

Berufsbedingte Atemwegserkrankungen

Ursula Gerhardinger, Florian Hiemeyer, Axel Kraus, Anja Kühnlein

19. Februar 2004

Inhaltsverzeichnis

1	Einführung	2
1.1	Inhalt und Durchführung der Studie	2
1.2	Datenstruktur und Variablenbeschreibung	2
2	Datenbearbeitung	3
2.1	Zusammenführung der Datensätze	3
2.2	Bildung von Scores	3
2.2.1	Exposition	3
2.2.2	Genetische Prädisposition	5
2.2.3	Passivrauchen	6
3	Deskriptive Analyse	6
3.1	Soziodemographische Daten	6
3.2	Symptomprävalenzen	8
3.3	Assoziation zwischen beruflicher Exposition und Atemwegserkrankungen	11
3.3.1	Allergischer Schnupfen	11
3.3.2	Rhinitis	11
4	Non-Response Analyse	13
5	Allgemeines zur Modellwahl	14
6	Umkodierungen	15
7	Der Weg vom Ausgangsmodell zum Endmodell	16
7.1	Ausgangsmodell	16
7.2	Interaktionen	16
7.3	Confounder	17

1	<i>EINFÜHRUNG</i>	2
7.4	Effektmodifizierer	18
7.5	Endmodelle	18
8	Interpretation der Modelle	26
9	Longitudinale Betrachtung	27
9.1	Vereinfachte longitudinale Betrachtung	27
9.2	Survivalanalyse	27
9.2.1	Konstruktion des Datensatzes	27
9.2.2	Berechnung des Survivalmodells	28
9.2.3	Interpretation der Survivalanalyse	30

1 Einführung

1.1 Inhalt und Durchführung der Studie

Beim European Community Respiratory Health Survey (ECRHS) handelt es sich um eine bevölkerungsbezogene internationale Kohortenstudie. In zwei Erhebungen aus den Jahren 1991/1992 und 2001/2002 wurden in den Studienzentren Hamburg und Erfurt Probanden untersucht, die zum Zeitpunkt der ersten Studie zwischen 20 und 44 Jahre alt waren. Bei der ersten Studie lag die Anzahl der befragten Personen in Hamburg bei 1259 und in Erfurt bei 732. Durch Umzüge, die nicht nachverfolgt werden konnten und durch Verweigerung der Teilnahme an einer zweiten Untersuchung reduzierte sich die Zahl der Probanden in der zweiten Studie in Hamburg auf 303 und in Erfurt auf 287. Der ECRHS interessiert sich für mögliche Zusammenhänge zwischen Schadstoffexposition am Arbeitsplatz und dem Auftreten von chronischen Atemwegserkrankungen. In diesem Projekt beschränken wir uns auf die Betrachtung von Allergien und Rhinitis und ihre Verbindung zur Belastung mit biologischen, mineralischen und gasförmigen Schadstoffen, sowie der Passivrauchexposition am Arbeitsplatz.

1.2 Datenstruktur und Variablenbeschreibung

In beiden Studien wurde eine Vielfalt von medizinischen und demographischen Variablen erhoben, aus der wir nur eine Auswahl betrachten. Bei den Erkrankungen setzen wir den Schwerpunkt auf das Auftreten allergischer Reaktionen und Rhinitis, da für chronische Bronchitis und Asthma die Fallzahlen zu gering sind, um aussagekräftige Ergebnisse zu erhalten. Als Prädiktoren und eventuelle Confounder werden Variablen zu Wohnort

(Studienzentrum in Hamburg oder Erfurt), Alter, Geschlecht, Bildungsniveau, Anzahl der Geschwister, Allergien bzw. Asthma in der Familie, den Lebensumständen im Kleinkindalter und zum aktuellen und evtl. früheren Rauchverhalten in Betracht gezogen. In der zweiten Erhebung wurde zusätzlich (retrospektiv) die Berufshistorie der Probanden erhoben. Zusammen mit einer Schadstoffcodierung konnten damit aus der Liste der ausgeübten Berufe drei Expositionsscores (getrennt nach biologischen, mineralischen und gasförmigen Schadstoffen) berechnet werden, die die Belastungen wiedergeben sollen, denen der Proband im Zeitraum zwischen den beiden Studien ausgesetzt war.

2 Datenbearbeitung

2.1 Zusammenführung der Datensätze

Zuerst wurden die, nach den Untersuchungszentren getrennten, Datensätze der zweiten Erhebung nach der Identifikationsnummer der Probanden geordnet zusammengesetzt. Anschließend wurden in diesem Datensatz und im Gesamtdatensatz der ersten Erhebung alle Variablen, die für unsere Analyse unwichtig sind, gelöscht. Die Datensätze enthielten nun nur noch die Krankheitsvariablen und die oben genannten Prädiktoren bzw. Confounder. Diese beiden gekürzten Datensätze, zusammengefügt mit Hilfe der Variablen *Id*, ergaben den Untersuchungsdatensatz mit allen für uns relevanten Variablen. Den endgültigen Datensatz für die Erstellung der deskriptiven Analyse erhielten wir nach Anfügen der Spalten für die drei Berufsexpositionsscores, die Scores für Allergie bzw. Asthma in der Familie, den Passivrauchscore (siehe Punkte 2.2.1 bis 2.2.3) und den Bildungsstand an den Datensatz.

Der Bildungsstand wurde aus einer Tabelle 'Alter in Jahren bei Ausbildungsabschluß' bestimmt, wobei ein Alter ≤ 20 Jahre bei Ausbildungsabschluß einem niedrigen Bildungsstand (Ausprägung 1) entspricht und ein Alter > 20 Jahren einem hohen Bildungsstand (Ausprägung 2).

2.2 Bildung von Scores

2.2.1 Exposition

Für die Berechnung des Berufsexpositionsscores wurde zu jedem Untersuchungsteilnehmer die Dauer aller seiner Beschäftigungen in Monaten einzeln ermittelt. Betrachtet wurden die Arbeitsverhältnisse seit Januar 1991. War der tatsächliche Arbeitsbeginn bereits früher, so wurde dieser auf Januar

1991 umdatiert, um unsere Analysen auf einem für alle Probanden einheitlichen Zeitrahmen aufbauen zu können.

Fehlten Werte bei der Angabe 'Monat Jobanfang bzw. -ende', wurden diese auf Juli bzw. Juni (Mitte des Jahres) gesetzt.

Für jede Beschäftigungsart lag eine Zuordnung der Stärke der verschiedenen Kategorien von Atemwegsbelastungen vor: Belastung durch biologischen Staub, durch mineralischen Staub und durch Gas/Rauch. Die Kodierung lautete für alle Schadstoffarten 'keine Belastung'=0, 'etwas Belastung'=1 und 'starke Belastung'=2. Somit konnte zu jeder Beschäftigung eine Maßzahl für die Belastung gefunden werden, die sich aus dem Produkt von Beschäftigungsdauer und Intensität der Belastung ergibt.

Die Scores der aufeinander folgenden Beschäftigungen, die ein Proband hatte, wurden über den Zeitraum zwischen den beiden Studien kumuliert, um jeweils einen Wert für die Gesamtexposition zu erhalten. Dies erfolgte getrennt für alle drei Arten von Belastung. War ein Proband zu einer Zeit mehrfach beschäftigt, so ging jede Beschäftigung mit vollem Gewicht in den Score ein.

