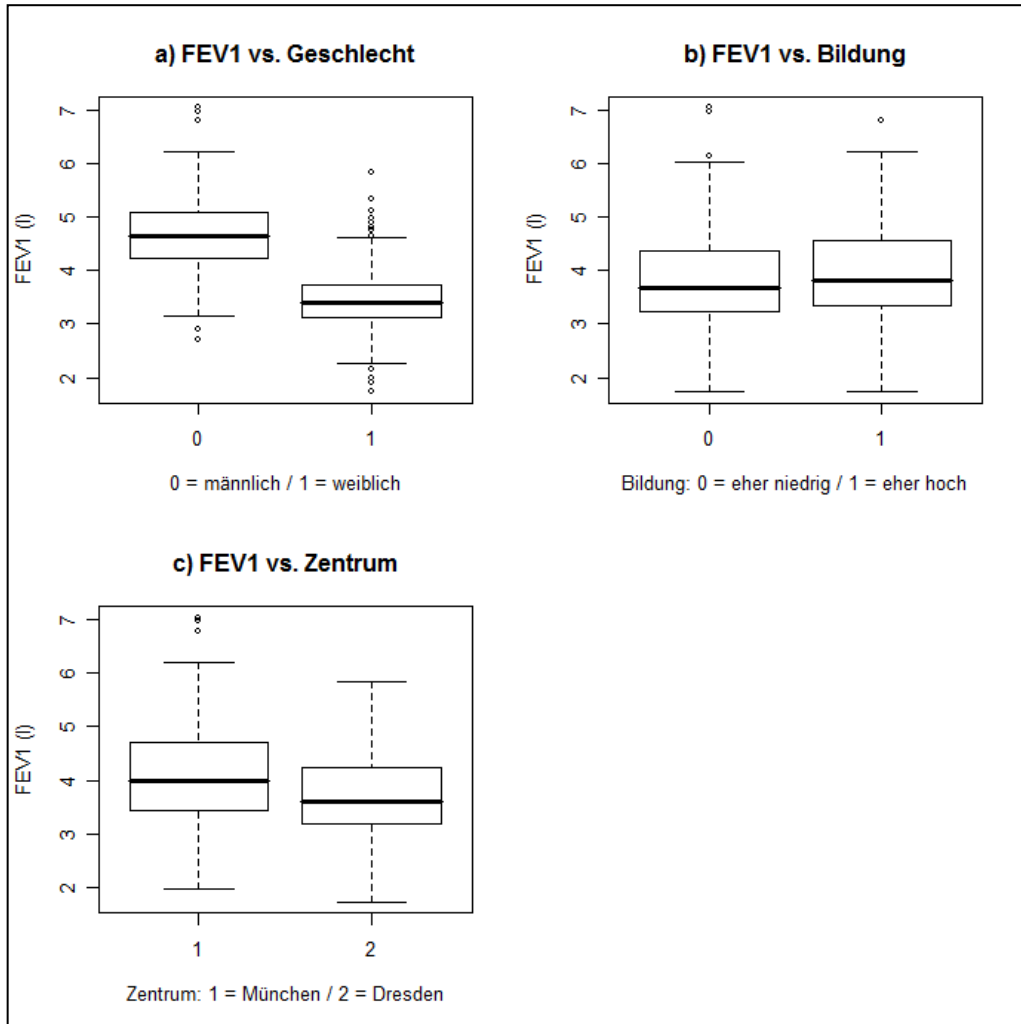


Abbildung 14: Geschlecht, Bildung und Erhebungszentrum

Verteilung der Einsekundenkapazität (FEV1) in Abhängigkeit vom Geschlecht ($n = 1.144$), der Bildung ($n = 1.142$) und dem Erhebungszentrum ($n = 1.144$)

Eine Korrelation zwischen der Einsekundenkapazität und der Bildung ist kaum erkennbar ($r_{\text{spearman}} = 0,08$). Der Median ist bei den Teilnehmern mit höherer Bildung mit 3,81 um 0,14 Liter höher als bei den Teilnehmern mit niedriger Bildung. Bei der Teilnehmergruppe aus Dresden wurden, bezogen auf den Median, geringere FEV1-Werte festgestellt. Der Median liegt bei Teilnehmern aus München bei 3,98 Litern und bei den Dresdner Teilnehmern bei 3,61 Litern. Die Korrelation nach Pearson (-0,22) deutet auf einen schwachen Zusammenhang zwischen dem Erhebungszentrum und dem Lungenfunktionsparameter (FEV1) hin.

Körpergröße, Gewicht und Lungenfunktionsparameter (ISAAC II-Studie)

Die *Abbildung 15* zeigt die Streudiagramme zwischen der Einsekundenkapazität und den metrischen Kovariablen „Körpergröße“ und „Gewicht“ sowie den Lungenfunktionsparametern (FEV1) aus der ISAAC II-Studie.

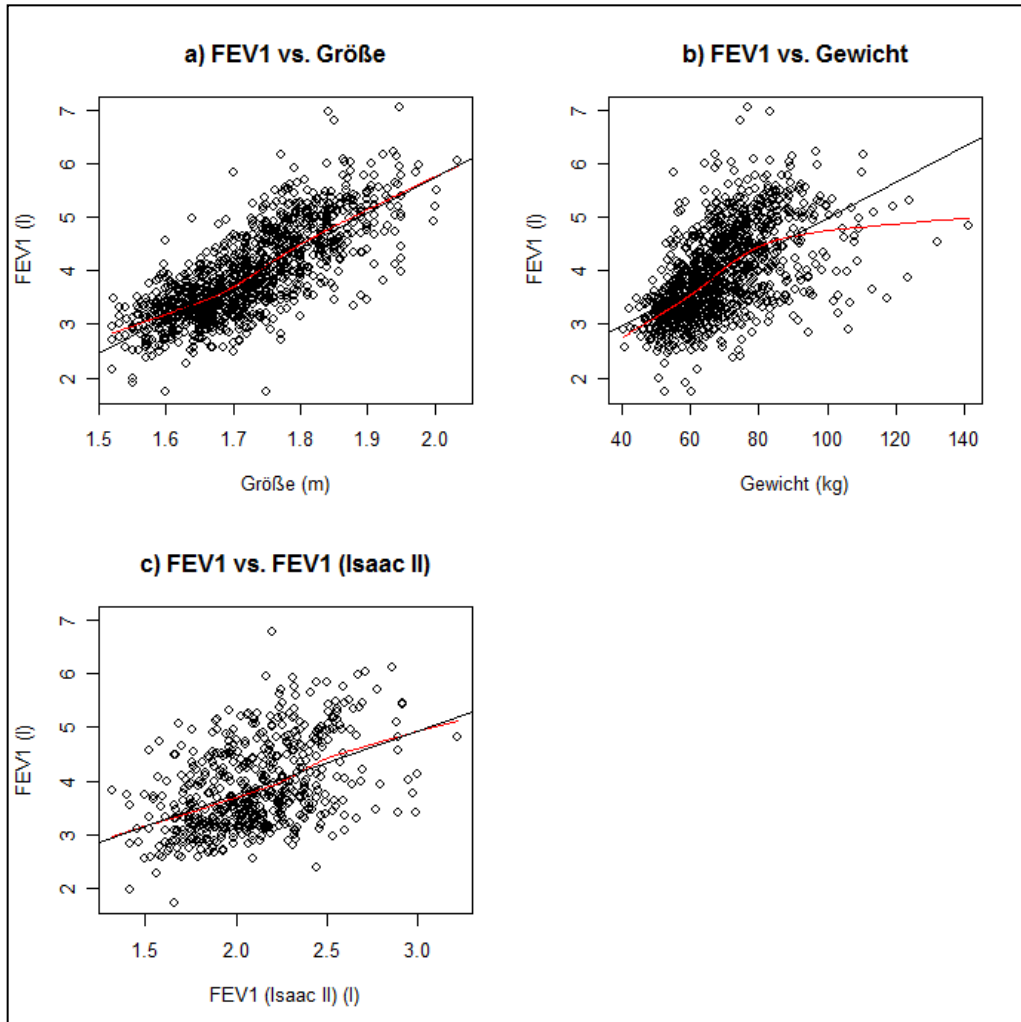


Abbildung 15: Körpergröße, Gewicht und Lungenfunktionsparameter (ISAAC II-Studie)
Streudiagramme zwischen der Einsekundenkapazität (FEV1) und
der Körpergröße ($n = 1.144$), dem Gewicht ($n = 1.143$) und
den FEV1-Werten der ISAAC II-Studie ($n = 565$)
mit Regressionsgerade (schwarz) und Streudiagramm-Glätter
(„lowess - locally weighted scatterplot smoothing“)

In *Abbildung 15 c*) ist der Zusammenhang zwischen den FEV1-Werten aus der SOLAR II-Studie und den Werten der Einsekundenkapazität aus der ISAAC II-Studie dargestellt. Dabei ist im Streudiagramm ein stark linear steigender Effekt der FEV1-Werte aus der ISAAC II-Studie zu sehen. D.h., dass bei Probanden, die im Alter von neun bis 11 Jahren

schon eine hohe Einsekundenkapazität aufwiesen, auch in der Follow-Up-Studie im Alter zwischen 21 und 23 Jahren tendenziell hohe FEV1-Werte gemessen wurden.

Die FEV1-Werte nehmen mit zunehmender Körpergröße linear zu. Der Zusammenhang zwischen dem Gewicht und den FEV1-Werten ist evtl. nicht linear. Bei Betrachtung der eingezeichneten Lowess-Kurve fällt auf, dass die FEV1-Werte mit zunehmendem Gewicht zunächst stark linear ansteigen. Ab einem Gewicht von ca. 80 kg flacht die Kurve ab, d.h. die Zunahme der FEV1-Werte fällt mit zunehmendem Gewicht geringer aus.

Asthma

Die *Abbildung 16 a)* zeigt die Verteilung der Einsekundenkapazität in Abhängigkeit von der Asthmad Diagnose der Teilnehmer. Anhand der Boxplots ist zu erkennen, dass die durchschnittlichen FEV1-Werte mit zunehmender Asthmad Diagnose fallen, d.h. je häufiger Asthma diagnostiziert wurde, desto schlechtere FEV1-Werte wurden in den medizinischen Untersuchungen gemessen. Allerdings fällt dieser Unterschied sehr minimal aus. Die Probanden, bei denen noch nie Asthma festgestellt wurde, weisen mediane FEV1-Werte von 3,78 Litern auf. Bei einmaliger Asthmad Diagnose lag der Median bei 3,67 und bei mehrmaliger Asthmad Diagnose bei 3,59 Litern. Die Korrelation nach Spearman ist mit -0,06 verschwindend gering.

Ein Zusammenhang zwischen beruflichem Asthmarisiko und der Einsekundenkapazität scheint ebenfalls nicht gegeben zu sein ($r_{\text{spearman}} = 0,05$). In *Abbildung 16 c)* und *d)* ist der Zusammenhang zwischen der elterlichen Asthmad Diagnose und den FEV1-Werten visualisiert. Die elterliche Asthmaanamnese scheint keinen Einfluss auf die Einsekundenkapazität der Kinder auszuüben. Der mediane FEV1-Wert (3,9 Liter) der Probanden bei einer positiven Asthmad Diagnose der Mutter unterscheidet sich kaum von dem Median (3,80 Liter) bei einer negativen Asthmad Diagnose der Mutter. Das gleiche Muster ist bei Betrachtung von den FEV1-Werten in Abhängigkeit von der väterlichen Asthmad Diagnose gegeben.

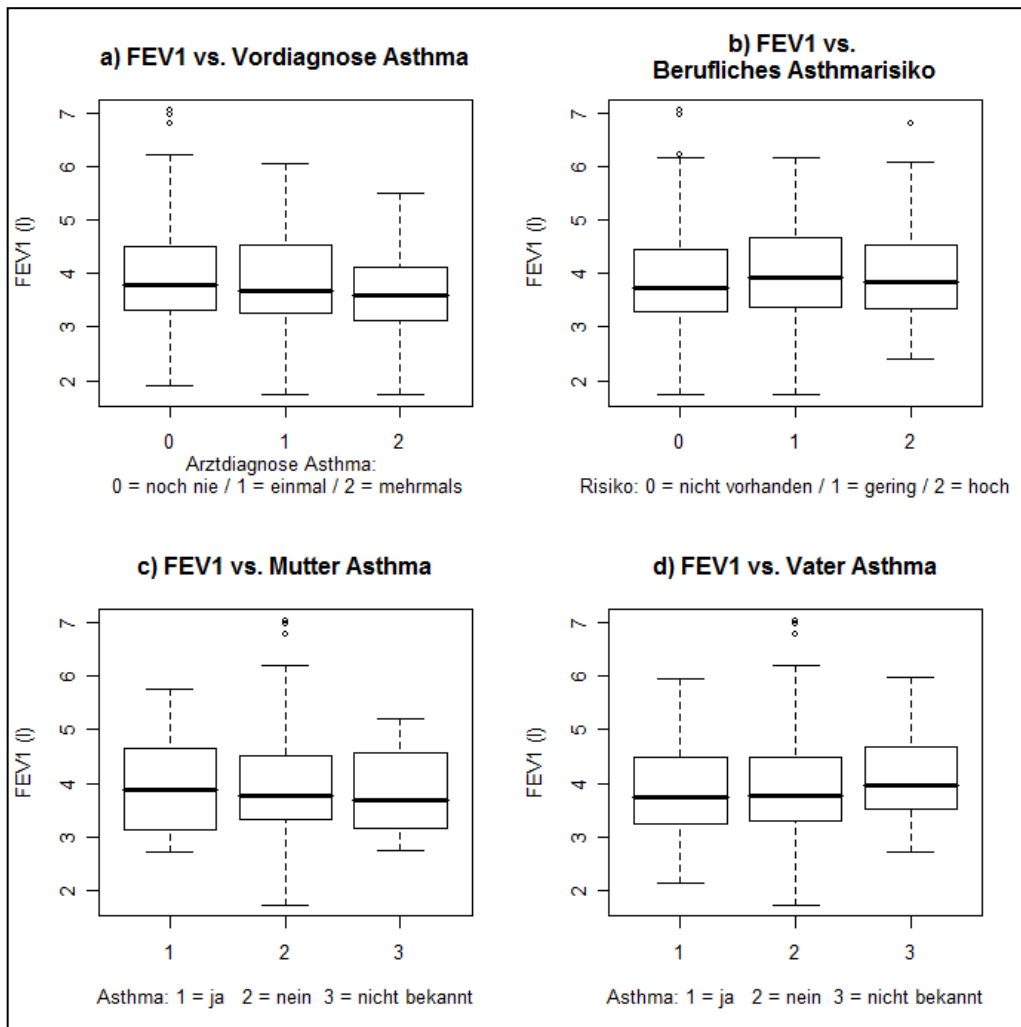


Abbildung 16: Asthma

Verteilung der Einsekundenkapazität (FEV1) in Abhängigkeit von der Asthmad Diagnose der Probanden ($n = 1.125$), dem beruflichem Asthmarisiko ($n = 1.143$) und den Variablen „Mutter Asthma“ ($n = 1.116$) und „Vater Asthma“ ($n = 1.116$)

3.2.2 Forcierte Vitalkapazität und potentielle Einflussgrößen/ Confounder

Aktives Rauchverhalten

Die graphische Darstellung des Zusammenhangs zwischen der forcierten Vitalkapazität und den Variablen, welche zum aktiven Rauchverhalten erhoben wurden, befindet sich in *Abbildung 17*.

Die Dauer des aktiven Rauchens und die Werte zur forcierten Vitalkapazität scheinen unkorreliert zu sein. Die Streubreite bleibt, bis auf extreme Werte, über die Jahre annähernd gleich. Der Mittelwert beträgt bei den Nichtraucher 4,58 Liter und im Bereich von sechs bis sieben Jahren bei der aktiven Rauchdauer 4,58 Liter. Die Lowess-Kurve nimmt

im Bereich von ca. fünf Jahren mit zunehmender Rauchdauer leicht linear ab. Eine Beeinträchtigung durch extreme Werte kann allerdings (vgl. *Abbildung 17 a*) nicht ausgeschlossen werden.

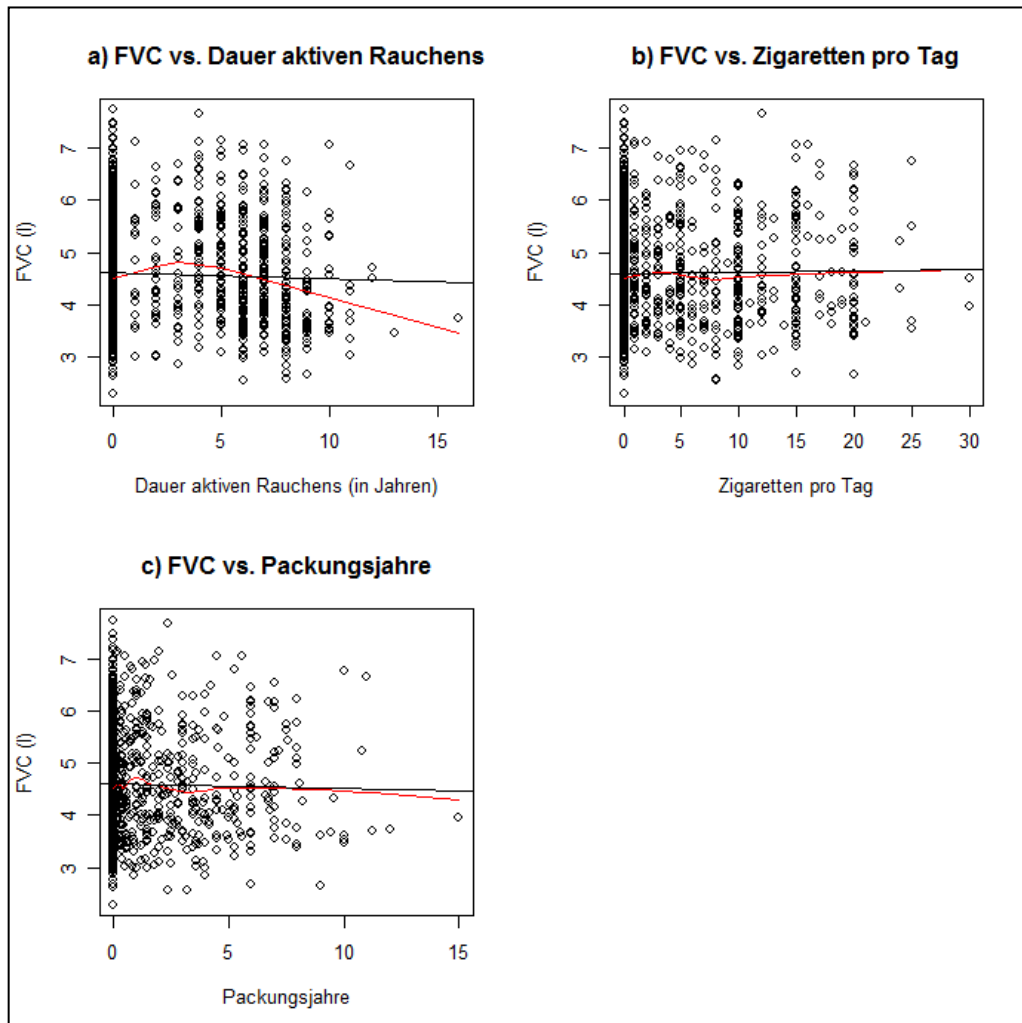


Abbildung 17: Aktives Rauchverhalten

Streudiagramme zwischen forcierter Vitalkapazität (FVC) und der Dauer des aktiven Rauchens ($n = 1.129$), der Anzahl Zigaretten pro Tag ($n = 1.124$) und den Packungsjahren ($n = 1.119$) mit Regressionsgerade (schwarz) und Streudiagramm-Glätter („lowess - locally weighted scatterplot smoothing“)

Auch zwischen der Anzahl an Zigaretten pro Tag und den FVC-Werten scheint kein Zusammenhang gegeben zu sein. Die Regressionsgerade und der Lowess-Schätzer verlaufen recht flach. Die Streubreite bleibt ebenfalls annähernd gleich. Der mittlere FVC-Wert beträgt bei den Nichtraucher 4,58 Liter. Bei Probanden, welche mehr als 10 Zigaretten pro Tag rauchten, beträgt der mittlere FVC-Wert 4,69 Liter.

