

**Ein Markov-Modell zur Ermittlung der  
Kosten-Nutzen-Relation von Risikoscores zur  
Prognose der koronaren Herzkrankheit**

**Dissertation**

zur Erlangung des Grades eines  
Doktors der Naturwissenschaften  
der Universität Dortmund

angefertigt am Fachbereich Statistik  
der Universität Dortmund

und am Institut für Gesundheitsökonomie  
und Klinische Epidemiologie  
der Universität zu Köln

vorgelegt von  
**Björn Stollenwerk**  
geboren in Dortmund

eingereicht am Fachbereich Statistik  
im Juni 2006

1. Gutachter: Prof. Dr. Joachim Hartung
  2. Gutachter: Prof. Dr. Wolfgang Urfer
  3. Gutachter: Prof. Dr. Dr. Karl W. Lauterbach
- Tag der mündlichen Prüfung: 18.09.2006

# Danksagung

An dieser Stelle möchte ich allen Personen danken, die mich bei der Erstellung meiner Dissertation unterstützt haben. Herrn Professor Hartung danke ich für die Betreuung meiner Dissertation. Herrn Professor Lauterbach danke ich für die Überlassung des Themas sowie dafür, dass ich diverse Datensätze des Instituts für Gesundheitsökonomie und Klinische Epidemiologie nutzen durfte. Meinen Arbeitskollegen Mathias Schaller (Statistiker) und Andreas Gerber (Arzt, Theologe und Gesundheitsökonom) sowie den Statistikern Verena Dlugay und Guido Knapp danke ich für das Korrektur lesen der Arbeit. Für anregende Diskussionen bedanke ich mich bei Guido Knapp, Andreas Gerber, Lars Stollenwerk (Physiker), Markus Längen (Volkswirt), Gabriele Klever-Deichert (Volkswirtin) und Afschin Gandjour (Arzt, Gesundheitsökonom). Herrn Professor Urfer danke ich für die Zweitbegutachtung der Dissertation.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Statistische und gesundheitsökonomische Grundlagen</b>	<b>5</b>
2.1	Verallgemeinerungen des linearen Regressionsmodells . . . . .	5
2.1.1	Das verallgemeinerte lineare Modell (GLM) . . . . .	5
2.1.2	Das verallgemeinerte additive Modell (GAM) . . . . .	8
2.1.3	Überstreuung („over-dispersion“) . . . . .	9
2.1.4	Quasi-Likelihood . . . . .	10
2.1.5	Splines . . . . .	11
2.1.6	„Thin plate regression splines“ . . . . .	11
2.1.7	Darstellung von GAMs als GLMs . . . . .	15
2.1.8	Die Wahl des Glättungsparameters bei GAMs - verallgemeinerte Kreuzvalidierung . . . . .	17
2.1.9	Überprüfung der Modellvoraussetzungen . . . . .	19
2.2	Delaunay Triangulation . . . . .	20
2.3	Markov-Modellierung . . . . .	21

2.4	Gesundheitsökonomische Evaluationsverfahren . . . . .	24
2.5	Diskontierung . . . . .	26
2.6	Sensitivitätsanalyse . . . . .	27
<b>3</b>	<b>Entwicklung des Markovmodells</b>	<b>28</b>
3.1	Das Ausgangsmodell . . . . .	28
3.2	Unterteilung des Ausgangsmodells bezüglich des Erkrankungsrisikos .	30
3.3	Das Präventionsmodell für Personen mit höherem Risiko . . . . .	31
3.4	Das finale Markovmodell . . . . .	33
3.5	Die Wahl der Präventionsmaßnahme . . . . .	33
3.6	Einmaliges Screening versus multiples Screening . . . . .	35
<b>4</b>	<b>Schätzung der Übergangswahrscheinlichkeiten</b>	<b>36</b>
4.1	Notationen und Vorbemerkungen . . . . .	36
4.2	Übergangswahrscheinlichkeiten des Ausgangsmodells . . . . .	37
4.3	Erste Unterteilung der Personen mit koronarer Herzkrankheit (KHK)	40
4.4	Unterteilung der Gesunden nach KHK-Erkrankungsrisiko . . . . .	44
4.5	Zeitlicher Verlauf der KHK-spezifischen Mortalität . . . . .	46
4.6	Zeitlicher Verlauf nicht KHK-bedingter Mortalität . . . . .	46
4.7	Zweite Unterteilung der Personen mit KHK . . . . .	51
4.8	Übergangswahrscheinlichkeiten bei Primärprävention . . . . .	53
<b>5</b>	<b>Quantifizierung der Input-Größen zur Bestimmung der Übergangs-</b>	