Formal zusammengefasst (z.B. für biologische Exposition):

Ein Proband hat seit Januar 1991 die Beschäftigungen (B_1, \dots, B_n) mit jeweiliger Schadstoffbelastung (S_1, \dots, S_n) ;

zu jedem $B_i \quad \exists \quad (\Delta t)_i$ für $i = 1, \dots, n$

$$b_{gesamt} = \sum_{i=1}^n S_i * (\Delta t)_i$$

Die Berechnung für m_{gesamt} und g_{gesamt} verläuft nach dem gleichem Schema.

Dazu ein kurzes Rechenbeispiel:

Ist ein Proband z.B. von Mai 1987 bis August 1992 als Fischer tätig gewesen, hat dann von August 1992 bis Juni 2001 als Waldarbeiter gearbeitet und hatte von März 1992 bis Oktober 1994 als Fensterputzer einen Nebenjob, so lassen sich für ihn die Schadstoffbelastungen wie folgt berechnen:

- Umkodierung von Mai 1987 auf Januar 1991 \Rightarrow Dauer von Beschäftigung 1: (August 1992 bis Januar 1991) = 19 Monate
- Beschäftigungsdauer 2: (Juni 2001 bis August 1992) = 106 Monate
- Beschäftigungsdauer 3: (Oktober 1994 bis März 1992) = 31 Monate

Beschäftigung 1 (=Fischer) hat nach entsprechender Tabelle folgende Schadstoffeinstufungen:

CODE	CLASSIFICATION	BIOLDUST	MINEDUST	GASFUMES
· 171	· Fishermen	· 1	· 0	· 1

⇒ Schadstoffbelastungen für Beschäftigung 1 sind

BIOLDUST	MINEDUST	GASFUMES
$19 \times 1 = 19$	$19 \times 0 = 0$	$19 \times 1 = 19$

Beschäftigung 2 (=Waldarbeiter) hat nach entsprechender Tabelle folgende Schadstoffeinstufungen:

CODE	CLASSIFICATION	BIOLDUST	MINEDUST	GASFUMES
· 170	· Forestry workers	· 2	· 2	· 1

⇒ Schadstoffbelastungen für Beschäftigung 2 sind

BIOLDUST	MINEDUST	GASFUMES
$106 \times 2 = 212$	$106 \times 2 = 212$	$106 \times 1 = 106$

Beschäftigungsdauer 3 (=Fensterputzer) hat nach entsprechender Tabelle folgende Schadstoffeinstufungen:

CODE	CLASSIFICATION	BIOLDUST	MINEDUST	GASFUMES
· 158	· window cleaners	· 1	· 2	· 2

⇒ Schadstoffbelastungen für Beschäftigung 3 sind

BIOLDUST	MINEDUST	GASFUMES
$31 \times 1 = 31$	$31 \times 2 = 62$	$31 \times 2 = 62$

Insgesamt ergeben sich für den Probanden folgende Scores:

bgesamt	mgesamt	ggesamt
$19 + 212 + 31 = \mathbf{262}$	$0 + 212 + 62 = \mathbf{274}$	$19 + 106 + 62 = \mathbf{187}$

2.2.2 Genetische Prädisposition

Die genetische Vorbelastung für Allergien und allergischen Schnupfen wurde sehr einfach quantifiziert: Besitzt ein Proband mindestens entweder einen

Elternteil oder eine Schwester bzw. einen Bruder mit Prävalenz von Allergien, so wird ihm in der Variable *Allergie in der Familie* die Ausprägung 1 zugewiesen, andernfalls die Ausprägung 0. *Allergie in der Familie*=0 bedeutet also, dass der Proband keinerlei genetische Vorbelastung hatte. Hierbei handelte es sich um Fragen nach dem Auftreten von Ekzemen, Hautallergien, allergischem Schnupfen oder „Heuschnupfen“.

2.2.3 Passivrauchen

Die Bildung dieses Scores beruht auf Querschnittsdaten, da Angaben über einen kontinuierlichen Zeitverlauf fehlen. Dies stellt die Verwertbarkeit der Variable im weiteren Verlauf in Frage. Schließlich kann nicht eindeutig festgestellt werden, wie lange sie dem Passivrauch effektiv ausgesetzt waren. Als Passivraucher werden Probanden bezeichnet, die auf die Frage 59 (alter Fragebogen) bzw. Frage 75 (neuer Fragebogen) „Sind sie regelmäßig in den letzten 12 Monaten Tabakrauch anderer ausgesetzt gewesen?“ und auf die Frage 59.2 bzw. 75.2. „Wird in dem Raum, in dem Sie arbeiten, regelmäßig von anderen geraucht“ mit „ja“ geantwortet haben. Die Zuweisungen erfolgen auf folgende Weise:

0 für Probanden, die bei beiden Untersuchungen keine „Passivraucher“ waren.

1 für Probanden, die bei der ersten Untersuchung „Passivraucher“ waren.

2 für Probanden, die bei der zweiten Untersuchung „Passivraucher“ waren.

3 für Probanden, die bei beiden Untersuchungen „Passivraucher“ waren.

Fehlende Werte blieben fehlend.

3 Deskriptive Analyse

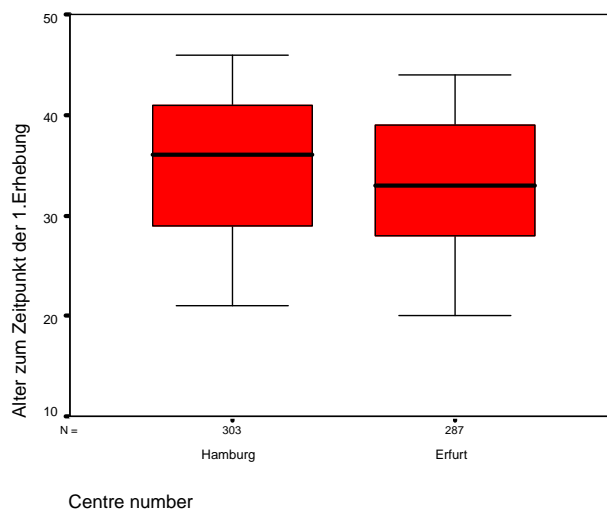
Bei der Betrachtung der Mittelwerte bei den Expositionsscores ist zu beachten, dass wir viele Probanden mit Expositionsscores = 0 haben, was die Mittelwerte verzerrend beeinflussen kann. Auf diese Problematik wird in dem Kapitel über die vorgenommenen Umkodierungen näher eingegangen (vgl. Kapitel 6).

3.1 Soziodemographische Daten

Die soziodemographischen Variablen der Untersuchungsgesamtheit werden in Tabelle 1 gezeigt.

Das *Alter* der Probanden zum Zeitpunkt der ersten Erhebung ist in Hamburg statistisch signifikant höher. Auch bei der Variable *Wohnort bis zum 5. Lebensjahr* findet man statistisch signifikante Unterschiede.

Altersvergleich



Etwa 40% der Probanden in beiden Zentren berichteten über *Passivrauchexposition am Arbeitsplatz*.

Variable	Hamburg	Erfurt
Alter zum Zeitpunkt der 1. Erhebung	35.06	33.02
Anzahl (%)		
Frauen	148 (48.8)	137 (47.7)
Raucher	200 (66.0)	184 (64.1)
Passivrauchexposition am Arbeitsplatz	132 (43.6)	110 (38.7)
Bildungsstand hoch	99 (32.7)	82 (28.6)
Allergie in der Familie	126 (47.0)	76 (28.9)
Wohnort bis 5. Lebensjahr:		
Bauernhof	16 (5.3)	25 (8.7)
ländliche Gegend	39 (12.9)	55 (19.2)
städtische Gegend	247 (81.8)	207 (72.1)

Tabelle 1: Beschreibung der soziodemographischen Variablen

3.2 Symptomprävalenzen

In Tabelle 2 sind die absoluten und die relativen Häufigkeiten der einzelnen Symptome aufgeführt.