Das Streudiagramm in *Abbildung 17 c)* zeigt, bezogen auf die Packungsjahre, ein ähnliches Bild. Auch zwischen der forcierten Vitalkapazität und den Packungsjahren ist kein Zusammenhang erkennbar.

Passivrauchexposition

In *Abbildung 18* wird die Verteilung der FVC-Werte in Abhängigkeit von der Passivrauchexposition betrachtet. Anhand der Boxplots ist kein Zusammenhang erkennbar. Die medianen FVC-Werte bei Teilnehmern, die zu Hause keinem Rauch anderer ausgesetzt sind, wurden mit 4,4 Litern ermittelt. In der Kategorie weniger als eine Stunde lag der Median 4,49 Litern und bei mehr als eine Stunde bei 4,22 Litern.

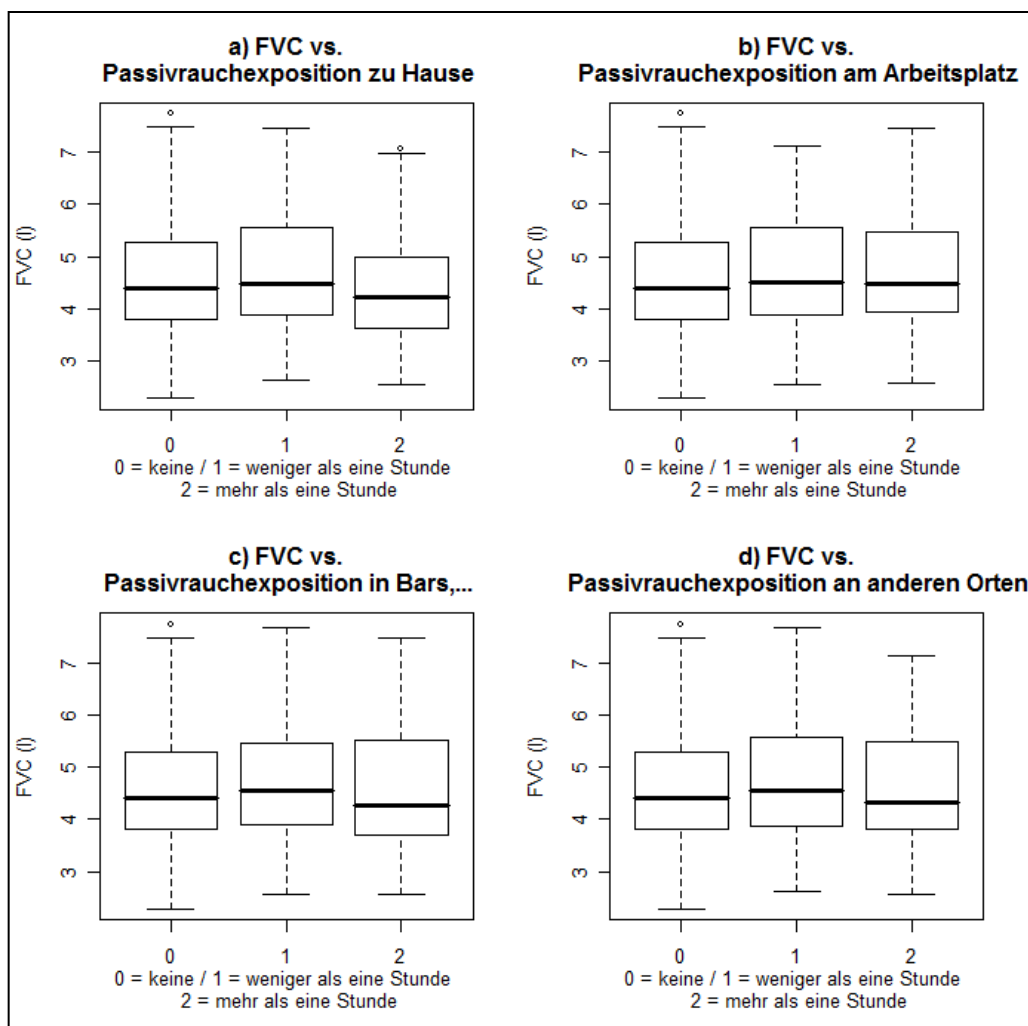


Abbildung 18: Passivrauchexposition

Verteilung der forcierten Vitalkapazität (FVC) in Abhängigkeit

von der Passivrauchexposition zu Hause (n = 945),

am Arbeitsplatz (n = 984), in Bars etc. (n = 1.073) und an anderen Orten (n = 1.000)

Der Korrelationskoeffizient nach Spearman betrug -0,01. D.h. man kann davon ausgehen, dass die Passivrauchexposition zu Hause in keinem Zusammenhang mit der forcierten Vitalkapazität steht. Bei Betrachtung der Verteilung der FVC-Werte in Abhängigkeit von der Passivrauchexposition am Arbeitsplatz ($r_{\text{spearman}} = 0,044$), in Bars etc. ($r_{\text{spearman}} = 0,007$) und an anderen Orten ($r_{\text{spearman}} = 0,024$) sind ähnliche Strukturen zu erkennen. Ebenso scheint hier kein Zusammenhang mit den FVC-Werten zu bestehen.

Auch die stetige Variable „kumulierte Passivrauchexposition“ scheint in keinem unmittelbaren Bezug zur forcierten Vitalkapazität zu stehen (vgl. *Abbildung 19*).

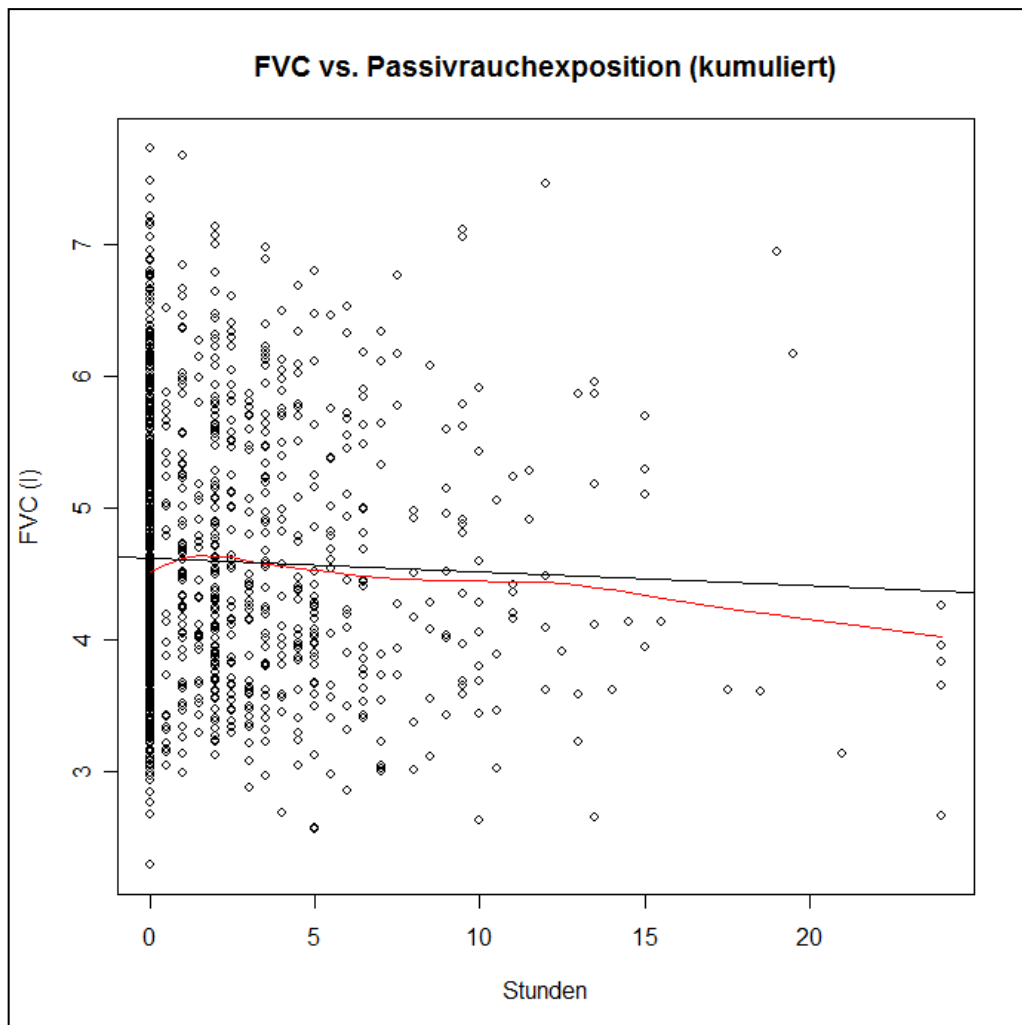


Abbildung 19: Passivrauchexposition (kumuliert)
Streudiagramme zwischen der forcierten Vitalkapazität (FVC) und
der kumulierten Passivrauchexposition (n = 1.131)
mit Regressionsgerade (schwarz) und Streudiagramm-Glätter
(„lowess - locally weighted scatterplot smoothing“)

Geschlecht, Bildung und Erhebungszentrum

Die Verteilung der forcierten Vitalkapazität in Abhängigkeit vom Geschlecht weist auf einen starken Zusammenhang hin (s.a. *Abbildung 20*). Der mediane FVC-Wert bei den Frauen ist mit 3,93 Litern wesentlich geringer als bei den Männern.

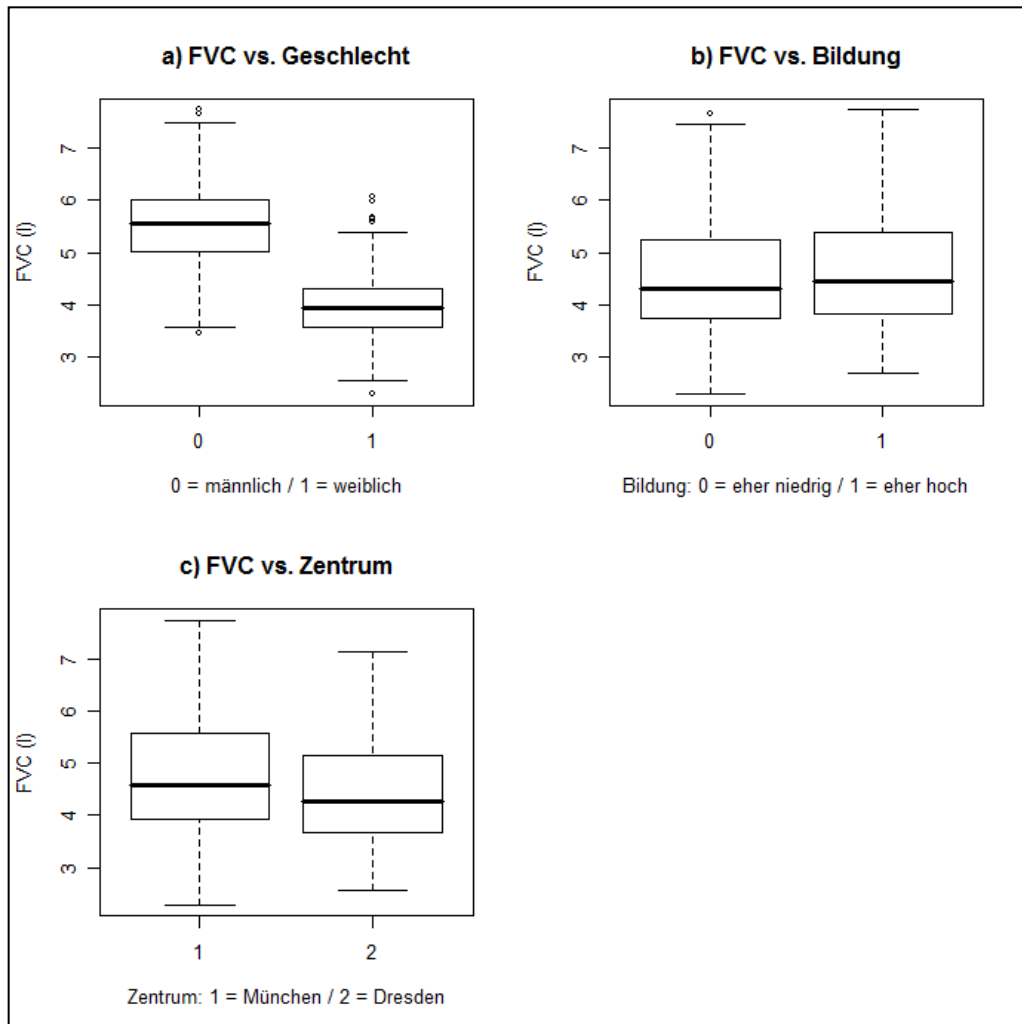


Abbildung 20: Geschlecht, Bildung und Erhebungszentrum

Verteilung der forcierten Vitalkapazität (FVC) in Abhängigkeit vom Geschlecht ($n = 1.041$), der Bildung ($n = 1.139$) und dem Erhebungszentrum ($n = 1.141$)

Die männlichen Teilnehmer weisen einen Median von 5,56 Litern auf. Die Berechnung der Korrelation nach Pearson ergab einen Wert von $-0,78$. Ein Zusammenhang zwischen dem Bildungsstand und der forcierten Vitalkapazität ist nicht erkennbar. Die Teilnehmer aus Dresden weisen geringere Medianwerte auf. Die medianen FVC-Werte lagen hier mit 4,26 um 0,325 Liter niedriger als die FVC-Werte der Münchener Gruppe ($r_{\text{pearson}} = -0,143$).

Körpergröße, Gewicht und Lungenfunktionsparameter (ISAAC II-Studie)

Die Streudiagramme in *Abbildung 21* zeigen die forcierte Vitalkapazität in Abhängigkeit von der Körpergröße, dem Gewicht und der forcierten Vitalkapazität aus der ISAAC II-Studie.

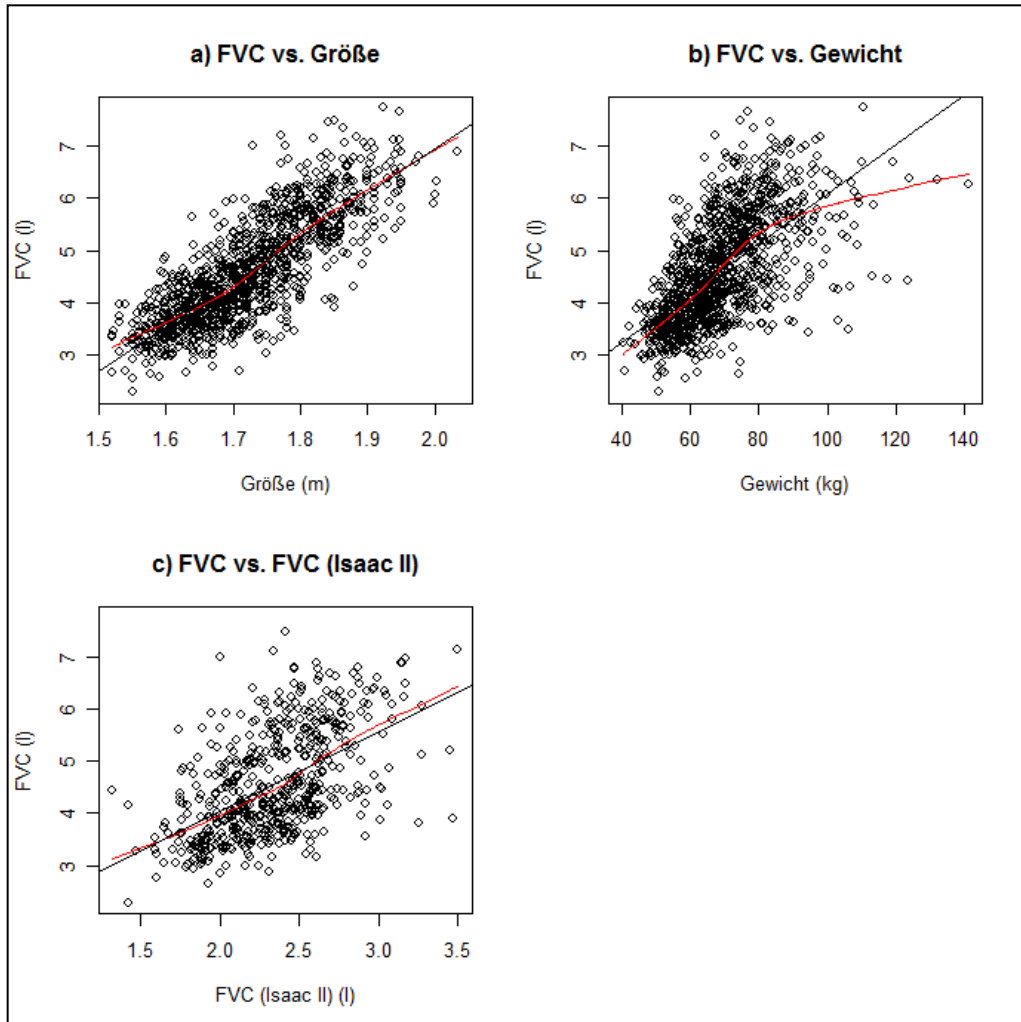


Abbildung 21: Körpergröße, Gewicht und Lungenfunktionsparameter (ISAAC II-Studie)

Streudiagramme zwischen der forcierten Vitalkapazität(FVC) und

der Körpergröße ($n = 1.141$), dem Gewicht ($n = 1.140$)

und den FVC-Werten der ISAAC II-Studie ($n = 564$)

mit Regressionsgerade (schwarz) und Streudiagramm-Glätter

Die FVC-Werte der SOLAR II-Studie scheinen mit der Körpergröße stark linear anzusteigen. Der Zusammenhang zwischen der forcierten Vitalkapazität und dem Gewicht ist evtl. nicht-linear. Wie bei der Einsekundenkapazität (FEV1) steigt die Lowess-Kurve zunächst linear an. Ab einem Gewicht von ca. 80 kg flacht die Kurve ab. Die FVC-Werte steigen nur noch leicht mit wachsendem Gewicht der Probanden. Die *Abbildung 21 c)* zeigt das

Streudiagramm zwischen den FVC-Werten, die in der klinischen Untersuchung der SOLAR II-Studie gemessen wurden und den FVC-Werten aus der ISACC II-Studie. Die beiden Größen sind deutlich positiv assoziiert.

Asthma

Die möglichen Einflussgrößen, welche im Bezug zum Asthma in das Datenmaterial aufgenommen wurden, sind in *Abbildung 22* dargestellt. Die medianen FVC-Werte in Bezug zur Asthmadiagnose der Teilnehmer sind in allen Kategorien mit ca. 4,5 Litern nahezu gleich.