Variable	Hamburg	Erfurt
Allergie	84 (27.7)	56 (19.5)
Chronische Bronchitis	11 (3.6)	8 (2.8)
Rhinitis	141 (46.5)	89 (31.0)

Tabelle 2: Absolute und relative Häufigkeiten der Symptome in beiden Untersuchungszentren

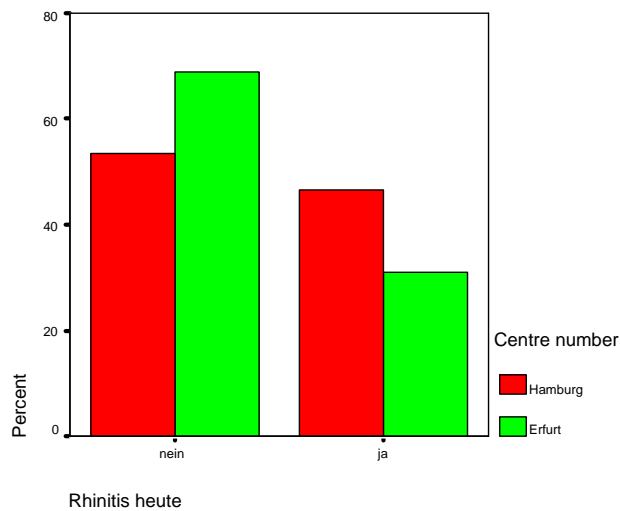
Die Variable *Allergie* beruht auf der Frage 15 „Haben Sie allergischen Schnupfen, z.B. Heuschnupfen?“. Vielfach ist die chronische Rhinitis die Vorstufe zum Asthma bronchiale, somit ist sowohl ein Zusammenhang mit der Allergie- als auch der Asthmaprädiposition vorhanden. Die Variablen *Allergie in der Familie* und *Asthma in der Familie* geben an, ob der Proband eine genetische Vorbelastung gegenüber zahlreichen allergischen Krankheiten bzw. Asthma hat. Um die Werte von *Allergie in der Familie* und *Asthma in der Familie* zu ermitteln, wurde - wie in 2.2.2 erläutert - ein Score ermittelt. In den folgenden Grafiken wird die unterschiedliche Aufteilung der Krankheiten auf die Untersuchungszentren Hamburg und Erfurt dargestellt.

Die *chronische Bronchitis* ist eine Dauererkrankung, des Bronchialsystems. Die WHO-Definition der Krankheit wurde im Fragebogen exakt abgefragt: Wer Frage 8.1 „Husten Sie - gewöhnlich im Winter während des Tages oder in der Nacht - mindestens 3 Monate jährlich?“ und Frage 10.1 „Haben Sie - im Winter gewöhnlich als erstes am Morgen - Auswurf?“ mit Ja beantwortete, wurde als erkrankt eingestuft.

Auch bei der *chronischen Rhinitis*, einer entzündlichen Erkrankung der Nasenschleimhaut - im allgemeinen Sprachgebrauch als „chronischer Schnupfen“ bezeichnet - wird die direkte Definition der Krankheit abgefragt: Wer die Frage 16.1 „Hatten Sie in den letzten 12 Monaten Probleme mit Niesanfällen oder einer laufenden, verstopften Nase ohne erkältet zu sein?“ bejahte, gilt als chronisch an Rhinitis erkrankt.

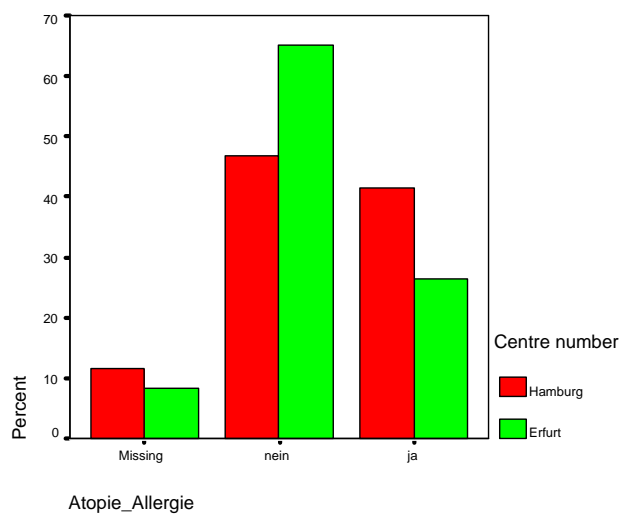
Statistisch signifikante Unterschiede zwischen beiden Zentren zeigen sich für das Auftreten des Merkmals *Rhinitis*. Der Anteil der Rhinitisfälle ist in Hamburg erhöht.

Rhinitis versus Zentrum

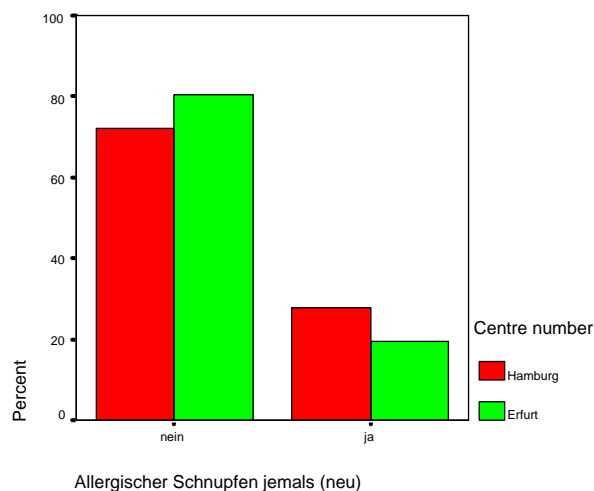


Auch bei den Merkmalen *Allergie* und *Allergie in der Familie* sind die Unterschiede signifikant (ebenfalls höherer Anteilswert in der Hansestadt).

Allergie in der Familie versus Zentrum

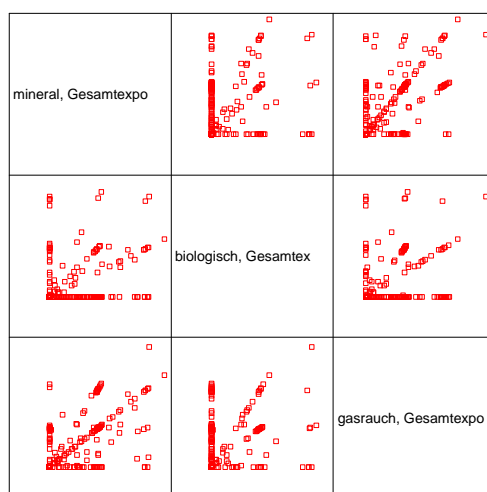


Allergie versus Zentrum



Auf Grund der sehr kleinen Fallzahl von *chronischer Bronchitis* ist es bei diesem Merkmal nicht sinnvoll, weitere statistische Zusammenhänge zu untersuchen.

Die folgende Grafik zeigt den Zusammenhang zwischen den verschiedenen Schadstoffexpositionen. Dabei ist zu erkennen, dass ein starker Zusammenhang zwischen mineralischer Staubexposition und Gas/Rauchexposition besteht.



In Tabelle 3 sind die Korrelationswerte zwischen den Expositionsscores aufgelistet. Alle Werte sind signifikant zum Niveau $\alpha = 0.01$.

Variable	BGesamt	MGesamt	GGesamt
BGesamt	1	0.343	0.383
MGesamt	0.343	1	0.692
GGesamt	0.383	0.692	1

Tabelle 3: Korrelationstabelle der Gesamtexpositionen

3.3 Assoziation zwischen beruflicher Exposition und Atemwegserkrankungen

In den folgenden Tabellen wird als Wert für die Schadstoffexposition immer der Mittelwert der Exposition über den gesamten Zeitraum angegeben.

3.3.1 Allergischer Schnupfen

Tabelle 4 zeigt die mittlere *Schadstoffexposition am Arbeitsplatz* von Probanden, die an einer *Allergie* erkrankt sind im Vergleich zur Exposition derjenigen Probanden, die nicht erkrankt sind.

Es gibt keine statistisch signifikanten Unterschiede zwischen der Schadstoffexposition von Probanden mit Allergie im Vergleich zur Schadstoffexposition von gesunden Probanden.

Exposition	Allergie ja	Allergie nein
biologische Expo.	12.27	13.80
mineralische Expo.	24.99	29.31
Gas/Rauchexpo.	33.53	30.87
Passivrauchexpo.	0.52	0.42

Tabelle 4: Zusammenhang zwischen Allergie und Schadstoffexposition am Arbeitsplatz

3.3.2 Rhinitis

Tabelle 5 zeigt die mittlere *Schadstoffexposition am Arbeitsplatz* von Probanden, die an *Rhinitis* erkrankt sind im Vergleich zur Exposition derjenigen Probanden, die nicht erkrankt sind.

Es gibt keine statistisch signifikanten Unterschiede zwischen der Schadstoffexposition von Probanden mit Rhinitis im Vergleich zur Schadstoffexposition von gesunden Probanden.

Exposition	Rhinitis ja	Rhinitis nein
biologische Expo.	12.11	14.29
mineralische Expo.	27.48	28.80
Gas/Rauchexpo.	28.46	33.44
Passivrauchexpo.	0.51	0.46

Tabelle 5: Zusammenhang zwischen Rhinitis und Schadstoffexposition am Arbeitsplatz

In Tabelle 6 wird die *Schadstoffexposition* bei gesunden und an *Rhinitis* erkrankten Probanden getrennt nach *Untersuchungszentrum* betrachtet.

Es treten bei der Schadstoffbelastung von an Rhinitis erkrankten Probanden signifikante Unterschiede zwischen Hamburg und Erfurt auf. In Erfurt ist bei den kranken Untersuchungsteilnehmern die biologische und mineralische Staubexposition sowie die Gas/Rauchexposition im Vergleich zu Hamburg signifikant erhöht.

Expo. bei Rhinitis ja/nein	Hamburg	Erfurt
biologische Expo.		
Rhinitis ja	5.43	22.74
Rhinitis nein	15.07	13.64
mineralische Expo.		
Rhinitis ja	19.24	40.58
Rhinitis nein	31.49	26.56
Gas/Rauchexpo.		
Rhinitis ja	21.35	39.77
Rhinitis nein	38.70	29.07
Passivrauchexpo.		
Rhinitis ja	0.46	0.47
Rhinitis nein	0.58	0.46

Tabelle 6: Unterschiede der Exposition bei Rhinitis zwischen Hamburg und Erfurt

4 Non-Response Analyse

In einer Non-Response Analyse soll nun untersucht werden, aus welchen Gründen es zwischen der ersten und der zweiten Studie zu Ausfällen bei den Fallzahlen kam, ob diese Ausfälle systematisch waren und in welcher Art man eine eventuelle Systematik bei der späteren Analyse und Modellbildung berücksichtigen muss.

Wie man Tabelle 7 entnehmen kann, haben bei der zweiten Untersuchung im Jahr 2002 nicht mehr alle Probanden der ersten Untersuchung teilgenommen.

	Erfurt	Hamburg	Erfurt und Hamburg
Anzahl Probanden			
response (= 1. und 2. Untersuchung)	287	303	590
nonresponse (= nur 1. Untersuchung)	445	956	1401
alle	732	1259	1991
response (in %)	39.2	24.1	29.6

Tabelle 7: Übersicht über das Antwortverhalten

Hier stellt sich nun die Frage, ob sich die Probanden, die nicht geantwortet haben, von den Antwortern unterscheiden. Dieser Punkt soll vor allem klären, ob bei den Ergebnissen aus unserer Stichprobe dadurch eine Verzerrung auftreten kann.

Deshalb werden an Hand der ersten Untersuchung Unterschiede in der Verteilung bestimmter für uns wichtiger Variablen zwischen den Antwortern und den Nichtantwortern betrachtet.

Getestete Variablen waren (nach Zentrum getrennt betrachtet): *Alter*, *Geschlecht*, *Rauchen*, *Bildungsstand*, *allergischer Schnupfen*, *allergischer Schnupfen in der Familie*, *Asthma in der Familie*, *Passivrauchen* (im Jahr 1991) und die drei Schadstoffexpositionen *biologischer Staub*, *mineralischer Staub* und *Gas/Rauch* (Querschnitt zum Jahr 1991).

Bei zwei Variablen ergaben sich signifikante Unterschiede in der Verteilung zwischen Antwortern und Nichtantwortern (siehe Tabelle 2 und 3).

In Erfurt liegt bei den Nichtantwortern eine andere Verteilung (Signifikanz-Niveau=95%) beim *Rauchen* vor als bei den Antwortern (getestet mit Hilfe Fisher's Vierfelder-Tafel). Auch bei der Variable *Alter* unterscheiden sich die

	Erfurt	Hamburg
in %		
response	58.9	67.7
nonresponse	67.6	67.9

Tabelle 8: Verteilung der Raucher

	Erfurt	Hamburg
arithmetisches Mittel		
response	33.02	35.06
nonresponse	32.00	33.07

Tabelle 9: Altersverteilung

beiden Verteilungen der Antwortter und Nichtantwortter , hier sowohl in Erfurt als auch in Hamburg (getestet mit t-Test). Hierbei ist allerdings anzumerken, dass möglicherweise die nötige Trennschärfe in den Daten nicht gegeben ist, um dieses Ergebnis als aussagefähig bezeichnen zu können. Einen Hinweis auf diese Problematik gibt die Tatsache, dass sich in Erfurt die Konfidenzintervalle für den Test auf Gleichheit des mittleren Alters in den beiden Gruppen *Antwortter* und *Nichtantwortter* überschneiden.

5 Allgemeines zur Modellwahl

Bei unserem Praktikumsthema liegen Daten eines Fragebogens zu allergischem Schnupfen, Rhinitis und der Schadstoffkonzentration bzgl. drei verschiedener Schadstoffarten am Arbeitsplatz und zudem Informationen über zahlreiche potentielle Confounder vor. Die Daten stammen von insgesamt 590 Probanden aus den beiden Untersuchungszentren Erfurt und Hamburg. Im Folgenden sollen Zusammenhangsstrukturen zwischen den Schadstoffexpositionen am Arbeitsplatz und Atemwegserkrankungen untersucht werden. Dabei betrachten wir ausschließlich die Erkrankungen *Rhinitis* und *allergischer Schnupfen*, da die Fallzahlen bei den anderen Krankheiten für die Anwendung der meisten Verfahren zu gering sind.

Zur Untersuchung des Zusammenhangs zwischen Krankheiten und Schadstoffexposition wird ein logistisches Regressionsmodell gewählt, da die Zielgrößen *allergischer Schnupfen* und *Rhinitis* jeweils in dichotomer Skalierung vorlie-

gen.

6 Umkodierungen

Um die kategorialen Variablen in die Regression einbinden zu können, wurden die in Tabelle 10 aufgezählten Dummy-Kodierungen vorgenommen.