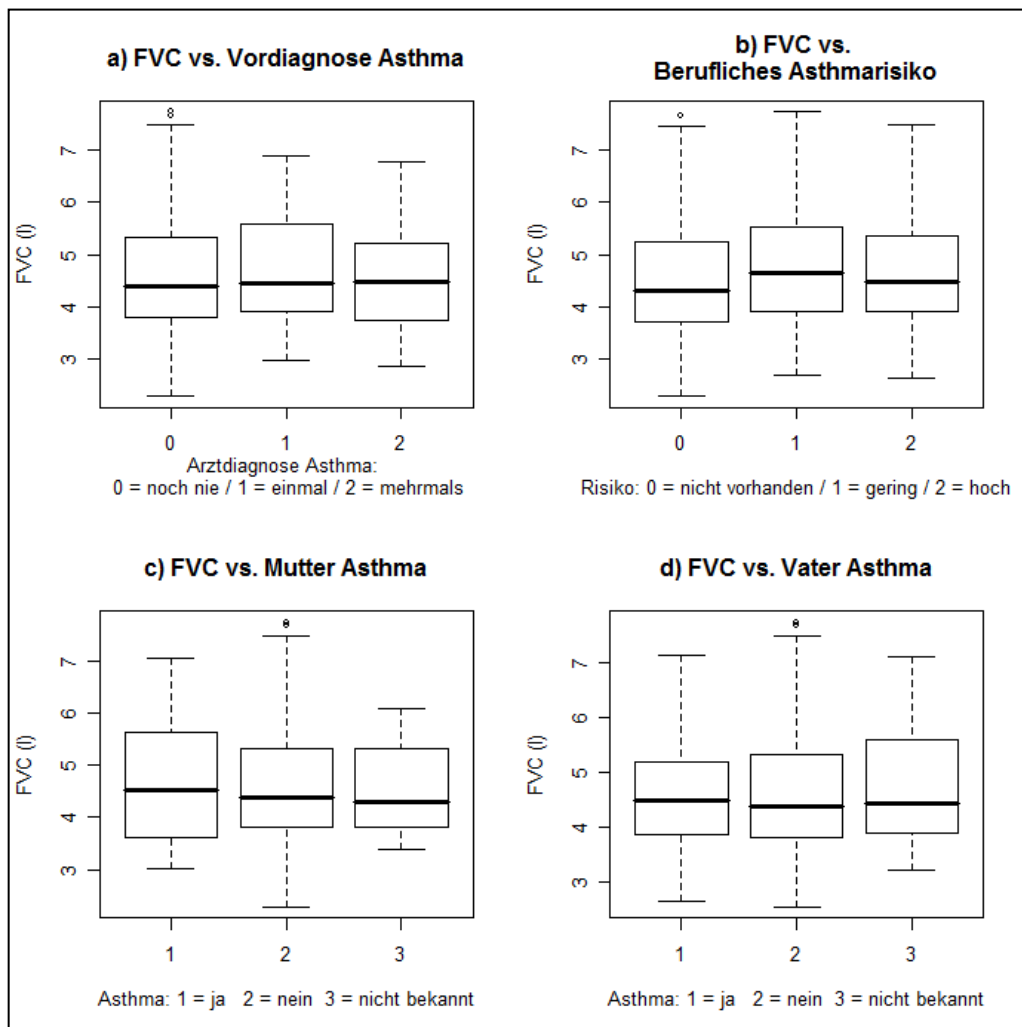


Abbildung 22: Asthma

Verteilung der forcierten Vitalkapazität (FVC) in Abhängigkeit von der Asthmadiagnose der Probanden ($n = 1.122$), dem beruflichen Asthmarisiko ($n = 1.140$) und den Variablen „Mutter Asthma“ ($n = 1.113$) und „Vater Asthma“ ($n = 1.113$)

Der Eindruck, dass kein Zusammenhang vorliegt, wird durch den berechneten Korrelationskoeffizienten nach Spearman von 0,01 untermauert. Auch die elterliche Asthmadia-

se und die forcierte Vitalkapazität scheinen unkorreliert zu sein. Die medianen FVC-Werte lagen bei einer positiven mütterlichen bzw. väterlichen Asthmadiagnose bei 4,5 Litern. Bei den Müttern bzw. den Vätern ohne Asthmadiagnose lag der mediane FVC-Wert bei 4,4 Litern. Das berufliche Asthmarisiko scheint ebenfalls keinen Einfluss auf die forcierte Vitalkapazität zu haben ($r_{\text{spearman}} = 0,08$).

3.2.3 Tiffeneau-Wert und potentielle Einflussgrößen/ Confounder

Aktives Rauchverhalten

Ein Zusammenhang zwischen dem Tiffeneau-Wert und den Variablen zum aktiven Rauchverhalten scheint nicht gegeben. Die *Abbildung 23* zeigt die Streudiagramme.

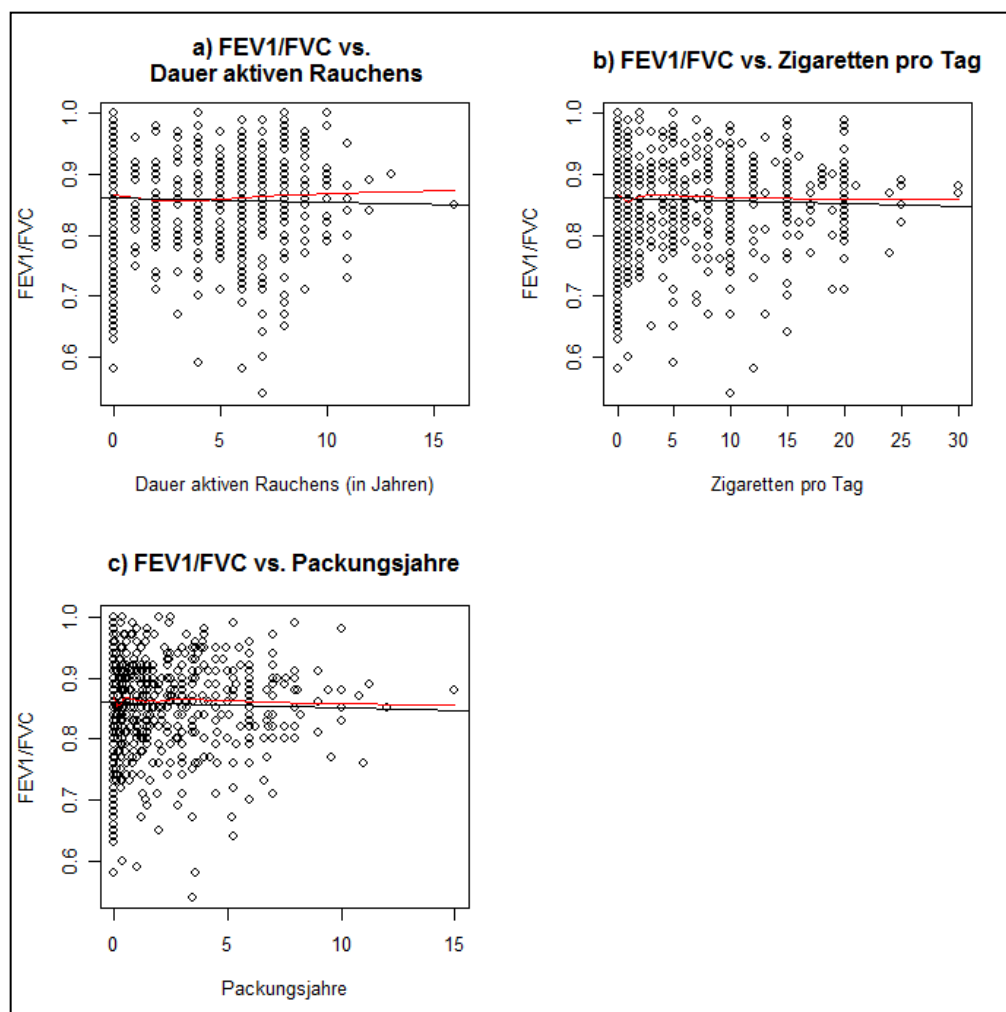


Abbildung 23: Aktives Rauchverhalten

Streudiagramme zwischen Tiffeneau-Wert und der Dauer aktiven Rauchens ($n = 1.129$), der Anzahl Zigaretten pro Tag ($n = 1.124$) und den Packungsjahren ($n = 1.119$) mit Regressionsgerade (schwarz) und Streudiagramm-Glätter („lowess - locally weighted scatterplot smoothing“)

Der mittlere Tiffeneau-Wert in Bezug zur Rauchdauer liegt bei ca. 86%. Eine Ab- bzw. Zunahme mit Anstieg der Rauchdauer ist nicht zu erkennen. Die Regressionsgerade und der Lowess-Schätzer verlaufen flach. Auch eine Betrachtung des Zusammenhangs zwischen der Anzahl an Zigaretten pro Tag und den Packungsjahren mit dem Tiffeneau-Wert zeigen ähnliche Muster. Die mittleren Tiffeneau-Werte bewegen sich hier ebenfalls gleichbleibend bei ca. 86%.

Passivrauchexposition

In *Abbildung 24* sind die Verteilungen der Tiffeneau-Werte in Abhängigkeit der Variablen, welche zur Passivrauchexposition erhoben wurden, dargestellt.

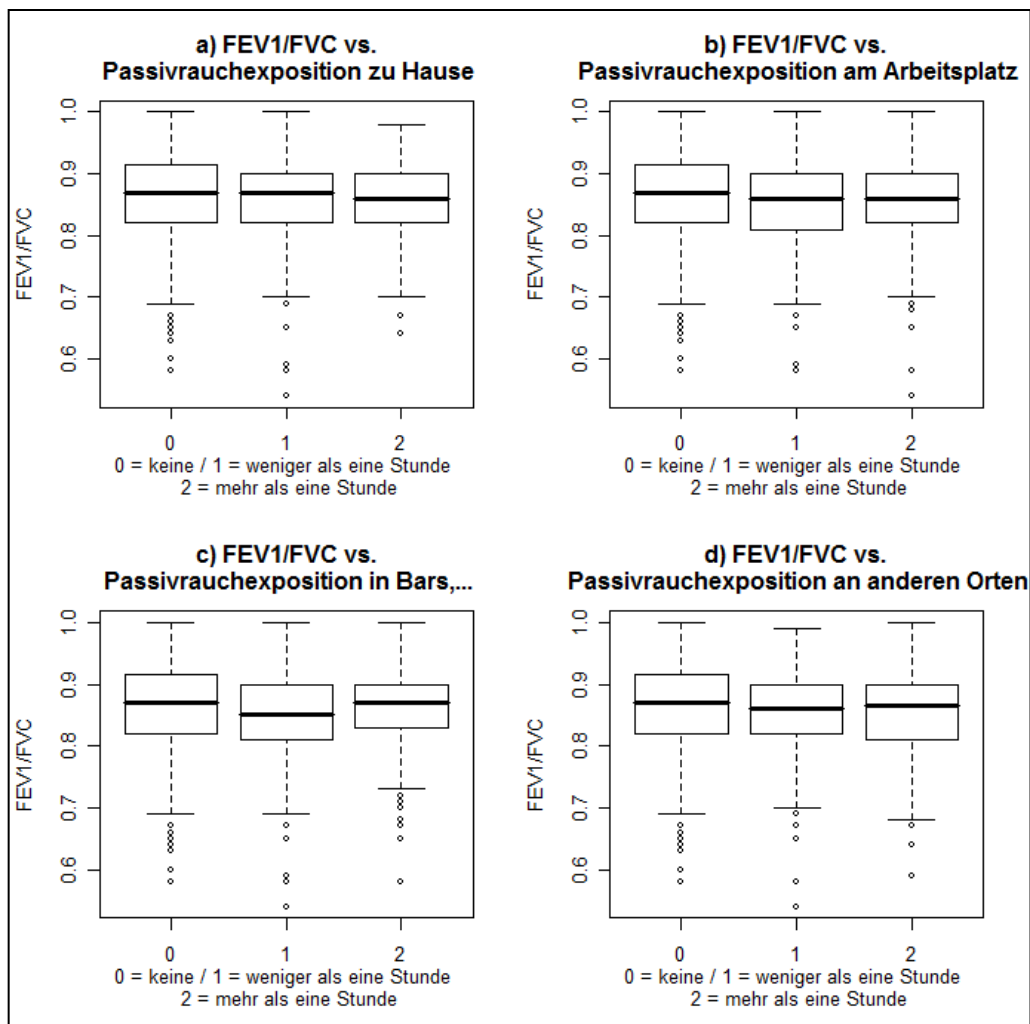


Abbildung 24: Passivrauchexposition

Verteilung der Tiffeneau-Werte in Abhängigkeit

von der Passivrauchexposition zu Hause (n = 945),

am Arbeitsplatz (n = 984), in Bars etc.(n = 1.073) und an anderen Orten (n = 1.000)

Alle Verteilungen weisen eine ähnliche Struktur auf. Bei jeder der abgefragten Kategorien lagen die Mediane des Tiffeneau-Wertes im Bereich von ca. 85% bis 87%. Auch die Streubreiten aller dargestellten Zusammenhänge zeigen keine extremen Unterschiede. Die errechneten Korrelationskoeffizienten nach Spearman liegen alle bei ca. -0,04. Auch eine Assoziation zwischen dem Tiffeneau-Wert und den einzelnen Variablen zur Passivrauchexposition scheint nicht gegeben zu sein. Auffallend sind extreme Werte in allen Kategorien unterhalb von ca. 70%.

In *Abbildung 25* ist das Streudiagramm zwischen der relativen Einsekundenkapazität und der kumulierten Passivrauchexposition graphisch dargestellt. Ein Zusammenhang zwischen den Variablen kann nicht abgeleitet werden.

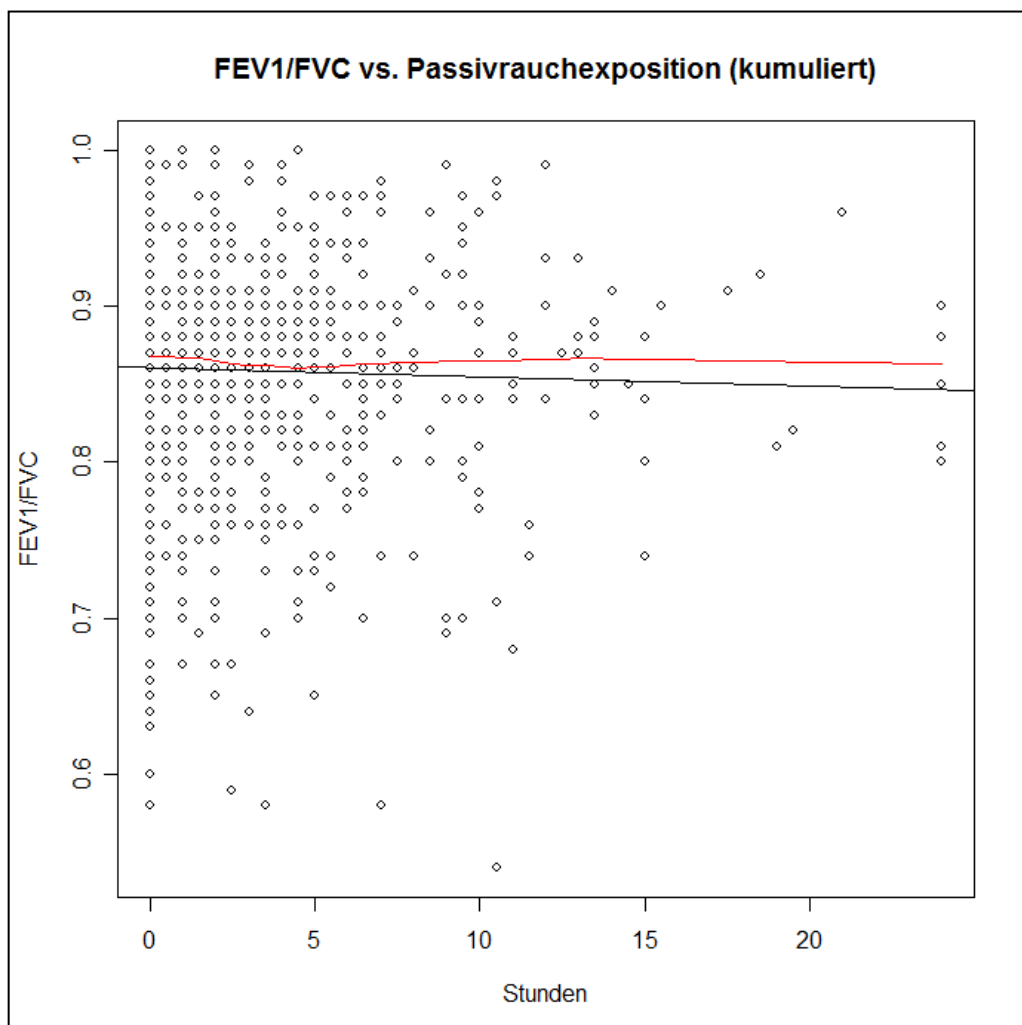


Abbildung 25: Passivrauchexposition (kumuliert)
Streudiagramme zwischen dem Tiffeneau-Wert (FEV1/ FVC) und
der kumulierten Passivrauchexposition (n = 1.131) mit Regressionsgerade (schwarz)
und Streudiagramm-Glätter („lowess - locally weighted scatterplot smoothing“)

Geschlecht, Bildung und Erhebungszentrum

Gemäß *Abbildung 26 a)* ergeben die stratifizierten Analysen nach dem Geschlecht für die weiblichen Probanden einen besseren Tiffeneau-Wert als bei den Männern.

Bezogen auf den Median weisen die männlichen Probanden einen Tiffeneau-Wert von 85% und die Frauen von 88% auf. Auch der Korrelationskoeffizient nach Pearson von 0,194 deutet auf einen Zusammenhang hin.

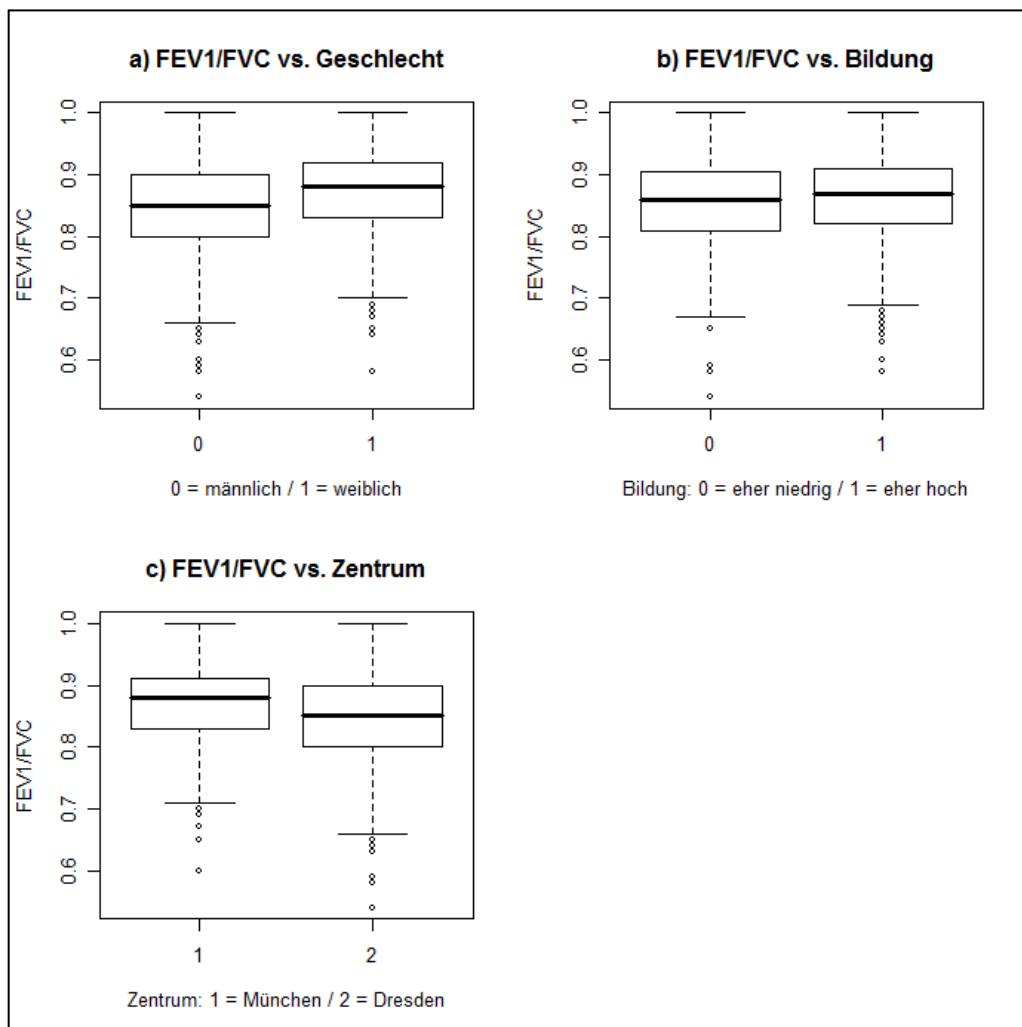


Abbildung 26: Geschlecht, Bildung und Erhebungszentrum

Verteilung der Tiffeneau-Werte in Abhängigkeit vom

Geschlecht ($n = 1.141$), der Bildung ($n = 1.139$) und dem Erhebungszentrum ($n = 1.141$)

Eine Assoziation mit der Bildungsvariablen scheint nicht vorhanden zu sein. Der Korrelationskoeffizient nach Spearman beträgt 0,045. Die Mediane der Kategorien mit niedriger und höherer Bildung liegen mit 86 % und 87 % dicht beieinander. Bei der Gruppe der aus Dresden stammenden Teilnehmer liegt der Median unterhalb der Münchner Gruppe. Die

Abbildung 26 c) und die Korrelation nach Pearson von -0,155 deuten hier auf einen Zusammenhang hin.