Variable	Referenzkategorie
Geschlecht	Männer
Bildungsstand	niedrige Bildung
Wohnort bis 5. LJ	städtische Umgebung
Rauchen	Nie geraucht
Passivrauchen	kein Passivrauch
Allergie in der Familie	keine Vorbelastung
Asthma in der Familie	keine Vorbelastung

Tabelle 10: Dummykodierung der Variablen

Die Kodierung der Passivrauchexposition wurde schon in Kapitel 2.2.3 erläutert. Außerdem liegen bei den drei Schadstoffvariablen (*biologische, mineralische Staub- und Gas-/Rauch-Exposition*) rechtsschiefe Verteilungen vor, so dass der Median in allen Fällen bei Null liegt. Aus diesem Grund werden diese Variablen kategorisiert, um sie sinnvoll in die Analyse einbinden zu können. Zur Kategorisierung der drei Schadstoffexpositionen wird nach dem folgenden Prinzip vorgegangen: Als Referenzkategorie werden alle Beobachtungen mit Exposition 0 verwendet. Der jeweilige Rest der Beobachtungen wird in Terzile aufgeteilt und mit den Zahlen 1, 2, 3 kodiert. Hierbei entstehen für jede der drei Schadstoffarten jeweils drei Expositionsvariablen mit den Namen BDUMMYU, BDUMMYM, BDUMMYO; MDUMMYU, MDUMMYM, MDUMMYO; GDUMMYU, GDUMMYM, GDUMMYO. Bei diesen Variablennamen stehen die Anfangsbuchstaben für biologische, mineralische Staub- bzw. Gas-/Rauch-Exposition und die Endbuchstaben für unteres, mittleres bzw. oberes Terzil.

Die vorgenommene Kategorisierung nach Terzilen wird in Tabelle 11 veranschaulicht.

Kategorie	0	1	2	3
	Exposition = 0	x_{33}	x_{66}	x_{100}
bgesamt_t	86.95%	4.07%	4.57%	4.41%
mgesamt_t	74.58%	8.47%	8.48%	8.47%
ggesamt_t	73.90%	8.81%	8.82%	8.47%

Tabelle 11: Verteilung der Expositionsvariablen

7 Der Weg vom Ausgangsmodell zum Endmodell

7.1 Ausgangsmodell

Im Folgenden sind in Klammern die SAS-Variablenamen angegeben.

Als Anfangsmodell wird für die beiden Zielgrößen *Rhinitis* und *allergischer Schnupfen* jeweils ein Modell mit den verschiedenen Expositionen (BDUMMYU, BDUMMYM, BDUMMYO; MDUMMYU, MDUMMYM, MDUMMYO; GDUMMYU, GDUMMYM, GDUMMYO), allen inhaltlich sinnvoll erscheinenden unabhängigen Variablen und einigen Interaktionen verwendet.

Die relevanten unabhängigen Variablen sind *Alter* (EL03ALT), *Geschlecht* (SEXDUMMY), *Bildungsstand* (BILDUMMY), *Passivrauchen* (PDUMMY1, PDUMMY2, PDUMMY3), *Rauchen* (RAUDUMMY), *Wohnort bis zum fünften Lebensjahr* (ORTDUM1 und ORTDUM2), *genetische Prädisposition für allergischen Schnupfen* (ATOP_ALL) und *genetische Prädisposition für Asthma* (ATOP_AST).

Außerdem werden Interaktionen zwischen dem *Zentrum* und den *Expositionen* und dem *Geschlecht* und den *Expositionen* verwendet. Sind diese Interaktionen nicht signifikant, so werden sie aus dem Modell entfernt.

7.2 Interaktionen

Da die Fallzahlen sowohl von *Rhinitis* als auch von *Allergie* bei dieser Untersuchung sehr klein sind, haben wir weitgehend auf die Aufnahme von Interaktionen in unser Modell verzichtet.

A priori wird lediglich auf Interaktionen zwischen dem *Geschlecht* und den *Expositionen* und dem *Zentrum* und den *Expositionen* getestet. Da sich aber in den Ausgangsmodellen keine der überprüften Interaktionen als signifikant

erweisen, werden diese im weiteren Vorgehen nicht mehr berücksichtigt.

7.3 Confounder

Die Wohnortsvariablen (*Wohnort bis zum fünften Lebensjahr*) werden als mögliche Confounder für beide Krankheiten näher betrachtet. Hierzu wird die Modellgüte des Ausgangsmodells (nun stets ohne Interaktionen) mit der Güte desselben Modells ohne Ortsvariablen verglichen (siehe Tabelle 12). Außerdem wird der Erklärungswert eines Modells, welches nur die Ortsvariablen enthält, untersucht (siehe Tabelle 13).

Rhinitis-Modell		
mit Ort	nur Intercept	Intercept und Kovariaten
AIC	680.087	685.181
ohne Ort	nur Intercept	Intercept und Kovariaten
AIC	680.087	682.458
Allergie-Modell		
mit Ort	nur Intercept	Intercept und Kovariaten
AIC	571.600	543.829
ohne Ort	nur Intercept	Intercept und Kovariaten
AIC	571.600	546.176

Tabelle 12: SAS Model Fit Statistics: Ausgangsmodell mit/ohne Ort

Rhinitis-Modell		
	nur Intercept	Intercept und Kovariaten
AIC	790.044	790.158
Allergie-Modell		
	nur Intercept	Intercept und Kovariaten
AIC	648.020	640.721

Tabelle 13: SAS Model Fit Statistics: Modell nur mit Ort

In dem Modell für *allergischen Schnupfen* tragen die Ortsvariablen zu einem besseren Erklärungswert des Modells bei und werden deshalb im Modell behalten. Im *Rhinitis*-Modell ändert sich der Erklärungswert des Modells mit Ortsvariablen kaum; er wird nur geringfügig schlechter. Auch im Modell für

Rhinitis wird die Ortsvariable beibehalten, da auf diese Weise beide Krankheitsmodelle besser miteinander verglichen werden können.

7.4 Effektmodifizierer

Die Variablen *Zentrum* und *Geschlecht* werden als mögliche Effektmodifizierer angesehen. Daher werden die Daten nach diesen beiden Variablen stratifiziert betrachtet. Dieses Vorgehen führt allerdings bei einigen wichtigen Variablen, wie z.B. der biologischen Schadstoffexposition, zu riesigen Konfidenzintervallen für die Odds Ratio-Schätzer. Damit ist offensichtlich die Gültigkeit des Modells nicht mehr gewährleistet.

Aus diesem Grund werden die Daten nun jeweils nur nach einer der beiden Variablen, entweder *Zentrum* oder *Geschlecht*, stratifiziert.

Bei der Zielgröße *allergischer Schnupfen* sind auch nach der getrennten Stratifizierung sowohl bei der Aufteilung nach *Zentrum* als auch bei der Aufteilung nach *Geschlecht* die Konfidenzintervalle für die Odds Ratio-Schätzer immer noch zu groß und die Modellgültigkeit geht verloren. Aus diesem Grund wird beim Modell für den *allergischen Schnupfen* vollkommen auf Stratifizierung verzichtet.

Bei der Zielgröße *Rhinitis* ist es möglich, ein nach *Geschlecht* stratifiziertes und ein nach *Zentrum* stratifiziertes Modell zu berechnen. Die Odds Ratio-Schätzer lassen auf den ersten Blick allerdings keine so großen Unterschiede zwischen den jeweiligen beiden Straten erkennen, als dass man *Geschlecht* oder *Zentrum* als Effektmodifizierer ausmachen könnte. Eine weitere Diskussion dieses Punktes folgt im Kapitel über die Interpretation der Modelle.