Körpergröße, Gewicht und Lungenfunktionsparameter (ISAAC II-Studie)

Die Ergebnisse der Zusammenhangsanalyse zur Körpergröße, dem Gewicht und der Tiffeneau-Werte aus der ISAAC II-Studie gegenüber der relativen Einsekundenkapazität aus der SOLAR II-Studie sind in *Abbildung 27* dargestellt.

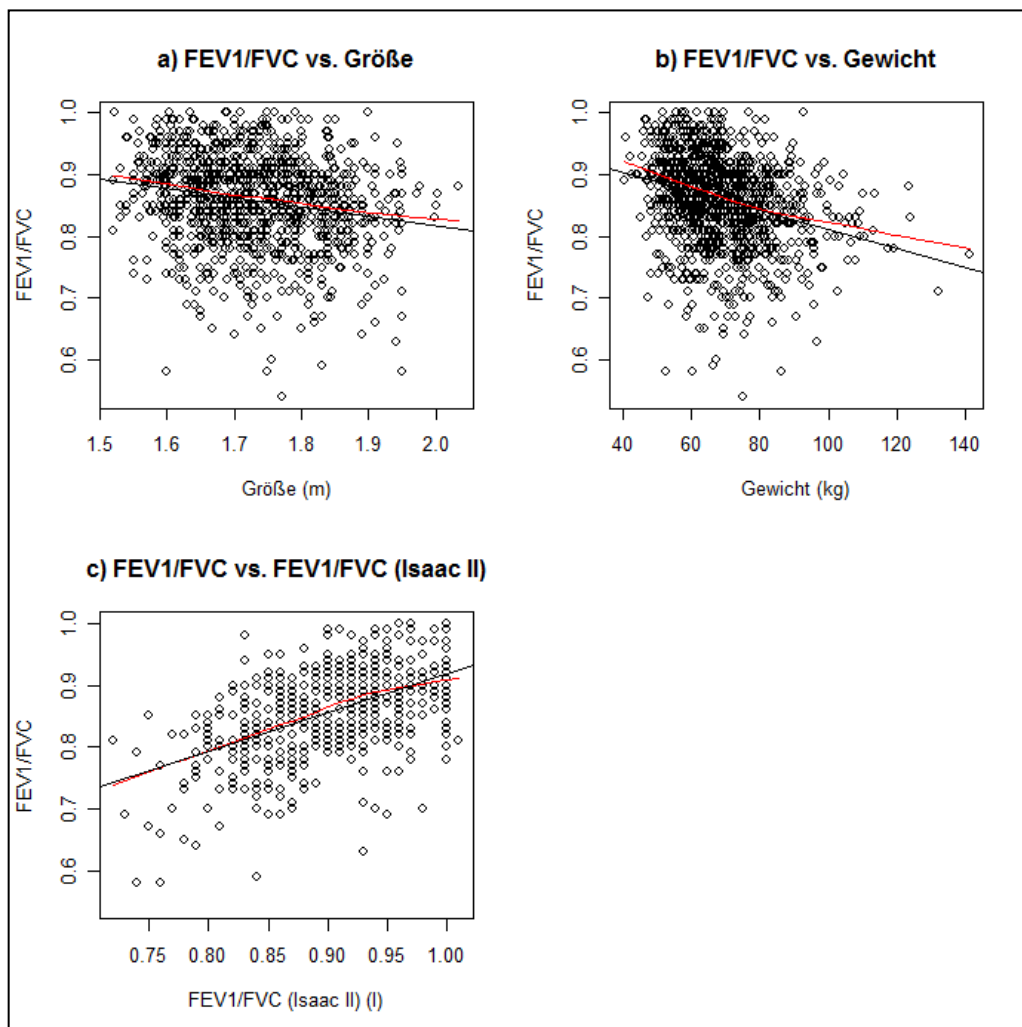


Abbildung 27: Körpergröße, Gewicht und Lungenfunktionsparameter (ISAAC II-Studie)

Streudiagramme zwischen den Tiffeneau-Werten und der Körpergröße ($n = 1.141$), dem Gewicht ($n = 1.140$) und den Tiffeneau-Werten der ISAAC II-Studie ($n = 564$) mit Regressionsgerade (schwarz) und Streudiagramm-Glätter („lowess - locally weighted scatterplot smoothing“)

Die KQ-Geraden und der geschätzte Verlauf der Streudiagramm-Glätter deuten auf lineare Zusammenhänge hin. Die Tiffeneau-Werte der SOLAR II-Studie fallen mit zunehmender Körpergröße bzw. mit zunehmendem Gewicht linear ab. Somit lässt sich sagen, dass sich die Tiffeneau-Werte mit zunehmender Körpergröße bzw. Gewicht verschlechtern. Allerdings steigen die Tiffeneau-Werte aus der SOLAR II-Studie mit wachsender relativer Einsekundenkapazität aus der Vorgängerstudie. D.h. je positiver die Tiffeneau-Werte im Alter zwischen neun und 11 Jahren im Rahmen der ISAAC II-Studie waren, desto besser fielen die mittleren Tiffeneau-Werte in der Follow-Up-Studie aus.

Asthma

Die nachfolgende *Abbildung 28* zeigt die Assoziationen zwischen dem Tiffeneau-Wert und den asthmarelevanten Fragestellungen.

Die *Abbildung 28 a)* zeigt, dass die mittleren Tiffeneau-Werte fallen, je häufiger Asthma bei den Probanden festgestellt wurde. Die Probanden, welche niemals eine Asthmadiagnose erhielten hatten einen medianen Wert von 87%. Bei einmaliger Diagnose lag der Median bei 83,5% und bei mehrmaliger Asthmadiagnose eines Arztes wurde ein Median von 82,5 % errechnet. Durch einen Korrelationskoeffizienten von -0,15 (Spearman) wird ein möglicher negativer Zusammenhang unterstützt.

Zwischen dem beruflichen Asthmarisiko und dem Tiffeneau-Wert scheint kein Zusammenhang zu bestehen. Auch die elterliche Asthmaanamnese scheint nicht mit der relativen Einsekundenkapazität assoziiert zu sein. Die jeweiligen Mediane lagen bei einer mütterlichen Asthmadiagnose bei 87% und bei keiner ärztlichen Diagnose bei 86%, bei der väterlichen Asthmadiagnose bei 85% und bei keiner medizinischen Asthmadiagnose bei 86%.

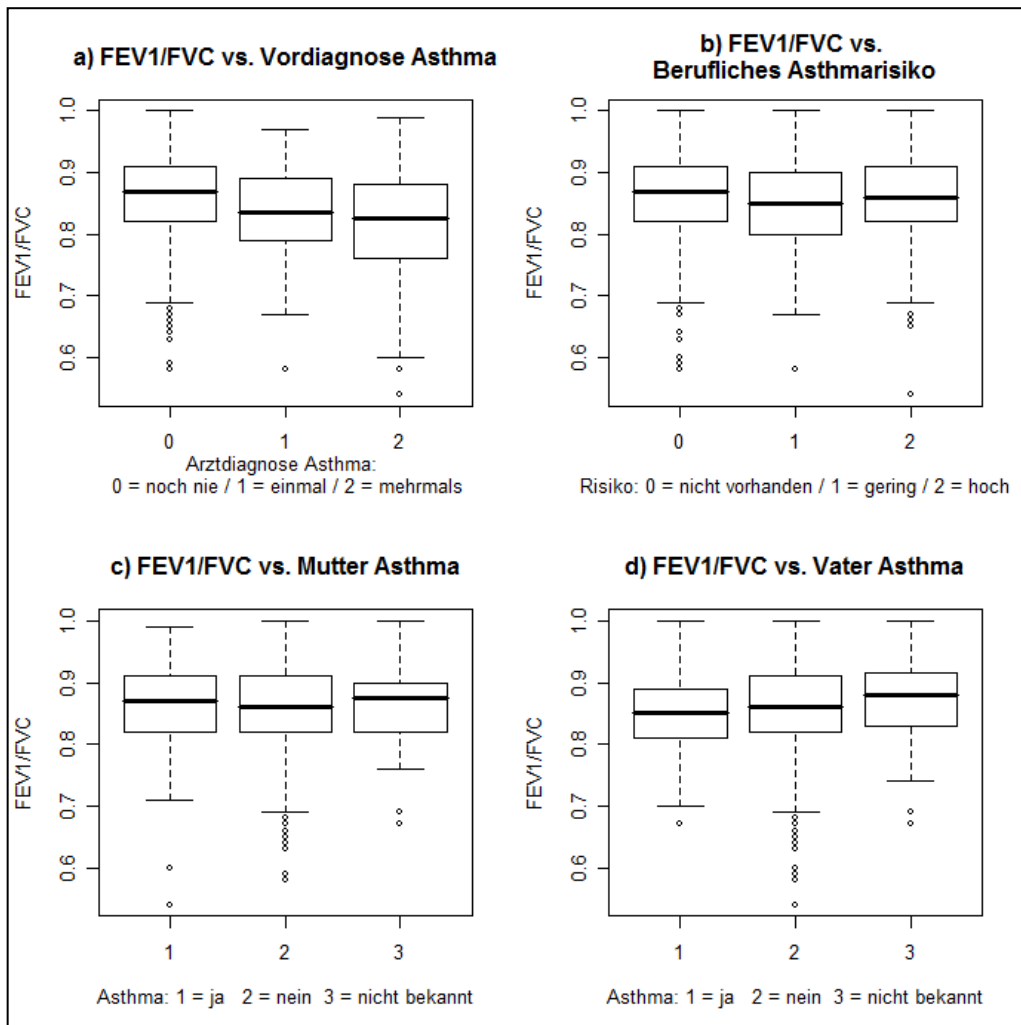


Abbildung 28: Asthma

Verteilung der Tiffeneau-Werte in Abhängigkeit von der Asthmadiagnose der Probanden ($n = 1.122$), dem beruflichen Asthmarisiko ($n = 1.140$) und den Variablen „Mutter Asthma“ ($n = 1.113$) und „Vater Asthma“ ($n = 1.113$)

4. *Multiple lineare Regression*

Das Ziel dieser Arbeit war die Anpassung von Regressionsmodellen für die stetigen Zielgrößen „Einsekundenkapazität (FEV1)“, „forcierte Vitalkapazität (FVC)“ und „Tiffeneau-Wert (FEV1/FVC)“. Mit den Ergebnissen aus dem dritten Kapitel wurde zunächst ein multiples lineares Regressionsmodell zugrunde gelegt. Die Modellierungen erfolgten auf Basis der kompletten Fälle („*complete cases*“), d.h. alle unvollständig beobachteten Datenzeilen wurden aus der Datenmatrix gelöscht. Da zwischen den Geschlechtern bereits anatomisch bedingte Unterschiede bzgl. der Lungenvolumina gegeben sind, wurden außerdem geschlechtsspezifische Regressionsmodelle berechnet. Zusätzlich zu den gewonnenen Resultaten aus der Zusammenhangsanalyse über lineare bzw. nicht-lineare Beziehungen zwischen Einfluss- und Zielgrößen wurde das GAM (generalisiertes additives Modell) genutzt, um zu entscheiden, welchen Einfluss eine Kovariable hat, wenn die Auswirkungen der anderen Kovariablen berücksichtigt werden. Da die gewählten Modelle möglichst keine irrelevanten Kovariablen enthalten sollten bzw. eine möglichst sparsame Modellspezifikation anzustreben ist (vgl. Fahrmeir et.al (2007, S. 157)), werden zwei verschiedene Modellwahlkriterien herangezogen. Zunächst wurde anhand des Informationskriteriums nach Akaike (AIC) ein möglichst gutes Modell bestimmt. In einem zweiten Schritt wurde die getroffene Modellauswahl mit dem Konzept der Kreuzvalidierung überprüft. Den nachfolgenden theoretischen Ausführungen zu den Modellannahmen, zur Parameterschätzung, zum generalisierten additiven Modell, zum AIC-Kriterium sowie zur Kreuzvalidierung ist vorzuschicken, dass diese - soweit nicht anders angegeben - aus der im Literaturverzeichnis angegebenen Quelle zur Regression (Fahrmeir et al (2007)) entnommen wurden.

4.1 Modellannahmen

Bei einem Regressionsmodell sollen Eigenschaften einer Zielvariablen y in Abhängigkeit von Kovariablen x_1, \dots, x_k beschrieben werden:

$$y = f(x_1, \dots, x_k) + \varepsilon$$

mit y als abhängige Variable,

$f(x_1, \dots, x_k)$ als systematische Komponente und

ε als Fehlerterm bzw. stochastische Komponente.

Das Ziel ist es, die systematische Komponente f aus den gegebenen Daten

$$y_i, x_{i1}, \dots, x_{ik}, i = 1 \dots n$$

zu schätzen und von der stochastischen Komponente zu trennen.

Wenn unterstellt wird, dass die Funktion f linear ist, ergibt sich die Klasse der linearen Regressionsmodelle.

Die unbekannte Funktion $f(x_1, \dots, x_k)$ wird als Linearkombination der Kovariablen modelliert, d.h.

$$f(x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k.$$

Die geschätzte lineare Funktion

$$\hat{f}(x_1, \dots, x_k) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k$$

kann als Schätzung $\hat{E}(y|x_1, \dots, x_k)$ für den bedingten Erwartungswert von y bei gegebenen Kovariablen x_1, \dots, x_k angesehen werden.

In Matrixschreibweise lautet das Modell:

$$y = X\beta + \varepsilon$$

Mit den Annahmen:

1. $E(\varepsilon) = 0$

2. $Cov(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I$

3. die Designmatrix X besitzt vollen Spaltenrang, d.h. $rg(X) = k + 1 = p$

wobei k = Anzahl der Einflussgrößen, 1 steht für den Intercept.

Zur Konstruktion von Konfidenzintervallen und Hypothesentests wird in vielen Fällen zusätzlich die folgende Annahme getroffen:

4. $\varepsilon \sim N(0, \sigma^2 I)$

4.2 Parameterschätzung

Das wohl am häufigsten verwendete Verfahren zur Schätzung der Regressionskoeffizienten ist die Methode der kleinsten Quadrate (KQ-Methode). Bei dieser Vorgehensweise werden die Werte von β gesucht, so dass

$$KQ(\beta) = \sum_{i=1}^n (y_i - x_i' \beta)^2 = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon$$

minimal wird.

Der Schätzer

$$\hat{\beta} = (X'X)^{-1}X'y$$

minimiert das KQ-Kriterium.

Ein anderes Verfahren stellt die Maximum-Likelihood-Methode dar. Die Maximum-Likelihood-Methode (ML-Methode) ist eine allgemeine Methode zur Bestimmung von Parametern aus Stichproben für beliebige Wahrscheinlichkeitsverteilungen. Die Parameter der Schätzfunktion werden dabei so gewählt, dass die Wahrscheinlichkeit, genau die Beobachtungspunkte der vorliegenden Stichprobe zu erhalten, maximal wird, d.h. die Schätzwerte maximieren die so genannte Likelihood-Funktion. Die Anwendung dieser Methode setzt allerdings voraus, dass die Wahrscheinlichkeitsverteilung der Störterme bekannt ist. Sind diese normalverteilt mit dem Erwartungswert Null und konstanter Varianz, so resul-

tiert daraus die gleiche Schätzfunktion für die Regressionskoeffizienten, wie bei der KQ-Methode.

Die Schätzung der Varianz der Störgrößen erfolgt mit der oben beschriebenen ML-Methode.

Als Schätzer resultiert:

$$\hat{\sigma}_{ML}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}$$

Dieser Schätzer ist jedoch nicht erwartungstreu für σ^2 . Eine geringfügige Modifikation liefert einen erwartungstreuen Schätzer:

$$\hat{\sigma}^2 = \frac{1}{n-p} \hat{\varepsilon}'\hat{\varepsilon}$$

4.3 Generalisiertes additives Modell (GAM)

Die generalisierten additiven Modelle (GAM) stellen eine Art der Erweiterung zu den generalisierten linearen Modellen (GLM) dar. Die GLMs umfassen in einem einheitlichen Rahmen viele Regressionsansätze und enthalten zudem das klassische lineare Modell als Spezialfall. Im klassischen linearen Regressionsmodell wird der Einfluss der metrischen Kovariablen auf die Zielgröße als linear angenommen. Eine rein lineare Modellierung ist jedoch in vielen praktischen Anwendungen nicht ausreichend. Die generalisierten additiven Modelle ermöglichen eine flexible Modellierung des Einflusses metrischer Kovariablen auf die abhängige Variable. Die allgemeine Form eines additiven Modells ist:

$$y_i = f_1(z_{i1}) + \dots + f_q(z_{iq}) + \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i$$

Die Funktionen f_1, \dots, f_q stellen die Effekte der metrischen Kovariablen z_1, \dots, z_q dar und werden mit nichtparametrischen Techniken geschätzt. Im Rahmen dieser Arbeit wurde das GAM dazu genutzt, um zu entscheiden, ob der potentielle Effekt der metrischen Kovariablen linear bzw. nicht-linear modelliert werden konnte.

4.4 Variablenselektion auf Basis des AIC-Kriteriums

Im Rahmen dieser Arbeit standen viele mögliche Einflussgrößen/ Confounder zur Verfügung. Um sicher zu gehen, dass nach dem Akaike Informationskriterium das „beste“ Modell für jede Zielgröße der Variablenteilmengen gefunden wird, wäre es erforderlich gewe-

sen, für jede mögliche Variablenkombination eine Regressionsrechnung durchzuführen. Die Anzahl an Kombinationen kann mit $2^k - 1$ berechnet werden (k = Anzahl an Kovariablen). Bei einer geringen Anzahl an potentiellen Einflussgrößen ist diese Vorgehensweise zwar aufwendig, aber durchführbar. Bei einer Anzahl von z.B. 10 Kovariablen ergeben sich aber bereits 1.023 verschiedene Kombinationen, was ein erhebliches Problem bei der Suche nach dem „besten Modell“ darstellt.

Neben der Betrachtung aller $2^k - 1$ Modelle, bieten Selektionsverfahren eine hilfreiche Alternative zur vollständigen Variablenselektion. Unter dem Begriff der Selektionsverfahren sind Prozesse zusammengefasst, in welchem jeweils eine Variable dem Modell hinzugefügt bzw. aus dem Modell entfernt wird, bis ein vorgegebenes Bewertungskriterium (hier AIC) sich nicht mehr verbessern lässt. Die Methode der Vorwärtsselektion beginnt mit einem Startmodell, von dem ausgehend mit jedem Schritt des Verfahrens eine weitere Kovariable aufgenommen wird, und zwar diejenige, welche die größte Reduktion nach dem AIC-Kriterium liefert. In der Rückwärtsselektion wird mit dem vollen Modell, in dem alle Kovariablen enthalten sind, gestartet. Anschließend wird in jeder Iteration diejenige Kovariable entfernt, welche die größte Reduktion nach dem AIC-Kriterium liefert. Weder Vorwärts- noch Rückwärtsselektion berücksichtigen den Effekt, dass das Hinzufügen oder Eliminieren einer Kovariablen den Beitrag anderer Variablen am Modell beeinflussen kann. Durch die schrittweise Selektion kann dieser Effekt vermieden werden. Sie ist eine Kombination aus Vorwärts- und Rückwärtsselektion. In jeder Iteration kann sowohl eine Variable aufgenommen, als auch entfernt werden. Um die relevanten Variablen auszuwählen, die ein möglichst gutes Modell widerspiegeln, kam die schrittweise Selektion im Rahmen dieser Arbeit zur Anwendung.

Das Akaike Informationskriterium (AIC) wird folgendermaßen berechnet:

$$AIC = -2 \cdot l(\hat{\beta}_M, \sigma^2) + 2p$$

wobei $l(\hat{\beta}_M, \sigma^2)$ der maximale Wert der Log-Likelihood ist. Der Term $2p$ „bestraft“ ein über-parametrisiertes Modell. Ein Modell ist umso besser, je kleiner der AIC-Wert ist. D.h., bei der Wahl zwischen verschiedenen Modellen wird dasjenige mit dem kleinsten AIC-Wert ausgewählt.

4.5 Kreuzvalidierung

Eine Alternative zur Modellbewertung ist die Methodik der k-fachen Kreuzvalidierung („Cross-Validation“ (CV)). Sie beruht grundsätzlich auf der Ermittlung des Prognosefehlers als erwartete quadratische Abweichung zwischen dem tatsächlichen Wert und dem Schätzwert des Regressionsmodells. Idealerweise erfolgt die Ermittlung des Prognosefehlers mit Hilfe neuer, unabhängiger Daten auf denen das geschätzte Modell angewandt wird. Da in den meisten Fällen keine zusätzlichen Daten zur Verfügung stehen, wird eine andere Methode herangezogen. Die Daten werden zufällig in Testdatensätze zur Schätzung der Modellparameter und einen Validierungsdatensatz zur Abschätzung der Prognosegüte unterteilt. Das weitere Vorgehen der k-fachen Kreuzvalidierung besteht dabei im Wesentlichen aus den folgenden Schritten (vgl. Efron und Tibshirani (1998, S. 240)):

- teile den Datensatz nach dem Zufallsprinzip in K möglichst gleich große Teil-datensätze,
- das mit den (K-1) Teilen geschätzte Modell wird auf den k-ten Teil des Testdatensatzes gefittet, indem der Prognosefehler als Abweichung zwischen dem Prognosewert und dem tatsächlichen Messwert $k(i)$ errechnet wird.
- Dieser Schritt wiederholt sich für alle $k = 1, 2, \dots, K$.

Der mittels Kreuzvalidierung geschätzte Wert des Prognosefehlers errechnet sich aus:

$$CV = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i^{-k(i)})^2$$

Neben dem mittleren quadratischen Prognosefehler (CV(SE)) können mit diesem Verfahren auch noch andere Kriterien, wie z.B. der absolute mittlere Prognosefehler (CV(L1)) basierend auf der L1-Norm ($\|x\|_1 = \sum_{i=1}^n |x_i|$) zur Beurteilung herangezogen werden. Das Modell mit dem jeweiligen kleinsten Wert bzw. Fehler wird bevorzugt. In den nachfolgenden Regressionsanalysen werden sowohl der mittlere quadratische Prognosefehler wie auch der absolute mittlere Prognosefehler als Modellwahlkriterien herangezogen.

4.6 *Potentielle Einflussgrößen/ Confounder und Datenbasis*

Als mögliche Einflussgrößen auf die Lungenfunktionsparameter „Einsekundenkapazität (FEV1)“, „forcierte Vitalkapazität (FVC)“ und „Tiffeneau-Wert (FEV1/ FVC)“ wurden die folgenden Variablen berücksichtigt (Haupteffektmodell):

- Dauer des aktiven Rauchens
- durchschnittliche Anzahl an Zigaretten pro Tag
- Passivrauchexposition (kumuliert)

Zudem wurden weitere potentielle Einflussgrößen/ Confounder in das Haupteffektmodell eingebunden:

- Lungenfunktionsparameter aus der ISAAC II-Studie
- Körpergröße
- Körpergewicht
- Geschlecht (männlich/ weiblich)
- Bildung (niedrig/ hoch)
- Erhebungszentrum (München/ Dresden)
- Asthmadiagnose der Probanden (nie/ einmal/ mehrmals)
- berufliches Asthmarisiko (keins/ niedriges/ hohes)
- Mutter Asthma (ja/ nein/ nicht bekannt)
- Vater Asthma (ja/ nein/ nicht bekannt)

Die Variable „Packungsjahre“ wurde als Interaktionsterm betrachtet. Sie entspricht der Interaktion zwischen den Variablen „Dauer des aktiven Rauchen“ und der „durchschnittlichen Anzahl an Zigaretten pro Tag“ mit dem Faktor 1/ 20.¹⁶ Des Weiteren wurden Interaktionen zwischen der (aktiven/ passiven) Rauchexposition und der Asthmadiagnose der Teilnehmer untersucht.

Die ursprünglichen kategorialen Variablen zur Passivrauchexposition (Frage 79) konnten nicht berücksichtigt werden, da sie untereinander hoch korreliert waren. In *Tabelle 16* sind diese Zusammenhänge dargestellt. Diese hohen Korrelationen können dazu führen, dass die Schätzungen der Regressionsparameter extrem unpräzise werden. Aus diesem Grund wurden die originären Informationen dieser Variablen aufsummiert (vgl. Abschnitt 2.3.2)

¹⁶ Die Berechnung der Variablen „Packungsjahre“ wurde bereits in Abschnitt 2.3.2 beschrieben.

und die kumulierte Passivrauchexposition an Stelle der vier Ausgangsvariablen in die Regressionsanalysen einbezogen.

	zu Hause	am Arbeitsplatz	in Bars etc.	andere Orte
zu Hause	1,000	0,963	0,955	0,964
am Arbeitsplatz	0,963	1,000	0,962	0,960
in Bars etc.	0,955	0,962	1,000	0,967
andere Orte	0,964	0,960	0,967	1,000

*Tabelle 16: Korrelationen unter den kategorialen Variablen zur Passivrauchexposition
(Korrelationskoeffizient nach Spearman)*

Hinsichtlich der nachträglichen Berechnung der kumulierten Passivrauchexposition bestehen aufgrund der im Erhebungsbogen der SOLAR II-Studie gewählten Frageformulierung verschiedene inhaltliche Bedenken, deren Auswirkungen auf die Regressionsanalysen nicht beurteilt werden können. Mit der **Frage 79** („Bitte beschreiben Sie dies genauer: Wie viele Stunden sind Sie pro Tag Tabakrauch von anderen an den folgenden Orten ausgesetzt?“) wird der Proband gebeten, die Aussetzung des Passivrauchs zu konkretisieren. Zu diesem Zweck gliedert sich die Hauptfrage in die Unterfragen „a) zu Hause“, „b) am Arbeitsplatz“, „c) in Bars, Restaurants, Kinos oder ähnlichem“ und „d) andere Orte“ auf. Der Proband soll zu diesen vier Unterfragen angeben, ob der Passivrauch weniger als eine Stunde pro Tag oder mehrere Stunden beträgt, wobei bei mehreren Stunden die genaue Stundenanzahl auszuweisen ist (Details s.a. Abschnitt 2.3.2 bzw. *Abbildung 8*).

Werden die vier Unterfragen übergreifend betrachtet, fallen zunächst die inhaltlichen Überschneidungen auf. Aus Perspektive der Probanden können die Kategorien „a) zu Hause“ und „b) am Arbeitsplatz“ z.B. für Heimarbeiter (Tele-Arbeitsplatz) oder Selbstständige identisch verstanden werden. Gleiches gilt für die Kategorien „b) am Arbeitsplatz“ und „c) in Bars, Restaurants, Kinos oder ähnlichen“, welche z.B. für Berufstätige der Gastronomie gleichbedeutend wären. Da den Probanden keine weitere Definition für derartige Überschneidungsfälle vorgegeben wird, ist unklar, welche Auswirkungen dies auf das Antwortverhalten genommen hat (z.B. Aufteilung der Stunden aufgrund einer scharfen Trennung

der Begriffe „zu Hause, am Arbeitsplatz, in Bars etc.“ oder einer mehrfachen Ausweisung der Stundenanzahl). Des Weiteren zeigt das Datenmaterial zur Passivrauchexposition, dass das Maximum der kumulierten Stundenangaben in einigen Fällen über 24 Stunden hinausgeht. Da der Hauptfrage nicht explizit zu entnehmen ist, dass die Summe über die vier Unterfragen maximal 24 Stunden ergeben kann, ist wiederum unklar, welche Gründe dafür sprechen, dass in einigen Fällen Mehrstunden ausgewiesen wurden. Handelt es sich um eine bewusste Doppelerfassung der Stundenangaben aufgrund der oben dargestellten Problematik der mehrdeutigen Antwortkategorien oder wurde die Fragestellung nicht aufmerksam und durchdacht gelesen?

Des Weiteren wurde die Zeitspanne in der Frage 79 nicht eindeutig definiert. Inhaltlich wird zwar auf die Aussetzung des Passivrauchs pro Tag verwiesen, allerdings kann sich diese Zeitspanne zu Hause bzw. in der Freizeit auf 7 x 24 Stunden und am Arbeitsplatz auf recht unterschiedliche Zeitspannen je nach Arbeitsmodell (z.B. Regelfall: 40 Stundenwoche mit 5 Arbeitstagen) beziehen.

Zudem ist den Regressionsanalysen vorzuschicken, dass sich die Datenbasis lediglich auf die Anzahl der kompletten Beobachtungsfälle (s.a. *Tabelle 17*) bezieht, d.h. der vollständigen Kombination von Beobachtungen aus der ISAAC II- und SOLAR II-Studie, bezogen auf die ausgewählten Variablen.

Datenbasis	n
Gesamt	523
Teildatensatz der Männer	211
Teildatensatz der Frauen	312

Tabelle 17: Aufteilung der Datenbasis („complete cases“)

Mit Hilfe der Zusammenhangsanalyse konnte bereits ein erster Überblick über die Art und Stärke des Zusammenhangs zwischen den Zielgrößen und den erklärenden Variablen gewonnen werden. In diesem Kontext konnten keine nicht-linearen Beziehungen identifiziert werden. Anzumerken ist allerdings, dass sich das dritte Kapitel lediglich auf eine reine bivariate Betrachtung beschränkt und simultane Effekte der Einflussgrößen mit dieser

Vorgehensweise nicht berücksichtigt werden können. Um festlegen zu können, ob nicht-lineare Terme bzgl. der potentiellen stetigen Einflussgrößen modelliert werden sollten, wurden vorab GAMs (generalisierte additive Modelle) gerechnet.

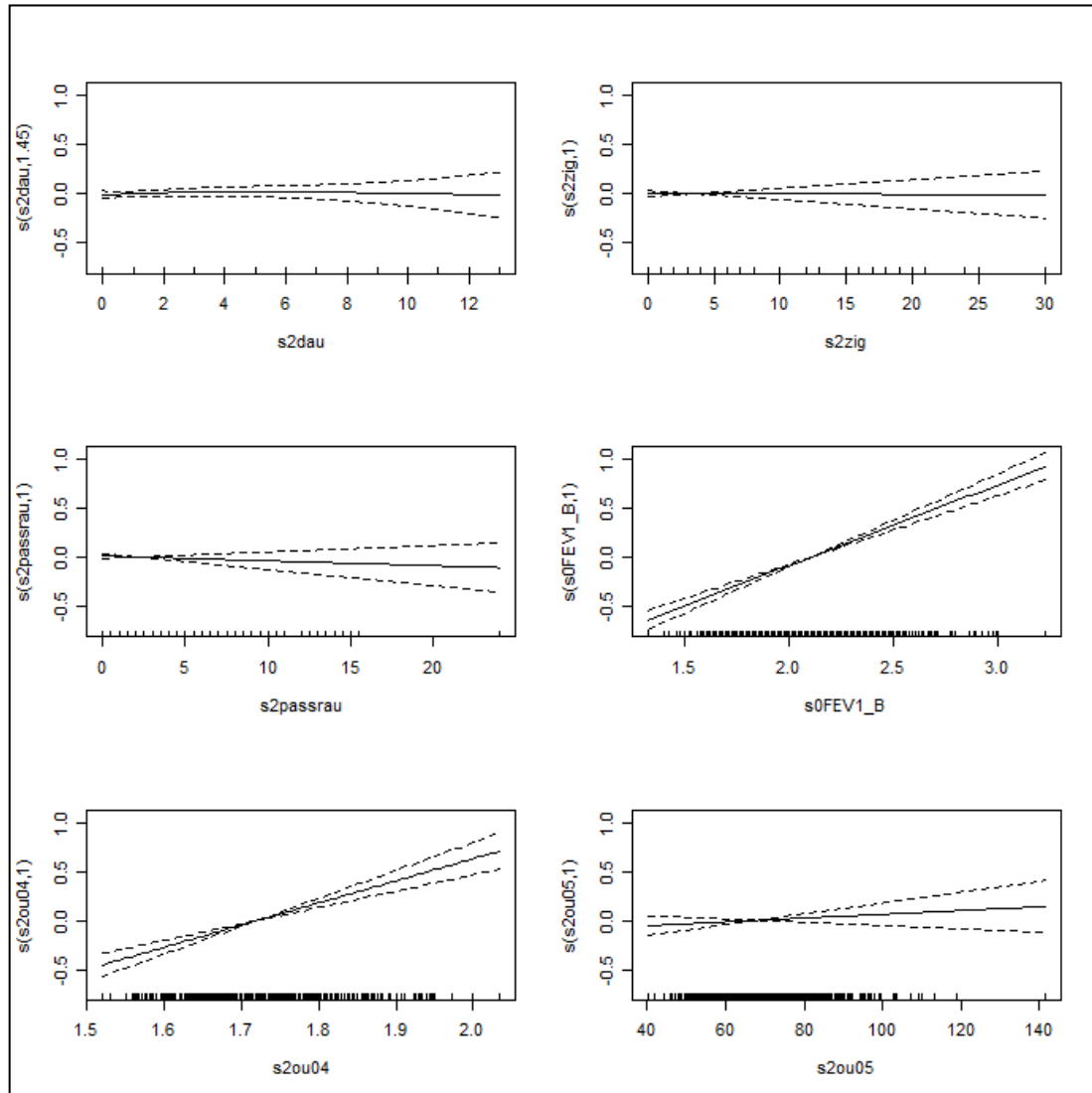


Abbildung 29: Geschätzte Funktionen für die stetigen Kovariablen

„Dauer des aktiven Rauchens“, „Anzahl der Zigaretten pro Tag“, „Passivrauchexposition (kumuliert)“, „FEV1 der ISAAC II-Studie“, „Körpergröße“ und „Körpergewicht“
im Haupteffektmodell für die Zielgröße „Einsekundenkapazität (FEV1)“ (n=523)

In *Abbildung 29* sind beispielhaft die geschätzten Funktionen für die stetigen Kovariablen im Haupteffektmodell für die Einsekundenkapazität (FEV1) auf Basis der kompletten Fälle (n = 523) dargestellt. Für alle Kovariablen zeigt sich ein linearer bzw. annähernd linearer Verlauf. Auch in allen weiteren Modellen wurden die metrischen Einflussgrößen linear modelliert.

4.7 Ergebnisse für die Zielgröße „Einsekundenkapazität“

4.7.1 Ergebnisse für die gesamte Datenbasis

In *Tabelle 18* ist das nach schrittweiser Variablenselektion verbleibende Modell für die Zielgröße „Einsekundenkapazität“ dargestellt. Die Variablen zur Aktiv- und Passivrauchexposition zeigten dabei keine Relevanz.

Variable	Koeffizient	Standard-abweichung	p-Wert	95% Konfidenz-intervall	
(Intercept)	-1,2661	0,4546	0,0056	-2,157	-0,375
FEV1 (ISAAC II)	0,8397	0,0596	< 0,001	0,723	0,957
Frauen (Referenz: Männer)	-0,7871	0,0504	< 0,001	-0,886	-0,688
Dresden (Referenz: München)	-0,3357	0,0343	< 0,001	-0,403	-0,268
Körpergröße	2,3514	0,2648	< 0,001	1,832	2,870
Einmalige Asthmadiagnose (Referenz: Keine Diagnose)	-0,1711	0,0715	0,0170	-0,311	-0,031
Mehrmalige Asthmadiagnose (Referenz: Keine Diagnose)	-0,2346	0,0927	0,0116	-0,416	-0,053
n = 523					
AIC: 495,21					

Tabelle 18: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell bezogen auf die gesamte Datenbasis und die Zielgröße „Einsekundenkapazität (FEV1)“

Die FEV1-Werte aus der ISAAC II-Studie waren deutlich positiv mit der Einsekundenkapazität assoziiert, ebenso wie die Körpergröße. Die geschätzten Parameter zeigten deutliche geschlechtsspezifische Unterschiede. Weibliche Probanden wiesen im Mittel geringere FEV1-Werte gegenüber den männlichen Probanden auf. Auch bei den Teilnehmern aus Dresden wurden im Durchschnitt geringere FEV1-Werte gegenüber den Teilnehmern aus München festgestellt. Eine vorherige Asthmadiagnose der Teilnehmer wirkte sich dabei negativ auf den Lungenfunktionsparameter aus. Je häufiger eine Asthmadiagnose gestellt wurde, umso geringer fielen die FEV1-Werte aus. Der Unterschied zwischen einer einmaligen Asthmadiagnose und einer nie festgestellten Erkrankung fällt um 0,17 Liter geringer aus. Bei mehrmaliger Asthmadiagnose war der Unterschied mit -0,23 Liter noch gravierender.

Beide gewählten Modellwahlverfahren (d.h. nach der schrittweisen Selektion und nach dem Verfahren zur 5-fachen Kreuzvalidierung) brachten das gleiche Ergebnis. Selektiert wurde zunächst nach dem Akaike Informationskriterium (AIC) aus dem vollen Modell mit den Interaktionstermen. Gemäß *Tabelle 19* wies das reduzierte Modell nach der schrittweisen Selektion den geringsten AIC-Wert auf. Durch eine anschließende Überprüfung mit dem Verfahren zur Kreuzvalidierung wurde die Modellwahl bestätigt. Das aufgeführte Modell wies demnach den geringsten mittleren quadratischen Prognosefehler (CV (SE)) bzw. den geringeren mittleren absoluten Prognosefehler (CV (L1)) gegenüber dem Haupteffekt- und Interaktionsmodell auf.

Modelle (n = 523)	AIC	5-fache Kreuzvalidierung	
		CV (SE)	CV (L1)
Interaktionsmodell	512,79	0,155430500	0,293711000
Haupteffektmodell	510,43	0,155251700	0,290543300
Schrittweise Selektion	495,21	0,149894200	0,282496100

Tabelle 19: Gegenüberstellung der Modellwahlkriterien

4.7.2 Ergebnisse für die männlichen Teilnehmer

Aus substanziell wissenschaftlichen Überlegungen und basierend auf dem Ergebnis aus Kapitel 4.7.1, wo geschlechtsspezifische Unterschiede festgestellt wurden, erfolgte zusätzlich eine Analyse stratifiziert nach dem Geschlecht. Hier war das Ziel herauszufinden, ob die aktive bzw. passive Rauchexposition unterschiedliche Wirkungen auf die Lungenfunktion bei männlichen bzw. weiblichen Probanden zeigte. Das nach dem AIC-Kriterium „beste“ Modell für die männlichen Teilnehmer ist in *Tabelle 20* abgebildet. In diesem verblieben drei Variablen. Die FEV1-Werte aus der Vorgängerstudie und die Körpergröße standen in einem deutlich positiven Zusammenhang mit der Einsekundenkapazität der männlichen Teilnehmer. Die Gruppe aus Dresden wies gegenüber den männlichen Teilnehmern aus München geringere Werte bzgl. der Einsekundenkapazität auf. Die erhobenen und berechneten Variablen zur Rauchexposition zeigten bei dieser Analyse ebenfalls keine Relevanz.

Multiple lineare Regression

Variable	Koeffizient	Standard-abweichung	p-Wert	95% Konfidenz-intervall	
(Intercept)	-1,0661	0,7449	0,1540	-2,526	0,394
FEV1 (ISAAC II)	1,0526	0,1125	< 0,001	0,832	1,273
Dresden (Referenz: München)	-0,4454	0,0633	< 0,001	-0,569	-0,321
Körpergröße	2,0067	0,4493	< 0,001	1,126	2,887
n = 211					
AIC: 264,99					

Tabelle 20: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell bezogen auf die männlichen Teilnehmer und die Zielgröße „Einsekundenkapazität (FEV1)“

Gemäß *Tabelle 21* lieferte das Modell nach schrittweiser Selektion ein deutlich besseres Ergebnis bzgl. des AIC-Kriteriums. Während sich das Haupteffektmodell und das Interaktionsmodell nur geringfügig unterschieden. Auch bei der 5-fachen-Kreuzvalidierung zeigte das Modell nach der schrittweisen Selektion die bessere Modellanpassung. Auffallend hier ist ein hoher mittlerer quadratischer Prognosefehler beim Interaktionsmodell im Vergleich zu beiden anderen Modellen.