7.5 Endmodelle

Bei unserem Endmodell für die Zielgröße *Rhinitis* wurden die Interaktionen zwischen den *Expositionen* und dem *Geschlecht* sowie zwischen den *Expositionen* und dem *Zentrum* weggelassen. Des Weiteren wurden die Daten bei dieser Zielgröße nicht nur insgesamt (siehe Tabellen 14 und 15), sondern auch einmal nach *Zentrum* (siehe Tabellen 16 und 17 bzw. Tabellen 18 und 19) und einmal nach *Geschlecht* (siehe Tabellen 20 und 21 bzw. Tabellen 22 und 23) stratifiziert betrachtet. Die unabhängigen Variablen im Endmodell für *Rhinitis* sind:

biologische-, mineralische- und Gas/Rauch-Exposition, Alter, Geschlecht, Bildungsstand, Rauchen, Passivrauchen, Wohnort bis zum fünften Lebensjahr sowie genetische Prädisposition für Allergie und Asthma.

	nur Intercept	Intercept und Kovariaten
AIC	680.087	685.181

Tabelle 14: Model Fit Rhinitis

Das Endmodell für die Zielgröße *allergischer Schnupfen* unterscheidet sich vom Anfangsmodell lediglich dadurch, dass auch hier die Interaktionen weggelassen wurden (siehe Tabellen 24 und 25). Somit sind die unabhängigen Variablen im Endmodell für *allergischen Schnupfen* die folgenden: *biologische-, mineralische- und Gas/Rauch-Exposition, Alter, Geschlecht, Bildungsstand, Rauchen, Passivrauchen, Wohnort bis zum fünften Lebensjahr* sowie *genetische Prädisposition für Allergie und Asthma*.

Effekt	Punktschätzer	untere KI-Grenze	obere KI-Grenze
BDUMMYU	1.271	0.405	3.986
BDUMMYM	1.077	0.414	2.804
BDUMMYO	0.670	0.247	1.820
MDUMMYU	1.092	0.423	2.820
MDUMMYM	1.256	0.520	3.034
MDUMMYO	1.073	0.418	2.756
GDUMMYU	0.564	0.232	1.372
GDUMMYM	1.258	0.551	2.872
GDUMMYO	0.564	0.208	1.528
ELO3ALT	0.980	0.953	1.007
SEXDUMMY	1.130	0.753	1.695
ZENTRUM	0.636	0.433	0.935
BILDUMMY	0.761	0.505	1.147
PDUMMY1	0.833	0.523	1.325
PDUMMY2	0.970	0.486	1.936
PDUMMY3	0.868	0.429	1.757
RAUDUMMY	1.030	0.691	1.535
ATOP_ALL	2.227	1.502	3.301
ATOP_AST	1.023	0.580	1.802
ORTDUM1	0.759	0.439	1.311
ORTDUM2	0.773	0.357	1.674

Tabelle 15: Odds Ratio-Schätzer für Rhinitis

	nur Intercept	Intercept und Kovariaten
AIC	323.612	340.647

Tabelle 16: Model Fit Rhinitis in Erfurt

Effekt	Punktschätzer	untere KI-Grenze	obere KI-Grenze
BDUMMYU	0.723	0.160	3.273
BDUMMYM	0.821	0.236	2.856
BDUMMYO	1.677	0.492	5.709
MDUMMYU	1.630	0.443	5.993
MDUMMYM	2.891	0.847	9.868
MDUMMYO	1.841	0.492	6.887
GDUMMYU	0.597	0.178	1.999
GDUMMYM	1.357	0.435	4.238
GDUMMYO	0.706	0.166	3.011
ELO3ALT	0.987	0.945	1.031
SEXDUMMY	1.466	0.788	2.727
BILDUMMY	0.497	0.256	0.965
PDUMMY1	1.008	0.516	1.966
PDUMMY2	1.271	0.424	3.814
PDUMMY3	0.992	0.335	2.938
RAUDUMMY	0.800	0.440	1.457
ATOP_ALL	2.071	1.111	3.860
ATOP_AST	1.064	0.419	2.702
ORTDUM1	0.561	0.266	1.185
ORTDUM2	0.420	0.138	1.276

Tabelle 17: Odds Ratio-Schätzer für Rhinitis in Erfurt

	nur Intercept	Intercept und Kovariaten
AIC	349.757	356.487

Tabelle 18: Model Fit Rhinitis in Hamburg

Effekt	Punktschätzer	untere KI-Grenze	obere KI-Grenze
BDUMMYU	7.857	0.546	113.014
BDUMMYM	1.319	0.222	7.827
BDUMMYO	0.110	0.011	1.075
MDUMMYU	0.701	0.154	3.184
MDUMMYM	0.459	0.109	1.929
MDUMMYO	0.503	0.118	2.148
GDUMMYU	0.338	0.074	1.551
GDUMMYM	1.381	0.371	5.141
GDUMMYO	0.672	0.158	2.848
ELO3ALT	0.976	0.939	1.015
SEXDUMMY	0.737	0.412	1.319
BILDUMMY	1.063	0.598	1.888
PDUMMY1	0.691	0.346	1.380
PDUMMY2	0.732	0.294	1.825
PDUMMY3	0.639	0.242	1.691
RAUDUMMY	1.072	0.599	1.918
ATOP_ALL	2.895	1.648	5.086
ATOP_AST	0.953	0.444	2.049
ORTDUM1	1.163	0.463	2.918
ORTDUM2	1.594	0.436	5.826

Tabelle 19: Odds Ratio-Schätzer für Rhinitis in Hamburg

	nur Intercept	Intercept und Kovariaten
AIC	336.771	346.270

Tabelle 20: Model Fit Rhinitis bei Männern

Effekt	Punktschätzer	untere KI-Grenze	obere KI-Grenze
BDUMMYU	0.976	0.141	6.741
BDUMMYM	0.767	0.198	2.967
BDUMMYO	0.537	0.159	1.805
MDUMMYU	1.211	0.389	3.768
MDUMMYM	1.626	0.569	4.645
MDUMMYO	1.231	0.401	3.786
GDUMMYU	0.578	0.196	1.700
GDUMMYM	1.202	0.384	3.761
GDUMMYO	0.321	0.093	1.104
ELO3ALT	0.939	0.901	0.980
ZENRUM	0.540	0.305	0.956
BILDUMMY	0.771	0.419	1.418
PDUMMY1	0.901	0.455	1.784
PDUMMY2	1.034	0.391	2.734
PDUMMY3	0.814	0.328	2.024
RAUDUMMY	1.359	0.722	2.558
ATOP_ALL	2.326	1.281	4.225
ATOP_AST	0.947	0.385	2.330
ORTDUM1	0.639	0.274	1.492
ORTDUM2	1.165	0.360	3.766

Tabelle 21: Odds Ratio-Schätzer für Rhinitis bei Männern

	nur Intercept	Intercept und Kovariaten
AIC	343.876	362.877

Tabelle 22: Model Fit Rhinitis bei Frauen

Effekt	Punktschätzer	untere KI-Grenze	obere KI-Grenze
BDUMMYU	1.634	0.364	7.342
BDUMMYM	1.371	0.274	6.861
BDUMMYO	1.488	0.199	11.112
MDUMMYU	2.097	0.134	32.746
MDUMMYM	0.685	0.077	6.050
MDUMMYO	0.254	0.029	2.233
GDUMMYU	0.307	0.027	3.438
GDUMMYM	1.696	0.435	6.613
GDUMMYO	2.841	0.364	22.169
ELO3ALT	1.018	0.978	1.060
ZENTRUM	0.720	0.413	1.253
BILDUMMY	0.830	0.458	1.504
PDUMMY1	0.763	0.385	1.513
PDUMMY2	0.822	0.286	2.364
PDUMMY3	0.832	0.236	2.934
RAUDUMMY	0.970	0.563	1.673
ATOP_ALL	2.366	1.342	4.173
ATOP_AST	1.167	0.525	2.593
ORTDUM1	0.916	0.426	1.968
ORTDUM2	0.471	0.157	1.412