Modelle (n = 211)	AIC	5-fache Kreuzvalidierung	
		CV (SE)	CV (L1)
Interaktionsmodell	283,06	0,725327200	0,433400200
Haupteffektmodell	281,68	0,223132400	0,353620700
Schrittweise Selektion	264,99	0,206552500	0,340462300

Tabelle 21: Gegenüberstellung der Modellwahlkriterien

4.7.3 Ergebnisse für die weiblichen Teilnehmer

Bei der Analyse der weiblichen Teilnehmer wurde bei der Variablenselektion ein Modell mit vier Variablen ausgewählt (s.a. *Tabelle 22*). Wie bei den männlichen Probanden scheint die Rauchexposition bei den weiblichen Probanden ebenfalls keine oder nur eine geringe Einflussnahme auf die Einsekundenkapazität auszuüben. Die jeweiligen Variablen sind nicht im Modell verblieben. Dagegen wirkt sich eine Asthmadiagnose negativ auf die FEV1-Werte der weiblichen Probanden aus. Im Vergleich zu Teilnehmerinnen, bei denen

noch nie Asthma diagnostiziert wurde, wiesen die weiblichen Probanden mit einer einmaligen Asthmadiagnose um 0,1719 Liter bzw. bei Teilnehmerinnen mit mehrmaliger Diagnose um 0,2638 Liter geringere FEV1-Werte auf. Die FEV1-Werte aus der Vorgängerstudie und die Körpergröße haben auch bei den Frauen einen stark positiven Effekt auf die Einsekundenkapazität. Zudem spielt der Erhebungsort bei den Frauen eine Rolle. Die weiblichen Teilnehmer aus Dresden hatten im Vergleich zu den Frauen aus München geringere FEV1-Werte.

Variable	Koeffizient	Standard-abweichung	p-Wert	95% Konfidenz-intervall	
(Intercept)	-2,0065	0,5187	0,0001	-3,023	-0,990
FEV1 (ISAAC II)	0,7222	0,0659	< 0,001	0,593	0,851
Dresden (Referenz: München)	-0,2691	0,0385	< 0,001	-0,345	-0,194
Körpergröße	2,4483	0,3213	< 0,001	1,819	3,078
Einmalige Asthmadiagnose (Referenz: Keine Diagnose)	-0,1791	0,0816	0,0289	-0,339	-0,019
Mehrmalige Asthmadiagnose (Referenz: Keine Diagnose)	-0,2638	0,1136	0,0208	-0,486	-0,041
n = 312					
AIC: 209,39					

*Tabelle 22: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell
bezogen auf die weiblichen Teilnehmer und die Zielgröße „Einsekundenkapazität (FEV1)“*

Die Modellwahl erfolgte auch hier zunächst nach dem AIC-Kriterium. In *Tabelle 23* sind die Ergebnisse abgebildet.

Modelle (n = 312)	AIC	5-fache Kreuzvalidierung	
		CV (SE)	CV (L1)
Interaktionsmodell	234,36	0,125507600	0,257933900
Haupteffektmodell	224,89	0,117366100	0,246471000
Schrittweise Selektion	209,39	0,114951800	0,242499100

Tabelle 23: Gegenüberstellung der Modellwahlkriterien

Das Modell nach der schrittweisen Selektion zeigte die beste Datenanpassung. Durch die 5-fache Kreuzvalidierung wurde die Wahl wiederum überprüft und bestätigt. Allerdings fallen die Unterschiede zwischen dem Haupteffektmodell und dem selektierten Modell nur sehr geringfügig aus.

4.8 Ergebnisse für die Zielgröße „forcierte Vitalkapazität“

4.8.1 Ergebnisse für die gesamte Datenbasis

Die Analyse für die Zielgröße „forcierte Vitalkapazität“ erfolgte zunächst wieder auf Basis der Gesamtdaten ($n = 523$). Das Ergebnis der Modellwahl mit dem Verfahren der schrittweisen Selektion ist aus *Tabelle 24* ersichtlich. Die FVC-Werte aus der ISAAC II-Studie und die Körpergröße sowie das Körpergewicht haben einen positiven Einfluss auf die forcierte Vitalkapazität. Ebenfalls zeigte sich hier wieder der geschlechtsspezifische Effekt. D.h., Frauen weisen gegenüber Männern wesentlich geringere FVC-Werte auf. Auch die Unterschiede bezüglich des Erhebungszentrums treten bei der Analyse der forcierten Vitalkapazität erneut auf. Die Parameterschätzungen bzgl. der Aktiv- und Passivrauchexposition waren hier ebenfalls nicht bedeutend. Außerdem scheinen die asthmarelevanten Variablen keinen oder nur einen geringen Einfluss auf die forcierte Vitalkapazität zu nehmen.

Variable	Koeffizient	Standard-abweichung	p-Wert	95% Konfidenz-intervall	
(Intercept)	-2,1901	0,5317	< 0,001	-3,232	-1,148
FVC (ISAAC II)	0,7838	0,0627	< 0,001	0,661	0,907
Frauen (Referenz: Männer)	-0,9245	0,0575	< 0,001	-1,037	-0,812
Dresden (Referenz: München)	-0,1674	0,0386	< 0,001	-0,243	-0,092
Körpergröße	2,9313	0,3316	< 0,001	2,281	3,581
Körpergewicht	0,0074	0,0020	< 0,001	0,003	0,011
n = 523					
AIC: 628,82					

Tabelle 24: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell bezogen auf die gesamte Datenbasis und die Zielgröße „forcierte Vitalkapazität (FVC)“

Um die relevanten Variablen auszuwählen, wurde wiederum das Modellwahlkriterium nach Akaike herangezogen. Die schrittweise Selektion lieferte den kleinsten Wert, wurde also als „bestes“ Modell gewählt. Durch Berechnung des mittleren quadratischen

Prognosefehlers und des mittleren absoluten Prognosefehlers mit dem Verfahren der Kreuzvalidierung wurde die Wahl erneut überprüft. Das gewählte Modell zeigte hier den kleinsten Fehler bzw. die beste Modellanpassung (vgl. *Tabelle 25*).

Modelle (n = 523)	AIC	5-fache Kreuzvalidierung	
		CV (SE)	CV (L1)
Interaktionsmodell	652,68	0,211370100	0,351066900
Haupteffektmodell	644,21	0,203945700	0,343411300
Schrittweise Selektion	628,82	0,194574000	0,331306500

Tabelle 25: Gegenüberstellung der Modellwahlkriterien

4.8.2 Ergebnisse für die männlichen Teilnehmer

Für die Zielgröße der „forcierten Vitalkapazität“ wurden ebenfalls Regressionsanalysen stratifiziert nach dem Geschlecht vorgenommen. Das resultierende bzw. nach AIC-Kriterium gewählte Modell ist in *Tabelle 26* ersichtlich.

Variable	Koeffizient	Standard- abweichung	p-Wert	95% Konfidenz- intervall	
(Intercept)	-2,5736	0,9130	0,0053	-4,363	-0,784
FVC (ISAAC II)	0,9439	0,1225	< 0,001	0,704	1,184
Dresden (Referenz: München)	-0,1815	0,0739	0,0149	-0,326	-0,037
Körpergröße	2,9626	0,5889	< 0,001	1,808	4,117
Körpergewicht	0,0078	0,0035	0,0265	0,001	0,015
Bildung eher hoch (Referenz: Bildung eher niedrig)	-0,1213	0,0811	0,1363	-0,280	0,038
n = 211					
AIC: 336,56					

*Tabelle 26: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell
bezogen auf die männlichen Teilnehmer und die Zielgröße „forcierte Vitalkapazität (FVC)“*

Die Variablen zur forcierten Vitalkapazität aus der ISAAC II-Studie, die Körpergröße, das Körpergewicht und das Erhebungszentrum, bleiben als Einflussgrößen im Modell enthalten. Verblieben ist ebenfalls die Variable „Bildung“. Allerdings scheint der Bildungsstand

keine große Rolle zu spielen, da der Effekt nicht signifikant ist. Die Rauchexposition scheint bei den männlichen Teilnehmern keinen bzw. nur einen geringen Einfluss auf die forcierte Vitalkapazität zu haben. Die Variablen, die in Bezug zum Asthma erhoben wurden, sind nicht mehr im gewählten Modell enthalten.

Die Entscheidung nach dem Modellwahlkriterium „AIC“ lieferte das Modell nach der schrittweisen Selektion mit dem kleinsten Wert. Die Unterschiede zwischen dem Interaktions-, dem Haupteffekt- und dem Selektionsmodell sind nicht gravierend. Die Bewertung mit dem Verfahren der 5-fachen Kreuzvalidierung zeigte gemäß *Tabelle 27* das gleiche Ergebnis, allerdings fällt ein besonders hoher mittlerer quadratischer Prognosefehler für das Interaktionsmodell auf.

Modelle (n = 211)	AIC	5-fache Kreuzvalidierung	
		CV (SE)	CV (L1)
Interaktionsmodell	354,49	1,413861700	0,526634400
Haupteffektmodell	349,85	0,314262000	0,423918500
Schrittweise Selektion	336,56	0,295696800	0,411189700

Tabelle 27: Gegenüberstellung der Modellwahlkriterien

4.8.3 Ergebnisse für die weiblichen Teilnehmer

Das nach dem AIC-Kriterium gewählte Modell für die Zielgröße der „forcierten Vitalkapazität“ unter Einbeziehung der Daten für die weiblichen Teilnehmer zeigt *Tabelle 28*. Das Passivrauchen wirkte sich negativ auf die FVC-Werte der Teilnehmerinnen aus. Allerdings ist der Effekt der Passivrauch-Tagesexposition nicht sehr groß. Ein positiver Zusammenhang besteht zwischen den FVC-Werten aus der ISAAC II-Studie, der Körpergröße, dem Körpergewicht und der forcierten Vitalkapazität der weiblichen Probanden. Die Teilnehmer aus Dresden wiesen im Vergleich zu den weiblichen Probanden aus München niedrigere Werte auf. Frauen mit hohem beruflichem Asthmarisiko hatten im Vergleich zu Probanden, welche keinem beruflichem Asthmarisiko ausgesetzt waren, höhere Werte.

Multiple lineare Regression

Variable	Koeffizient	Standard-abweichung	p-Wert	95% Konfidenz-intervall	
(Intercept)	-2,6005	0,5745	< 0,001	-3,726	-1,475
FVC (ISAAC II)	0,6845	0,0659	< 0,001	0,555	0,814
Passivrauchen kumuliert	-0,0140	0,0062	0,0235	-0,026	-0,002
Dresden (Referenz: München)	-0,1806	0,0412	< 0,001	-0,261	-0,100
Körpergröße	2,7722	0,3790	< 0,001	2,029	3,515
Körpergewicht	0,0071	0,0023	0,0021	0,003	0,012
geringes berufliches Asthmarisiko (Referenz: kein Risiko)	0,0439	0,0583	0,4518	-0,070	0,158
hohes berufliches Asthmarisiko (Referenz: kein Risiko)	0,0995	0,0491	0,0436	0,003	0,196
n = 312 AIC: 256,18					

Tabelle 28: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell bezogen auf die weiblichen Teilnehmer und die Zielgröße „forcierte Vitalkapazität (FVC)“

Sowohl nach dem AIC-Kriterium, als auch nach dem Verfahren der 5-fachen-Kreuzvalidierung zeigt das Modell nach Variablenselektion die beste Datenanpassung (vgl. *Tabelle 29*).

Modelle (n = 312)	AIC	5-fache Kreuzvalidierung	
		CV (SE)	CV (L1)
Interaktionsmodell	276,8	0,156989000	0,298545400
Haupteffektmodell	268,58	0,135365400	0,283165500
Schrittweise Selektion	256,18	0,129028500	0,276685000

Tabelle 29: Gegenüberstellung der Modellwahlkriterien

4.9 Ergebnisse für die Zielgröße „Tiffeneau-Wert“

4.9.1 Ergebnisse für die gesamte Datenbasis

Die Regressionsanalysen für die Zielvariable der „relativen Einsekundenkapazität“ (Tiffeneau-Wert) wurden entsprechend den vorherigen Beschreibungen durchgeführt. In dem resultierenden Modell nach der schrittweisen Selektion auf Basis der Gesamtdaten

sind insgesamt vier Variablen verblieben. Ein positiver Zusammenhang war zwischen den Tiffeneau-Werten aus der Vorgängerstudie und der relativen Einsekundenkapazität zu verzeichnen. Ein negativer Effekt ging vom Körpergewicht aus. Ebenfalls negativ wirkte sich eine Asthmaerkrankung der Teilnehmer auf den Tiffeneau-Wert aus. Je häufiger eine Asthmad Diagnose ausgestellt wurde, desto stärker ist der Effekt im Vergleich zu den Probanden, bei denen noch nie eine Diagnose gestellt wurde. Der Unterschied zwischen den Erhebungszentren war zudem vorhanden. Die Variablen, die zu Aktiv- und Passivrauchexposition erhoben wurden, zeigten sich bei der Modellwahl im Sinne des AIC-Kriteriums als nicht relevant (s.a. *Tabelle 30*).

Variable	Koeffizient	Standardabweichung	p-Wert	95% Konfidenzintervall	
(Intercept)	0,3722	0,0417	< 0,001	0,290	0,454
FEV1/ FVC (ISAAC II)	0,6370	0,0423	< 0,001	0,554	0,720
Dresden (Referenz: München)	-0,0441	0,0049	< 0,001	-0,054	-0,034
Körpergewicht	-0,0009	0,0002	< 0,001	-0,001	-0,001
Einmalige Asthmad Diagnose (Referenz: Keine Diagnose)	-0,0237	0,0102	0,0205	-0,044	-0,004
Mehrmalige Asthmad Diagnose (Referenz: Keine Diagnose)	-0,0329	0,0133	0,0140	-0,059	-0,007
n = 523					
AIC: -1548,2					

Tabelle 30: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell bezogen auf die gesamte Datenbasis und die Zielgröße „Tiffeneau-Wert (FEV1/ FVC)“

Gemäß *Tabelle 31* besitzt das Modell nach der schrittweisen Selektion wiederum den kleinsten AIC-Wert.

Modelle (n = 523)	AIC	5-fache Kreuzvalidierung	
		CV (SE)	CV (L1)
Interaktionsmodell	-1.522,10	0,003294592	0,044486180
Haupteffektmodell	-1.530,50	0,003153370	0,043686600
Schrittweise Selektion	-1.548,20	0,003041582	0,042880786

Tabelle 31: Gegenüberstellung der Modellwahlkriterien

Den schlechtesten Wert liefert das volle Modell mit den Interaktionstermen. Auch mit der Methode der Kreuzvalidierung wird die getroffene Modellwahl untermauert. Das selektierte Modell besitzt den kleinsten mittleren bzw. absoluten Prognosefehler. Die Unterschiede sind allerdings geringfügig.

4.9.2 Ergebnisse für die männlichen Teilnehmer

Um eine mögliche differenzierte Auswirkung der Einflussgrößen auf die Zielgröße der „relativen Einsekundenkapazität“ bzgl. des Geschlechts identifizieren zu können, wurden zunächst wieder die Daten der männlichen Probanden betrachtet. Nach der schrittweisen Selektion resultierte ein Modell mit den Einflussgrößen Tiffeneau-Wert der ISAAC II-Studie, Körpergewicht, Studienzentrum sowie Asthmadiagnose der Probanden. Die Koeffizientenschätzung für die Kategorie „mehrmalige Asthmadiagnose“ war hier allerdings nicht signifikant. Eine bedeutende Auswirkung des Rauchens bzw. Passivrauchens auf die relative Einsekundenkapazität konnte bei den männlichen Probanden nicht festgestellt werden (s.a. *Tabelle 32*).

Variable	Koeffizient	Standard-abweichung	p-Wert	95% Konfidenzintervall	
(Intercept)	0,4012	0,0729	< 0,001	0,258	0,544
FEV1/ FVC (ISAAC II)	0,6053	0,0731	< 0,001	0,462	0,749
Dresden (Referenz: München)	-0,0512	0,0003	< 0,001	-0,068	-0,034
Körpergewicht	-0,0009	0,0003	0,0107	-0,002	0,000
Einmalige Asthmadiagnose (Referenz: Keine Diagnose)	-0,0369	0,0178	0,0396	-0,072	-0,002
Mehrmalige Asthmadiagnose (Referenz: Keine Diagnose)	-0,0333	0,0218	0,1278	-0,076	0,009
n = 211 AIC: -568,32					

Tabelle 32: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell bezogen auf die männlichen Teilnehmer und die Zielgröße „Tiffeneau-Wert (FEV1/ FVC)“

Das gewählte Modell zeigt entsprechend der *Tabelle 33* ebenfalls die beste Prognosegüte.

Multiple lineare Regression

Modelle (n = 211)	AIC	5-fache Kreuzvalidierung	
		CV (SE)	CV (L1)
Interaktionsmodell	-543,49	0,005468378	0,055284201
Haupteffektmodell	-550,90	0,004520480	0,052072030
Schrittweise Selektion	-568,32	0,003955358	0,048183480

Tabelle 33: Gegenüberstellung der Modellwahlkriterien

4.9.3 Ergebnisse für die weiblichen Teilnehmer

Bei der Betrachtung der Daten zu den weiblichen Teilnehmern resultierte aus den Regressionsanalysen ein etwas komplexeres Modell. Wiederum verblieben nach der schrittweisen Selektion die Variablen zur relativen Einsekundenkapazität aus der ISAAC II-Studie, das Körpergewicht und das Erhebungszentrum (s.a. *Tabelle 34*).

Variable	Koeffizient	Standard- abweichung	p-Wert	95% Konfidenz- intervall	
(Intercept)	0,3410	0,0501	0,0000	0,243	0,439
FEV1/ FVC (ISAAC II)	0,6639	0,0512	< 2e-16	0,564	0,764
Dresden (Referenz: München)	-0,0388	0,0059	0,0000	-0,050	-0,027
Körpergewicht	-0,0008	0,0003	0,0025	-0,001	0,000
geringes berufliches Asthmarisiko (Referenz: kein Risiko)	-0,0045	0,0080	0,5766	-0,020	0,011
hohes berufliches Asthmarisiko (Referenz: kein Risiko)	-0,0169	0,0067	0,0122	-0,030	-0,004
Einmalige Asthmadiagnose (Referenz: Keine Diagnose)	-0,0109	0,0121	0,3673	-0,035	0,013
Mehrmalige Asthmadiagnose (Referenz: Keine Diagnose)	-0,0377	0,0168	0,0254	-0,071	-0,005
Passivrauchen kumuliert	0,0016	0,0008	0,0661	0,000	0,003
n = 312					
AIC: -985,73					

*Tabelle 34: Schätzergebnisse für das gemäß AIC-Kriterium „beste“ Modell
bezogen auf die weiblichen Teilnehmer und die Zielgröße „Tiffeneau-Wert (FEV1/ FVC)“*

Ein hohes berufliches Asthmarisiko im Vergleich zu keinem beruflichen Asthmarisiko hatte hier eine negative Auswirkung auf die relative Einsekundenkapazität. Negativ wirkte sich des Weiteren eine mehrmalige Asthmadignose aus (im Vergleich zu keiner Asthmadignose). Die Passivrauchexposition ist ebenfalls im Modell verblieben. Allerdings war die Schätzung nicht signifikant.

Die Gegenüberstellung der berechneten Modellwahlkriterien wird in *Tabelle 35* veranschaulicht. Das Interaktionsmodell zeigt wiederum die schlechteste Anpassung, während das gewählte Modell nach dem Kriterium des AIC- und der Kreuzvalidierung die „beste“ Anpassung an die Daten liefert.