Tabelle 23: Odds Ratio-Schätzer für Rhinitis bei Frauen

	nur Intercept	Intercept und Kovariaten
AIC	571.600	543.829

Tabelle 24: Model Fit Allergischer Schnupfen

Effekt	Punktschätzer	untere KI-Grenze	obere KI-Grenze
BDUMMYU	1.333	0.310	5.727
BDUMMYM	1.525	0.486	4.783
BDUMMYO	0.859	0.246	2.998
MDUMMYU	0.842	0.268	2.645
MDUMMYM	1.209	0.419	3.483
MDUMMYO	0.572	0.171	1.915
GDUMMYU	0.829	0.283	2.427
GDUMMYM	1.001	0.357	2.807
GDUMMYO	1.659	0.527	5.226
ELO3ALT	0.981	0.950	1.014
SEXDUMMY	1.313	0.811	2.128
ZENTRUM	0.794	0.488	1.291
BILDUMMY	0.791	0.485	1.290
PDUMMY1	0.834	0.480	1.449
PDUMMY2	1.449	0.649	3.234
PDUMMY3	0.556	0.217	1.423
RAUDUMMY	0.637	0.402	1.008
ATOP_ALL	4.136	2.620	6.530
ATOP_AST	0.722	0.368	1.415
ORTDUM1	0.640	0.328	1.249
ORTDUM2	0.273	0.077	0.973

Tabelle 25: Odds Ratio-Schätzer für allerg. Schnupfen

8 Interpretation der Modelle

In diesem Abschnitt beschäftigen wir uns mit der inhaltlichen Interpretation unserer Modelle.

In dem Gesamtmodell für *Rhinitis* ist zu sehen, dass gegenüber *allergischem Schnupfen* genetisch prädisponierte Probanden ein höheres Krankheitsrisiko haben. Außerdem gibt es einen Unterschied, je nachdem aus welchem Untersuchungszentrum die Person kommt. Kommt sie aus Erfurt, so ist das Erkrankungsrisiko im Vergleich zu einer Person aus Hamburg deutlich niedriger.

In ähnlicher Weise verhält sich der Einfluss einer genetischen Prädisposition gegenüber allergischem Schnupfen in dem Gesamtmodell für *allergischen Schnupfen*. Hier ist das Risiko allerdings stärker erhöht als bei *Rhinitis*. Dies erscheint plausibel, da es sich hier um die für die Zielvariable *allergischer Schnupfen* spezifische Atopie handelt. Zudem sinkt das Allergierisiko, sowohl wenn die ersten fünf Lebensjahre auf einem Bauernhof verbracht wurden als auch wenn die untersuchte Person nie geraucht hat.

Bei den nach dem Zentrum stratifizierten Modellen für *Rhinitis* ist für die genetische Prädisposition jeweils der selbe Effekt zu sehen wie im Gesamtmodell. In Hamburg ist außerdem auffällig, dass hohe Werte der biologischen Exposition scheinbar das Krankheitsrisiko senken. Dieser Effekt heißt 'Healthy-Worker-Effekt' und besagt, dass Beschäftigte, die auf Grund von starker Schadstoffexposition am Arbeitsplatz erkranken, versuchen, den Arbeitsplatz zu wechseln. In Erfurt kann dieser Effekt vermutlich deshalb nicht beobachtet werden, weil viele Arbeitnehmer aus Angst keine neue Stelle zu bekommen, den Arbeitsplatz nicht wechseln oder ihn nicht wechseln können. Bei den nach dem Geschlecht stratifizierten Modellen für *Rhinitis* ist zu sehen, dass bei Frauen der Effekt der biologischen Exposition auf das Krankheitsrisiko höher ist als bei Männern. Ein möglicher Schluss daraus wäre, dass Frauen empfindlicher auf biologische Schadstoffe reagieren als Männer. Interessante Ergebnisse könnten sich außerdem ergeben, wenn man mit größeren Fallzahlen zweimal -zuerst nach Zentrum und dann jeweils nach Geschlecht-stratifizierte Modelle berechnen und vergleichen könnte.

9 Longitudinale Betrachtung

9.1 Vereinfachte longitudinale Betrachtung

Bei der vereinfachten longitudinalen Betrachtung beziehen wir die Information mit ein, dass nicht nur die Höhe der Exposition zum Zeitpunkt der zweiten Untersuchung bekannt ist, sondern - bei den im Zeitraum zwischen 1991 und 2001 Erkrankten - auch die Höhe der Exposition zum Zeitpunkt ihrer Erkrankung. Deshalb stutzen wir jetzt die Expositionswerte auf diesen eigentlich relevanten Wert und berechnen mit diesen gekürzten Expositionsvariablen das selbe Allergie-Modell wie in Kapitel 7.5 noch einmal.

Hierbei entstehen im Gegensatz zum bisherigen Allergie-Modell riesige Konfidenzintervalle und die Maximum-Likelihood-Schätzer können auf Grund numerischer Probleme nicht mehr geschätzt werden. Somit ist dieses Modell nicht sinnvoll interpretierbar.

Eine anschließende Berechnung mit nach Median statt nach Terzilen aufgeteilten Expositionsvariablen brachte keine Verbesserung der Ergebnisse.

9.2 Survivalanalyse

9.2.1 Konstruktion des Datensatzes

In dieser Auswertungsstufe wird die longitudinale Datenstruktur miteinbezogen: Für den allergischen Schnupfen liegen die Erkrankungsjahre vor und aus den Expositionen und der Berufsmatrix lassen sich die jeweiligen Expositionen zum 31.12. der Jahre 1992-2001 konstruieren. Dann kann für alle 590 Individuen für jedes der zehn Jahre ein Datensatz erstellt werden, der - neben den zeitunabhängigen Variablen - auch die Information enthält, ob der Proband bis zum 31.12. des jeweiligen Jahres erkrankt ist, und außerdem wie groß seine Expositionswerte in diesem Jahr sind. Des weiteren enthält jeder Datensatz eine konstante Variable *Jahr*, die später im Gesamtdatensatz eine Identifikation der einzelnen Jahre ermöglicht. Anschließend werden diese zehn Jahresdatensätze untereinandergehängt. Da im Anschluss eine Survivalanalyse durchgeführt werden soll, werden die Probanden, die vor Beginn der Studie schon krank waren, aus dem Datensatz entfernt. Zudem wird jede Person nur solange 'beobachtet', bis sie krank wird, d.h. in den Folgejahren nach dem Auftreten der Krankheit werden die Beobachtungen des entsprechenden Probanden aus dem Datensatz gelöscht. Der fertige Datensatz auf dem alle folgenden Berechnungen beruhen enthält 4728 Beobachtungen (vgl. Tabelle 26).

id	Jahr	Allergie	Bio	Staub	Gas	Alter	zeitunabh. Varn.
1	9	0	12	6	0		...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
485	9	0	6	12	6		...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	0	0	52	34	66		...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
456	0	1	18	24	6		...

Tabelle 26: Struktur des Longitudinal-Datensatzes

9.2.2 Berechnung des Survivalmodells

In dem Survivalmodell ist die abhängige Variable eine binäre Größe, die angibt, ob ein Proband in dem jeweils beobachteten Jahr krank wird oder nicht. Als Einflussgrößen werden dieselben Variablen wie in den vorangegangenen logistischen Regressionsmodellen und zudem eine kategoriale Variable *Jahr* mit den Ausprägungen 9(1992),...,0(2001) verwendet. Das *Alter* geht zeitabhängig ins Modell ein.

Berechnet wird ein logistisches Regressionsmodell der Form

$$P(\text{allergie}_t = 1 | \text{allergie}_{t-1} = 0, \dots, \text{allergie}_1 = 0) = \text{hazard}(\alpha_t + x' \beta)$$

Es wird also die Wahrscheinlichkeit modelliert, mit der ein Proband im Jahr t erkrankt. α_t gibt die sogenannte „base-line“, eine Art Intercept, für das Jahr t an und β ist ein Vektor, der die anderen Regressionsparameter enthält. Als Link-Funktion wird zunächst die cloglog-Funktion verwendet. Wegen der kleinen Fallzahlen (31 Erkrankungen im Beobachtungszeitraum) ergeben sich aber mit der logit-Funktion approximativ dieselben Parameterschätzungen, weshalb die Berechnungen letztendlich mit der logit Funktion als Link-Funktion durchgeführt wurden.

Um die Plausibilität des Modells zu prüfen, werden im Folgenden einige alternative Modelle betrachtet.

Drei Alternativmodelle enthalten die Exposition des aktuellen Jahres. Die Expositionswerte werden also ab jetzt nicht mehr bis zum Erkrankungszeitpunkt aufaddiert. Diese Modelle unterscheiden sich dadurch, dass die Variablen *Alter* und *Jahr* jeweils auf unterschiedliche Art ins Modell eingehen: Entweder

- *Alter* als zeitunabhängige Variable

Oder

- *Jahr* als lineare Einflußgröße

Oder

- *Alter* als zeitunabhängige Variable und *Jahr* als lineare Einflußgröße

Vier weitere Alternativmodelle enthalten als Exposition nur die Exposition des Vorjahres.

Sie beinhalten neben den üblichen Variablen noch einen der folgenden Punkte:

Entweder

- *Alter* als zeitabhängige Variable und *Jahr* als kategoriale Variable (wie bei dem oben beschriebenen Ausgangsmodell)

Oder

- *Alter* als zeitunabhängige Variable

Oder

- *Jahr* als lineare Einflußgröße

Oder

- *Alter* als zeitunabhängige Variable und *Jahr* als lineare Einflußgröße

Da sich die Parameterschätzer in sämtlichen Alternativmodellen kaum von denen des zuerst beschriebenen Modells unterscheiden, werden die Zusatzmodelle nicht mehr weiter betrachtet.

9.2.3 Interpretation der Survivalanalyse

In den Ergebnissen des Survivalmodells ist -analog zu den früheren Endmodellen- zu sehen, dass gegenüber *allergischem Schnupfen* genetisch prädisponierte Probanden ein höheres Krankheitsrisiko haben.

Erstaunlicherweise hat die Variable Alter einen starken Effekt auf das Erkrankungsrisiko: Je älter der Proband ist, desto schwächer ist sein Risiko, zu erkranken. Lässt man die Variable Alter aus der Berechnung heraus, entstehen unrealistische Schätzer und sehr große Konfidenzintervalle, so dass keine sinnvolle Interpretation der Ergebnisse möglich ist.

Alle anderen Variablen haben keinen signifikanten Effekt auf das Erkrankungsrisiko.

Effekt	Punktschätzer	untere KI-Grenze	obere KI-Grenze
Jahr 9 vs 0	0.109	0.012	0.948
Jahr 8 vs 0	0.232	0.040	1.355
Jahr 7 vs 0	0.796	0.199	3.178
Jahr 6 vs 0	0.024	< 0.001	3.127
Jahr 5 vs 0	0.188	0.022	1.617
Jahr 4 vs 0	1.165	0.293	4.624
Jahr 3 vs 0	0.255	0.030	2.177
Jahr 2 vs 0	0.539	0.094	3.096
Jahr 1 vs 0	2.030	0.529	7.798
baktuell	0.944	0.818	1.091
maktuell	0.896	0.764	1.051
gaktuell	0.865	0.722	1.036
alter	0.884	0.864	0.905
EH22SEX 2 vs 1	1.319	0.605	2.878
ZENTRUM 2 vs 1	1.327	0.593	2.971
EH321 2 vs 1	0.592	0.249	1.409
PASSIVR 3 vs 0	3.928	1.216	12.689
PASSIVR 2 vs 0	2.015	0.449	9.041
PASSIVR 1 vs 0	1.436	0.608	3.394
M74 (Rauchen) 2 vs 1	1.056	0.473	2.357
ATOP_ALL 1 vs 0	3.225	1.442	7.212
ATOP_AST 1 vs 0	1.721	0.652	4.541
ORT 2 vs 0	0.930	0.180	4.800
ORT 1 vs 0	1.295	0.441	3.800

Tabelle 27: Odds Ratio-Schätzer für das Survivalmodell

Literatur

- Büsching, Kerstin (2003), *Berufsbedingte Atemwegsbeschwerden in Ost- und Westdeutschland - Eine Querschnittsstudie*, Dissertation. München.
- Blanc, P.D., Ellbjär, S., Janson, Ch., Norbäck, D., Norrman, E., Plaschke, P. and Torén, K. (1999). Asthma-related Work Disability in Sweden. *Am. J. Respir. Care Med.*, Vol. 160, 2028-2033.
- Dwyer, J.H., Feinleib, M., Lippert, P. and Hoffmeister, H. (1992). Statistical Models for Longitudinal Studies of Health. Chapter 7: *Use of the logistic and related models in longitudinal studies of chronic disease risk*, by Norman Breslow, New York: Oxford University Press.
- Clayton, David and Hills, Michael (1993). Statistical Models in Epidemiology. Chapter 27: *Choice and interpretation of models*, New York: Oxford University Press.
- Coding and Instructions for main questionnaire - ECRHSII, 2001.
- Cody, Ronald P. and Smith, Jeffrey K. (1997), *Applied Statistics and the SAS Programming Language*, Fourth Edition. New Jersey: Prentice Hall.
- Fragebogen zur ECRHSII, 2001.
- Gordis, Leon (2001). *Mehr über kausale Schlussfolgerungen: Bias, Confounding und Interaktionen* in Epidemiologie, Marburg: Verlag im Kilian.
- Heinrich, J. et al (2002). Die Europäische Studie zu Atemwegserkrankungen bei Erwachsenen (ECRHS). *Pneumologie*, Vol. 56, 297-303. Stuttgart, Thieme Verlag.
- Kleinbaum, David G. and Klein, Mitchel (2002). *Logistic Regression: A self-learning text*, Second Edition. New York: Springer.
- Radon, Katja, et al (2002). Passive Smoking Exposure, A Risk for Chronic Bronchitis and Asthma in Adults? *CHEST*, Vol. 122, Issue 3, 1086-1090.
- SAS Institute Inc. (1989). *SAS Guide to the SQL Procedure: Usage and Reference*, Version 6, First Edition. Cary, SAS Institute Inc.
- SAS Institute Inc. (1990). *SAS Procedures Guide*, Version 6, Third Edition. Cary, SAS Institute Inc.
- Ulm/Bickeböller (2001). *Skriptum zur Vorlesung „Epidemiologie“*. München.