Modelle (n = 312)	AIC	5-fache Kreuzvalidierung	
		CV (SE)	CV (L1)
Interaktionsmodell	-958,67	0,003555852	0,044021028
Haupteffektmodell	-972,05	0,002740677	0,040778252
Schrittweise Selektion	-985,73	0,002626553	0,039827699

Tabelle 35: Gegenüberstellung der Modellwahlkriterien

4.10 Überprüfung der Modellannahmen

Die Normalverteilungsannahme wurde anhand von Normal-Quantil-Quantil-Diagrammen graphisch überprüft, wobei die empirischen Quantile gegen die theoretischen Quantile der Verteilung abgetragen wurden.

Die gewählten Modelle für die Zielgröße der „relativen Einsekundenkapazität“ (Tiffeneau-Wert bzw. FEV1/ FVC) zeigten zum Teil Abweichungen von der Normalverteilungsannahme im Bereich negativer Residuen (vgl. *Abbildung 30*). Hier traten die Abweichungen bei der Gruppe der männlichen Probanden am deutlichsten hervor. Diesbezüglich stellte sich nun die Frage, ob die Auswirkungen der Einflussgrößen über die gesamte Verteilung der Werte stabil sind bzw. die Ergebnisse der linearen multiplen Regression als robust einzustufen sind. Im nachfolgenden Kapitel wird dieser Fragestellung mit der Methode der Quantilregression nachgegangen.

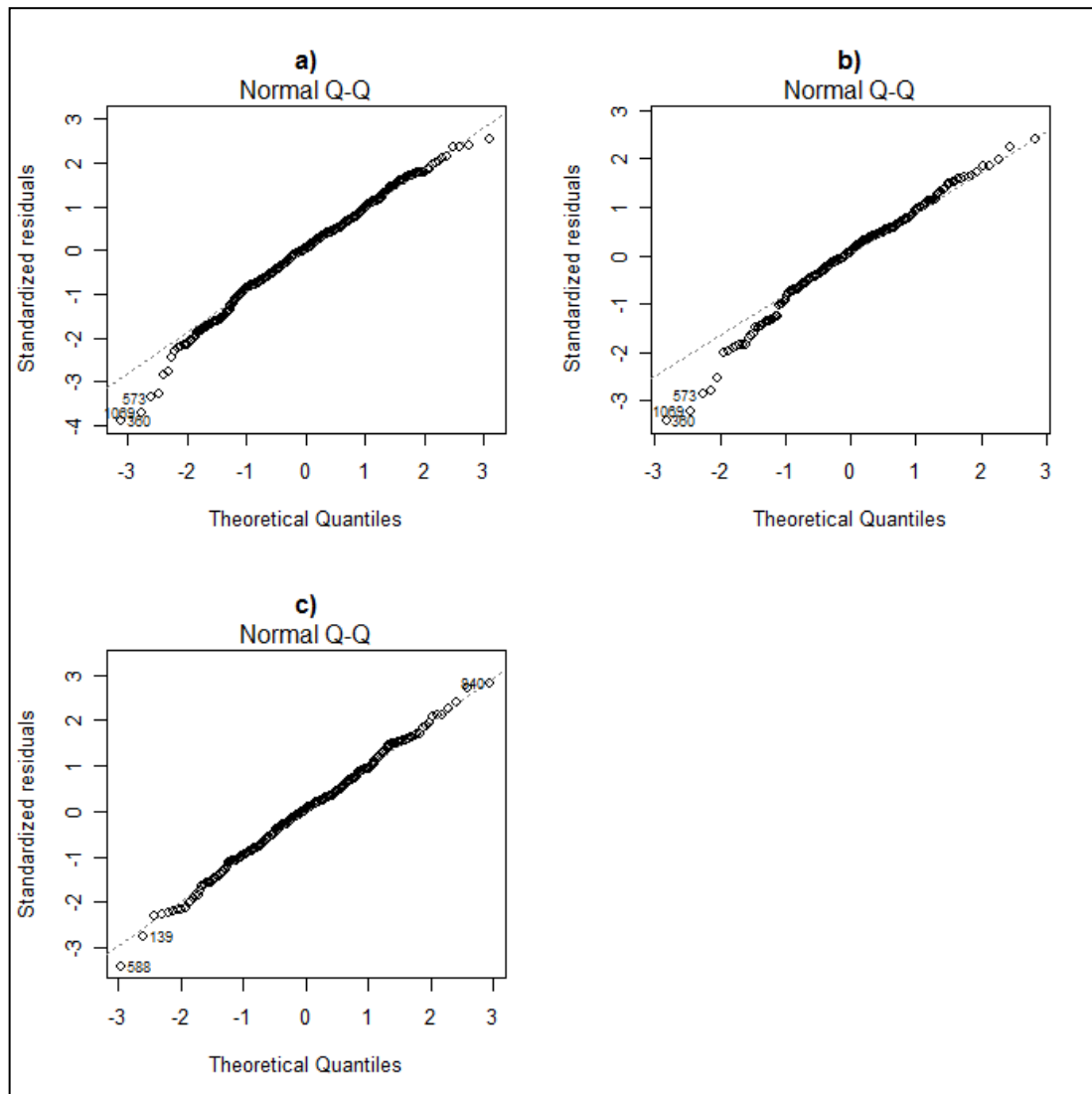


Abbildung 30: Quantil-Quantil Diagramme
für die gewählten Modelle bzgl. der Zielgröße „Tiffeneau-Wert“
bezogen auf a) die gesamte Datenbasis ($n = 523$) b) die männlichen Teilnehmer ($n = 211$)
und c) die weiblichen Teilnehmer ($n = 312$)

5. *Quantilregression*

Im Rahmen der multiplen linearen Regressionsanalyse wurden für die Zielgröße „relative Einsekundenkapazität (FEV1/ FVC)“ Abweichungen von der Normalverteilung im Bereich negativer Residuen beobachtet. Eine Quantilregression (QR) ermöglicht einen genaueren Einblick in die Lungenfunktionskapazität der Studienteilnehmer. Sie zeigt auf, ob und wie Kovariablen in verschiedenen Bereichen der Verteilung der Lungenfunktionsparameter unterschiedlichen Einfluss haben. Dabei kommt das Verfahren ohne die Annahme normalverteilter Residuen aus. Durch die Nutzung der Quantilregression kann man, falls die Störterme nicht normalverteilt sind, eine substanziell bessere Schätzung erreichen als mit der kleinsten Quadrate Schätzung (OLS). Im Gegensatz zur klassischen linearen Regression, wo nur der bedingte Erwartungswert charakterisiert wird, lässt sich mit Hilfe der Quantilregression die gesamte bedingte Verteilung modellieren.

5.1 *Modell und Schätzverfahren*

Das Modell der linearen Quantilregression¹⁷ lautet:

$$y_i = x_i' \beta_\tau + \varepsilon_{\tau i}$$

mit unabhängigen, aber möglicherweise heteroskedastischen $\varepsilon_{\tau i}$.

Die einzige Forderung an die $\varepsilon_{\tau i}$ ist, dass unabhängig von der Fehlerverteilung das τ -te Quantil gleich 0 angenommen wird:

$$\int_{-\infty}^0 f_{\varepsilon_\tau}(\varepsilon_\tau) d\varepsilon_\tau = F_{\varepsilon_\tau}(0) = \tau$$

¹⁷ Für eine ausführliche Darstellung wird auf Koenker (2005) verwiesen.

Das heißt

$$F_{\varepsilon_\tau}^{-1}(\tau) = 0.$$

Das τ -te Quantil der bedingten Verteilungsfunktion von y_i gegeben x_i ist dann:

$$Q_\tau(y_i|X = x_i) = x_i' \beta_\tau + F_{\varepsilon_\tau}^{-1}(\tau) = x_i' \beta_\tau$$

Einen Schätzer $\hat{\beta}_\tau$ für β_τ erhält man als Lösung des Minimierungsproblems

$$\min_{\beta_\tau} \sum_{i=1}^n \rho_\tau(y_i - x_i' \beta_\tau)$$

($0 < \tau < 1$) mit der so genannten Check-Funktion,

$$\rho_\tau(u) = u(\tau - I(u < 0)) = \begin{cases} |u| = \tau u & , u \geq 0 \\ (1 - \tau)|u| = (\tau - 1)u & , u < 0 \end{cases}$$

welche positiven Residuen das Gewicht τ zuweist, während die Absolutwerte negativer Residuen mit dem Gewicht $(1 - \tau)$ eingehen. Der Schätzer $\hat{\beta}_\tau$ ist konsistent und asymptotisch normalverteilt. Die Schätzung der Varianz-Kovarianz des QR-Schätzers erfolgt im Fall von unabhängigen, aber nicht gleichverteilten Fehlern mit dem so genannten „Huber-Sandwich“:

$$\sqrt{n}(\hat{\beta}_\tau - \beta_\tau) \rightarrow N(0, \tau(1 - \tau) \underbrace{H^{-1}(\tau)J(\tau)H^{-1}(\tau)}_{\text{„Huber Sandwich“}})$$

mit

$$J(\tau) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i x_i'$$

$$H(\tau) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i x_i' \cdot f_i(y_{i\tau})$$

Dabei ist $f_i(y_{i\tau})$ die bedingte Dichte von y_i an der Stelle $y_{i\tau}$.

Als weiteres Verfahren zur Schätzung der Varianz-Kovarianz-Matrix des QR-Schätzers wird u.a. das Bootstrapping vorgeschlagen.¹⁸ Die Standardfehler sind dabei allerdings nicht

¹⁸ Weitere Details zum Bootstrapping stellen z.B. Efron, Tibshirani (1998) bereit.

exakt bestimmbar, da sie anhand von zufälligen Teilstichproben, die aus den Daten durch „Ziehen mit Zurücklegen“ generiert werden, berechnet werden.

Nachfolgend werden die Ergebnisse der Quantilregression bezogen auf die Zielgröße „relative Einsekundenkapazität (FEV1/ FVC bzw. Tiffeneau-Wert)“ eingehender beschrieben.

5.2 *Ergebnisse der Quantilregression*

Um auch die möglichen Effekte der verschiedenen Rauchvariablen zu untersuchen, erfolgte die Quantilregression auf Basis des Haupteffektmodells und nicht für das im Sinne des AIC-Kriteriums gewählte „beste“ Modell. Die Berechnungen erfolgten wieder jeweils auf Basis der Gesamtdaten ($n = 523$) sowie für die männlichen und weiblichen Teilnehmer ($n = 211$ bzw. $n = 312$).

Es wurden jeweils mit drei Quantilen, $\tau = 0,25$, $\tau = 0,50$ und $\tau = 0,75$, simultane Quantilregressionen geschätzt. Wobei die Standardfehler auf Basis des „Huber-Sandwich-Schätzers“ ermittelt wurden. Zum Vergleich wurden ebenfalls OLS-Schätzungen durchgeführt. Die Ergebnisse der OLS-Schätzung sollten nicht direkt mit den Ergebnissen der Quantilregression verglichen werden, da sich die beiden Schätzfunktionen unterscheiden. In den *Tabellen 36 bis 38* sind die Ergebnisse für die beiden Regressionsmethoden abgetragen. Dargestellt sind jeweils immer die Koeffizienten, die berechneten Standardfehler sowie die p-Werte für die 25%-, 50%- und 75%-Quantile. In Fettdruck sind diejenigen Koeffizientenschätzungen hervorgehoben, die auf dem 5%-Niveau statistisch signifikant von Null verschieden sind.

Zunächst werden Gemeinsamkeiten aller Tabellen bzw. aller drei Gruppen herausgearbeitet, um anschließend auf die Besonderheiten der einzelnen Datenbasen einzugehen. Bei der übergreifenden Betrachtung der drei Tabellen fällt auf, dass die Dauer des aktiven Rauchens, die durchschnittliche Anzahl an Zigaretten pro Tag, die Körpergröße, die Bildung und die väterliche Asthmaerkrankung über alle Quantile hinweg keine Bedeutung haben, wohingegen zwei Variablen, der Tiffeneau-Wert der ISAAC II-Studie und das Erhebungszentrum, über alle Quantile hinweg hochsignifikante Ergebnisse liefern. Dies entspricht auch dem Ergebnis der OLS-Schätzung.

Anhand von *Tabelle 36* wird nun zunächst auf das Ergebnis der Berechnungen für die relative Einsekundenkapazität, bezogen auf die gesamte Datenbasis ($n = 523$) eingegangen. Die Variablen, die zur Aktiv- und Passivrauchexposition erhoben/ berechnet wurden, sind

über alle Quantile hinweg deutlich nicht signifikant. Der Tiffeneau-Wert, das Körpergewicht und das Erhebungszentrum sind dahingegen über alle drei Quantile signifikant. Der positive Effekt der FEV1/ FVC-Werte aus der ISAAC II-Studie ist an den Rändern der Verteilung der Tiffeneau-Werte etwas stärker ausgeprägt. Die Koeffizienten der Körpergewichtsvariablen sind negativ. Der (geringe) Effekt ist am 25%-Quantil schwächer und nimmt dann entlang der Verteilung zu. D.h., je schwerer die Probanden werden desto negativer wirkt sich dies auf den Tiffeneau-Wert aus, wobei sich eine Gewichtsveränderung am 75%-Quantil stärker auf die relative Einsekundenkapazität auswirkt als am 25%-Quantil. Die Teilnehmer aus Dresden zeigten über alle Quantile schlechtere Tiffeneau-Werte als Teilnehmer aus München. Eine einmalige Asthmadiagnose wirkte sich bei Probanden mit geringen Tiffeneau-Werten gegenüber Teilnehmern ohne Asthma negativ aus. Die Schätzungen in den oberen beiden Quantilen sind nicht signifikant. Die Ergebnisse bzgl. des beruflichen Asthmarisikos sind ebenfalls sehr heterogen verteilt.

In der *Tabelle 37* sind die Ergebnisse für die männlichen Teilnehmer ($n = 211$) abgebildet. Die Rauchexposition scheint bei den männlichen Probanden keine Rolle zu spielen. Die FEV1/ FVC-Werte aus der ISAAC II-Studie und der Erhebungsort haben über alle Quantile hinweg einen bedeutenden Einfluss. Das Körpergewicht hat im Bereich des Medians und am 75%-Quantil einen bedeutenden negativen Effekt, während das Ergebnis am 25%-Quantil auf dem 10%-Niveau signifikant ist. Eine einmalige Asthmadiagnose wirkte sich am 50%-Quantil negativ aus. Die OLS-Schätzungen für die Variablen zum Tiffeneau-Wert aus der ISAAC II-Studie, zur Asthmadiagnose der Probanden und zum Erhebungszentrum liegen alle in der Größenordnung der Schätzungen für das 50%-Quantil. Dies gilt ebenfalls für den Koeffizienten zum Körpergewicht entsprechend der kleinsten Quadrate Schätzung auf dem 10%-Niveau.

Die Ergebnisse der einzelnen Quantilregressionen und die OLS-Schätzungen für die weiblichen Teilnehmer sind in der *Tabelle 38* abgebildet. Das Passivrauchen hat am 25%-Quantil und am 75%-Quantil einen positiven Einfluss auf die relative Einsekundenkapazität. Dieser Effekt ist im oberen Bereich der Verteilung etwas stärker ausgeprägt. Im Hinblick auf die Ergebnisinterpretation wäre diesbezüglich ein negativer Einfluss auf die Zielgröße zu erwarten gewesen. Zudem sind die Effekte der Tiffeneau-Werte aus der ISAAC II-Studie an den Rändern der Verteilung bezogen auf die Zielgröße etwas stärker ausgeprägt. Entgegen der Annahme wirkte sich eine einmalige Asthmadiagnose am 75%-Quantil positiv aus, während eine mehrmalige Asthmadiagnose erwartungsgemäß negative Effekte

auf die Tiffeneau-Werte am 25%-Quantil zur Folge hat. Die Ergebnisse für die Variablen zum beruflichen Asthmarisiko und zur Asthmad Diagnose der Mutter sind wiederum heterogen.

Zusammenfassend kann gesagt werden, dass auch unter Verwendung der Quantilregression bzw. der Fokussierung auf spezifische Quantile der Verteilung des Tiffeneau-Wertes sich hinsichtlich der Variablen zur aktiven und passiven Tabakrauchexposition keine signifikanten Ergebnisse und damit zusätzliche Erkenntnisse gegenüber den Resultaten der multiplen linearen Regression ergeben haben. Dieses gilt sowohl für die Datenbasis insgesamt wie auch für die männlichen und weiblichen Probanden. Des Weiteren liefert die Quantilregression bzgl. der zusätzlichen Einflussgrößen Ergebnisse, welche die bisherigen Resultate der OLS-Schätzung bestätigen.

Quantilregression

FEV1/ FVC (n = 523)				
Quantile	25%	50%	75%	OLS
	Koeffizient Standardfehler P-Wert	Koeffizient Standardfehler P-Wert	Koeffizient Standardfehler P-Wert	Koeffizient Standardfehler P-Wert
(Intercept)	0,42400 0,10224 0,00004	0,44927 0,08284 0,00000	0,37401 0,08945 0,00003	0,40490 0,08270 < 0,00001
FEV1/ FVC (ISAAC II)	0,63702 0,05427 0,00000	0,60715 0,04399 0,00000	0,63396 0,05127 0,00000	0,63540 0,04317 < 0,00001
Passivrauch kumuliert	0,00117 0,00088 0,18281	-0,00018 0,00096 0,85497	0,00034 0,00069 0,61831	0,00022 0,00082 0,78817
Dauer aktiven Rauchens	0,00005 0,00132 0,96869	0,00068 0,00107 0,52676	0,00027 0,00128 0,83512	0,00074 0,00110 0,49902
Zigaretten pro Tag	0,00016 0,00079 0,84089	0,00013 0,00060 0,82867	0,00012 0,00065 0,85915	0,00011 0,00064 0,86804
Körpergröße	-0,04308 0,04960 0,38552	-0,02485 0,04554 0,58554	0,02098 0,04223 0,61960	-0,02076 0,04121 0,61468
Körpergewicht	-0,00068 0,00024 0,00482	-0,00094 0,00025 0,00021	-0,00105 0,00015 0,00000	-0,00082 0,00025 0,00109
Einmalige Asthmad Diagnose (Referenz: keine Diagnose)	-0,02240 0,01019 0,02836	-0,01382 0,01344 0,30447	-0,00709 0,00797 0,37399	-0,02219 0,01047 0,03446
Mehrmalige Asthmad Diagnose (Referenz: keine Diagnose)	-0,05642 0,02873 0,05008	-0,04268 0,02566 0,09687	0,00721 0,04281 0,86628	-0,03417 0,01373 0,01314
Frauen (Referenz: Männer)	-0,00492 0,00899 0,58424	-0,00808 0,00794 0,30915	-0,00883 0,00845 0,29663	< 0,00001 0,00720 0,99965
Dresden (Referenz: München)	-0,04928 0,00613 0,00000	-0,03830 0,00523 0,00000	-0,03476 0,00544 0,00000	-0,04361 0,00503 < 0,00001
Hohe Bildung (Referenz: niedrige Bildung)	0,00582 0,00653 0,37332	0,00959 0,00609 0,11560	-0,00364 0,00653 0,57777	0,00307 0,00577 0,59442
Geringes berufliches Asthma- risiko (Referenz: kein Risiko)	-0,00164 0,00681 0,81007	-0,01587 0,00668 0,01788	-0,00193 0,01051 0,85458	-0,00661 0,00664 0,32013
Hohes berufliches Asthma- risiko (Referenz: kein Risiko)	-0,00219 0,00726 0,76270	-0,02021 0,00571 0,00044	-0,01309 0,00619 0,03486	-0,00850 0,00577 0,14121
Mutter Asthma - nein (Referenz: ja)	-0,01032 0,00615 0,09400	-0,00246 0,01025 0,81064	0,00658 0,00851 0,43977	-0,00369 0,01267 0,77083
Mutter Asthma - unbekannt (Referenz: ja)	-0,06787 0,04572 0,13829	-0,01398 0,01844 0,44860	-0,01364 0,01283 0,28809	-0,02376 0,02142 0,26793
Vater Asthma - nein (Referenz: ja)	-0,01869 0,01734 0,28171	0,00185 0,01283 0,88528	0,00720 0,00693 0,29896	-0,00042 0,01250 0,97320
Vater Asthma - unbekannt (Referenz: ja)	-0,00595 0,02126 0,77975	0,00453 0,01410 0,74829	0,00516 0,01663 0,75628	0,00285 0,01628 0,86128

Tabelle 36: Ergebnisse simultaner Quantilregressionen und einer OLS-Regression
bezogen auf die gesamte Datenbasis

Quantilregression

FEV1/ FVC (n = 211)				
Quantile	25%	50%	75%	OLS
	Koeffizient Standardfehler P-Wert	Koeffizient Standardfehler P-Wert	Koeffizient Standardfehler P-Wert	Koeffizient Standardfehler P-Wert
(Intercept)	0,52171 0,18375 0,00500	0,42592 0,12614 0,00089	0,39277 0,15423 0,01165	0,45587 0,13802 0,00114
FEV1/ FVC (ISAAC II)	0,51503 0,11846 0,00002	0,57418 0,06545 0,00000	0,55108 0,09381 0,00000	0,60269 0,07517 < 0,00001
Passivrauch kumuliert	-0,00193 0,00250 0,44217	-0,00112 0,00100 0,26425	-0,00118 0,00182 0,51853	-0,00158 0,00148 0,28680
Dauer aktiven Rauchens	0,00385 0,00305 0,20877	0,00199 0,00222 0,37093	-0,00108 0,00244 0,66063	0,00217 0,00230 0,34699
Zigaretten pro Tag	-0,00100 0,00147 0,49732	-0,00124 0,00124 0,31910	0,00034 0,00097 0,72544	-0,00070 0,00114 0,54358
Körpergröße	-0,02213 0,10102 0,82680	-0,00779 0,05832 0,89390	0,03874 0,07102 0,58605	-0,03593 0,06753 0,59525
Körpergewicht	-0,00104 0,00061 0,08938	-0,00073 0,00035 0,03925	-0,00099 0,00040 0,01435	-0,00080 0,00041 0,05616
Einmalige Asthmediagnose (Referenz: keine Diagnose)	-0,03712 0,02933 0,20723	-0,05121 0,01258 0,00007	-0,05311 0,03942 0,17950	-0,03659 0,01846 0,04883
Mehrmalige Asthmediagnose (Referenz: keine Diagnose)	-0,05954 0,03049 0,05228	-0,02415 0,03128 0,44098	-0,03841 0,03630 0,29124	-0,03483 0,02320 0,13486
Dresden (Referenz: München)	-0,05028 0,01347 0,00025	-0,04977 0,00774 0,00000	-0,03509 0,01275 0,00649	-0,05078 0,00896 < 0,00001
Hohe Bildung (Referenz: niedrige Bildung)	0,00385 0,01947 0,84334	0,01164 0,00904 0,19944	-0,00883 0,01454 0,54432	0,00683 0,01057 0,51858
Geringes berufliches Asthma- risiko (Referenz: kein Risiko)	-0,00997 0,01861 0,59280	-0,02471 0,01305 0,05987	0,00159 0,01695 0,92515	-0,00754 0,01169 0,51989
Hohes berufliches Asthma- risiko (Referenz: kein Risiko)	0,00125 0,01651 0,93964	-0,00601 0,00745 0,42119	-0,00611 0,01196 0,61004	0,00296 0,01053 0,77913
Mutter Asthma - nein (Referenz: ja)	-0,01205 0,02195 0,58369	0,01760 0,02755 0,52371	0,02695 0,01378 0,05199	0,00628 0,02921 0,83003
Mutter Asthma - unbekannt (Referenz: ja)	-0,07020 0,06925 0,31192	0,02774 0,02888 0,33796	0,00612 0,07357 0,93381	-0,02948 0,04894 0,54758
Vater Asthma - nein (Referenz: ja)	-0,01176 0,02400 0,62456	-0,00693 0,02067 0,73784	0,01536 0,01843 0,40569	-0,00487 0,02114 0,81814
Vater Asthma - unbekannt (Referenz: ja)	0,01415 0,03403 0,67806	-0,00722 0,02074 0,72825	0,01931 0,03613 0,59373	0,01080 0,02766 0,69676

Tabelle 37: Ergebnisse simultaner Quantilregressionen und einer OLS-Regression
bezogen auf die männlichen Teilnehmer

Quantilregression

FEV1/ FVC (n = 312)				
Quantile	25%	50%	75%	OLS
	Koeffizient Standardfehler P-Wert	Koeffizient Standardfehler P-Wert	Koeffizient Standardfehler P-Wert	Koeffizient Standardfehler P-Wert
(Intercept)	0,43111 0,10823 0,00009	0,41923 0,08718 0,00000	0,24127 0,11161 0,03144	0,33193 0,10165 0,00122
FEV1/ FVC (ISAAC II)	0,69353 0,05361 0,00000	0,64607 0,04449 0,00000	0,69798 0,06311 0,00000	0,66974 0,05318 < 0,00001
Passivrauch kumuliert	0,00168 0,00082 0,03966	0,00084 0,00121 0,48779	0,00313 0,00078 0,00008	0,00146 0,00098 0,13860
Dauer aktiven Rauchens	-0,00007 0,00201 0,97204	0,00122 0,00083 0,14341	0,00037 0,00113 0,74522	0,00018 0,00121 0,88047
Zigaretten pro Tag	0,00022 0,00114 0,84519	-0,00025 0,00050 0,61325	-0,00006 0,00066 0,92826	0,00014 0,00078 0,85745
Körpergröße	-0,08673 0,05322 0,10425	-0,04115 0,05036 0,41447	0,08658 0,06055 0,15381	0,01101 0,05322 0,83628
Körpergewicht	-0,00049 0,00026 0,05788	-0,00062 0,00028 0,02565	-0,00153 0,00030 0,00000	-0,00086 0,00031 0,00664
Einmalige Asthmadignose (Referenz: keine Diagnose)	-0,00487 0,02920 0,86760	0,01632 0,01228 0,18496	0,01935 0,00836 0,02125	-0,00736 0,01271 0,56311
Mehrmalige Asthmadignose (Referenz: keine Diagnose)	-0,07301 0,02249 0,00130	-0,05923 0,03515 0,09307	0,01179 0,06265 0,85082	-0,04105 0,01750 0,01961
Dresden (Referenz: München)	-0,04477 0,00706 0,00000	-0,03299 0,00493 0,00000	-0,04195 0,00684 0,00000	-0,03818 0,00601 < 0,00001
Hohe Bildung (Referenz: niedrige Bildung)	0,00746 0,00664 0,26216	0,00871 0,00661 0,18815	-0,00160 0,00813 0,84437	0,00061 0,00695 0,93018
Geringes berufliches Asthma- risiko (Referenz: kein Risiko)	-0,00873 0,00798 0,27467	-0,01179 0,00604 0,05193	-0,01751 0,00874 0,04597	-0,00432 0,00809 0,59358
Hohes berufliches Asthma- risiko (Referenz: kein Risiko)	-0,00711 0,00651 0,27518	-0,02615 0,00550 0,00000	-0,01725 0,01181 0,14526	-0,01644 0,00681 0,01631
Mutter Asthma - nein (Referenz: ja)	-0,01665 0,00633 0,00899	-0,00322 0,00787 0,68247	-0,00169 0,00673 0,80239	-0,00783 0,01328 0,55573
Mutter Asthma - unbekannt (Referenz: ja)	-0,00903 0,05495 0,86958	-0,01577 0,02366 0,50560	-0,03169 0,01891 0,09472	-0,02897 0,02250 0,19895
Vater Asthma - nein (Referenz: ja)	-0,01722 0,02551 0,50016	-0,00855 0,00810 0,29199	0,00256 0,00886 0,77281	-0,00677 0,01580 0,66855
Vater Asthma - unbekannt (Referenz: ja)	-0,01905 0,02799 0,49658	-0,01206 0,01314 0,35958	0,01221 0,01908 0,52270	-0,00942 0,02058 0,64764

Tabelle 38: Ergebnisse simultaner Quantilregressionen und einer OLS-Regression
bezogen auf die weiblichen Teilnehmer

6. *Zusammenfassung*

Eine Vielzahl von Lungen- und Atemweg Erkrankungen wie z.B. chronisch obstruktive Bronchitis, Lungenemphysem (COPD), die Bildung von Lungenkarzinomen oder chronisch-entzündlichen Erkrankungen wie Asthma bronchiale werden durch den Konsum von Tabak verursacht oder negativ beeinflusst. Mit Blick auf die Todesursachenstatistik gingen dabei im Jahr 2007 allein ca. 5,2% aller Todesfälle auf Krebserkrankungen zurück, welche in Bezug zum Tabakkonsum gebracht werden können. Die Zahl derartiger Todesursachen ist - wie bereits einleitend beschrieben - in den vergangenen Jahren bedauerlicherweise angestiegen (vgl. Statistisches Bundesamt (2007, 2008)). Im Gegensatz dazu ist die Quote jugendlicher Raucher entsprechend der Ergebnisse der Langzeitstudie der Bundeszentrale für gesundheitliche Aufklärung (BZgA) in den vergangenen Jahren erfreulicherweise gesunken (vgl. BZgA (2008)). Hervorzuheben ist in diesem Kontext dennoch, dass Raucher nicht nur ihre eigene Gesundheit gefährden, sondern auch Passivrauchern teilweise schwerwiegende gesundheitliche Schäden zufügen.

Vor diesem Hintergrund war es die Zielsetzung der vorliegenden Arbeit den Einfluss des aktiven und passiven Rauchens auf die Gesundheit, insbesondere auf die Lungenfunktionsfähigkeit heranwachsender Jugendlicher eingehender zu untersuchen. Die empirische Analyse der zu bearbeitenden Fragestellung, sollte mit Hilfe der Daten erfolgen, welche im Rahmen der ISAAC II- und SOLAR II-Studie erhoben wurden. Der Einfluss auf die Lungenfunktion sollte dabei mit den in der medizinischen Untersuchung erhobenen Parametern zur Einsekundenkapazität und zur forcierten Vitalkapazität sowie bzgl. der daraus berechneten relativen Einsekundenkapazität („Tiffeneau-Wert“) beurteilt werden.

In einem ersten Schritt wurde zunächst eine univariate Datenbetrachtung durchgeführt. Darauf aufbauend wurde der Zusammenhang zwischen den jeweiligen Zielgrößen und den

potentiellen Einflussgrößen/ Confoundern z.B. anhand von Streudiagrammen und Boxplots untersucht. Mit Hinblick auf die eigentliche Zielsetzung, die zentrale Fragestellung mit Hilfe der Regressionsanalyse zu untersuchen, dienten diese deskriptiven Analysen dazu, einen ersten Überblick über die zugrunde liegende Datenstruktur zu gewinnen. Hinsichtlich der zentralen Fragestellung konnte kein Zusammenhang zwischen den jeweiligen Lungenfunktionsparametern und den Variablen zum aktiven und passiven Rauchverhalten, resp.

- zur Dauer des aktiven Rauchens
- zur durchschnittlichen Anzahl an Zigaretten pro Tag
- zu den Packungsjahren
- zum Passivrauchen
- zur kumulierten Passivrauchexposition

identifiziert werden.

Beobachtet werden konnten dagegen Zusammenhänge zwischen

- dem Geschlecht
- der Körpergröße
- dem Körpergewicht
- der Asthmavorerkrankung
- den Lungenfunktionsparametern der ISAAC II-Studie
- dem Erhebungszentrum

bezüglich der Einsekundenkapazität, der forcierten Vitalkapazität und der relativen Einsekundenkapazität. Des Weiteren hat die graphische Betrachtung gezeigt, dass die stetigen potentiellen Einflussgrößen/ Confounder zu den jeweiligen stetigen Zielgrößen eher in einem linearen Zusammenhang zueinander stehen. Unter Verwendung des generalisierten additiven Modells wurde dieser erste Eindruck aus der deskriptiven Analyse bestätigt.

Vor diesem Hintergrund wurde die zentrale Fragestellung unter Verwendung der multiplen linearen Regression methodisch weiter bearbeitet. Dabei wurde das im Sinne des AIC-Kriteriums „beste“ Modell ausgewählt und unter Einsatz des k-fachen Kreuzvalidierungsverfahrens bzw. unter Betrachtung der Prognosegüte die jeweilige Modellwahl bestätigt. Wie einleitend bereits beschrieben, kann das aktive und passive Rauchen mit einer Reihe von Lungenerkrankungen in Verbindung gebracht werden. Hinsichtlich der zu bearbeiten-

den Fragestellung wäre unter Verwendung der ISAAC II- und SOLAR II-Daten hier ein Effekt zu erwarten gewesen. Allerdings konnte im Rahmen der Regressionsanalyse eine Beeinträchtigung der Lungenfunktionstätigkeit durch die aktive und passive Rauchexposition grundsätzlich nicht nachgewiesen werden. Dieses Ergebnis kann auch für die geschlechtsspezifischen Regressionsanalysen festgehalten werden.

Wie bereits durch die deskriptiven Analysen beobachtet, zeichnete sich bei den zusätzlichen Variablen (z.B. Körpergröße, Gewicht) durchaus ein Einfluss auf die drei Lungenfunktionsparameter ab. Besonders hervorzuheben ist in diesem Kontext die Variable zur Asthmadiagnose der Patienten, welche in verschiedenen Modellkonstellationen den genau zu erwartenden Effekt zeigte. So kann bspw. gesagt werden, dass eine Asthmaerkrankung bei den Frauen einen negativen Einfluss auf die Einsekundenkapazität ausübt. Des Weiteren stellte sich bei der Überprüfung der Modellannahmen heraus, dass das Ergebnis für die männlichen Probanden bezogen auf den Tiffeneau-Wert im Bereich negativer Residuen Abweichungen von der Normalverteilungsannahme zeigte.

Demzufolge sollte mit Hilfe der Quantilregression untersucht werden, inwieweit die Auswirkungen der Einflussgrößen für die Modelle zum Tiffeneau-Wert über die gesamte Verteilung der Werte stabil bleiben bzw. die Ergebnisse der linearen multiplen Regression als robust einzustufen sind. Wobei der Fokus insbesondere auf dem Modell für die männlichen Probanden lag. Anzumerken ist in diesem Kontext, dass sich die Quantilregression zur Bearbeitung dieser Fragestellung angeboten hat, da sie ohne die Annahme normalverteilter Residuen auskommt und mit ihrer Hilfe sich zudem die gesamte bedingte Verteilung modellieren lässt.

Auch unter Verwendung der Quantilregression bzw. der Fokussierung auf spezifische Quantile der Verteilung des Tiffeneau-Wertes ergeben sich hinsichtlich der Variablen zur aktiven und passiven Tabakrauchexposition keine signifikanten Ergebnisse und damit zusätzliche Erkenntnisse gegenüber den Resultaten der multiplen linearen Regression. Dieses gilt sowohl für die Datenbasis insgesamt, wie auch für die männlichen und weiblichen Probanden. Des Weiteren liefert die Quantilregression bzgl. der zusätzlichen Einflussgrößen Ergebnisse, welche die bisherigen Resultate der OLS-Schätzungen bestätigen.

Zusammenfassend kann gesagt werden, dass die ISAAC II- und SOLAR II-Studien hinsichtlich ihrer Forschungszielrichtung wesentlich breiter angelegt waren, als auf die explizite Betrachtung der Tabakrauchexposition heranwachsender Jugendlicher. Zielsetzung beider Studien war im Wesentlichen die Untersuchung der weltweiten Prävalenz von Symptomen allergischer und asthmatischer Erkrankungen bei Kindern und Jugendlichen. Vor diesem Hintergrund sollten die Ergebnisse dieser Arbeit nicht verallgemeinert betrachtet werden.

Literaturverzeichnis

Andreas, S.; Hering, T.; Mühlig, S.; Nowak, D.; Raupach, T.; Worth, H. (2009): Tabakentwöhnung bei chronisch obstruktiver Lungenerkrankung – Eine effektive und sinnvolle medizinische Intervention, Deutsches Ärzteblatt, Jg. 106, Heft 16, S. 276 – 282.

Baur, X (2001): Arbeitsmedizinische Leitlinie der Deutschen Gesellschaft für Arbeitsmedizin und Umwelt e.V. (DGAUM) – Arbeitsgruppe „Arbeitsbedingte Gefährdungen und Erkrankungen der Lunge und Atemwege“ – Lungenfunktionsprüfungen in der Arbeitsmedizin, Arbeitsmedizin, Sozialmedizin, Umweltmedizin (ASU), Jg. 36, Heft 10, S. 498 – 505.

Bundeszentrale für gesundheitliche Aufklärung BZgA (2001): Jugendliche Raucher - Veränderung des Rauchverhaltens und Ansätze für die Prävention – Ergebnisse der Wiederholungsbefragung „Drogenaffinität Jugendlicher in der Bundesrepublik Deutschland 2001“, <http://www.bzga.de/studien>, Zugriff: 09. November 2009.

Bundeszentrale für gesundheitliche Aufklärung BZgA (2008): Die Drogenaffinität Jugendlicher in der Bundesrepublik Deutschland 2008 – Alkohol-, Tabak- und Cannabiskonsum – Erste Ergebnisse zu aktuellen Entwicklungen und Trends, <http://www.bzga.de/studien>, Zugriff: 05. November 2009.

Deutsches Krebsforschungszentrum dkfz (2003): Passivrauchende Kinder in Deutschland – Frühe Schädigung für ein ganzes Leben, Heidelberg.

Efron, B.; Tibshirani, R. J. (1998): An Introduction to the Bootstrap, CRC Monographs on Statistics & Applied Probability, Chapman & Hall, New York.

Fachgruppe „Suchtprävention“ im Auftrag des Drogen- und Suchtrates (Juni 2008): Empfehlungen des Drogen- und Suchtrates an die Drogenbeauftragte der Bundesregierung für ein Nationales Aktionsprogramm zur Tabakprävention, <http://www.bmg.bund.de>, Zugriff: 07. November 2009.

Fahrmeier, L.; Kneib, T.; Lang, S. (2007): Regression – Modelle, Methoden und Anwendungen, 5. Auflage, Springer Verlag, Berlin et al.

Koenker, R. (2005): Quantile Regression, Cambridge University Press, Cambridge et al.

Kraus, L.; Pabst, A.; Steiner, S. (2008): Europäische Schülerstudie zu Alkohol und anderen Drogen 2007 (ESPAD), Institut für Therapieforchung (IFT), München.

Lampert, T. (2008): Tabakkonsum und Passivrauchbelastung von Jugendlichen – Ergebnisse des Kinder- und Jugendgesundheitssurvey (KiGGS), Deutsches Ärzteblatt, Jg. 105, Heft 12, S. 265 – 270.

Radon, K.; Dressel, H.; Hümmer, S.; Riu, E.; Nowak, D.; Weinmayr, G.; Genuneit, J.; Weiland, St.; Vogelberg, Ch.; Leupold, W.; Windstetter, D.; von Mutius, E.; Kupfer, J. (2005): Berufliche Allergierisiken – Die SOLAR-Kohortenstudien; Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund et al.

Radon, K.; Heinrich, S.; Kellberger, J.; Peter, A.; Brückner, U.; Kolb, S.; Braun, E.; Nowak, D.; Brunner, U.; Brzozowski, B.; von Mutius, E.; Ellenberg, D.; Kirsten, L.; Podworny, U.; Vogelberg, C.; Weinmayr, G.; Genuneit, J.; Firscher, T. (2008): Studie in Ost- und Westdeutschland zu beruflichen Allergierisiken – SOLAR II – Sachstandsbericht, München.

Statistisches Bundesamt (2007): Gesundheit – Todesursachen in Deutschland, Fachserie 12, Reihe 4.

Statistisches Bundesamt (2008): Aktuelle Daten zum Rauchverhalten in Deutschland, Wirtschaft und Statistik, Heft 5, S. 375 – 376.

Anhang I: CD-Inhalt

Auf dem beiliegenden Datenträger ist eine digitale Ausgabe der vorliegenden Bachelorarbeit (pdf-Format) enthalten. Die zugrunde liegenden Erhebungsbögen der ISAAC II- und SOLAR II-Studien, die verwendeten Daten, der R-Code sowie die Literatur (z.B. elektronische Aufsätze) befinden sich ebenfalls auf der beiliegenden CD.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbstständig verfasst und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

München, den 02. Dezember 2009

(Martin Bartke)