---

<b>wahrscheinlichkeiten</b>	<b>58</b>
5.1 Verwendete Datensätze . . . . .	58
5.1.1 Die Daten des Bundes-Gesundheitssurveys 1998 . . . . .	58
5.1.2 Der SOEP-Datensatz . . . . .	59
5.1.3 Der RSA-Datensatz von sechs großen Krankenkassen . . . . .	59
5.1.4 Der RSA-Datensatz des Bundesversicherungsamtes (BVA) . . . . .	61
5.2 Die Prävalenz der koronaren Herzkrankheit . . . . .	61
5.2.1 Prävalenz der KHK aus dem Bundesgesundheitsurvey . . . . .	62
5.2.2 Prävalenz der KHK aus Kassendaten . . . . .	67
5.2.3 Die Prävalenz der KHK im Markovmodell . . . . .	67
5.3 Ursachenspezifische Mortalität . . . . .	68
5.4 Schätzung des Parameters $\lambda$ . . . . .	69
5.5 Mortalitätsspezifische Parameter . . . . .	71
5.5.1 Die Wahl von $\psi_t^{\text{sup}}$ . . . . .	73
5.5.2 Die Wahl von $\psi_i^{\text{sup}}$ . . . . .	80
5.5.3 Die Wahl von $\omega_k$ und $\omega_t$ . . . . .	80
5.5.4 Die Wahl von $\omega_i$ . . . . .	83
5.6 KHK-bedingte Herzinsuffizienz . . . . .	85
5.6.1 Modellierung der Prävalenz der Herzinsuffizienz . . . . .	87
5.6.2 Quantifizierung vermeidbarer Fälle von Herzinsuffizienz . . . . .	93
5.7 Effekt der Primärprävention . . . . .	98

<b>6</b>	<b>Zuweisung der Kosten und Nutzen</b>	<b>100</b>
6.1	Der Nutzen . . . . .	100
6.2	Zweige der Sozialversicherungen . . . . .	100
6.3	Die SOEP-Daten als Einnahmen- und Ausgabenquelle . . . . .	102
6.4	Einnahmen der Sozialversicherungen . . . . .	103
6.5	Ausgaben der Renten- und der Arbeitslosenversicherung . . . . .	107
6.6	Ausgaben der Pflegeversicherung . . . . .	114
6.7	Ausgaben der gesetzlichen Krankenversicherung . . . . .	114
6.7.1	GAM zur Modellierung der Gesundheitsausgaben . . . . .	116
6.7.2	Kostenproportionen . . . . .	123
6.7.3	Anwendung der Quotienten . . . . .	131
6.8	Screeningkosten . . . . .	133
6.9	Kosten der Statintherapie . . . . .	133
<b>7</b>	<b>Sensitivität und Spezifität des Modells</b>	<b>134</b>
7.1	Modell zur Bestimmung der Sensitivität und der Spezifität . . . . .	134
7.2	Steuerung der Sensitivität und der Spezifität im Markovmodell . . . . .	138
7.3	Berechnung der Sensitivität und der Spezifität anhand der Framing- hamrisiken . . . . .	139
7.4	Analyse der Sensitivität und Spezifität aus Framinghamrisiken . . . . .	141
<b>8</b>	<b>Durchführung der Kosten-Nutzwert-Analyse</b>	<b>145</b>
8.1	Evaluation des Markovmodells . . . . .	145

---

8.2	Kann sich Primärprävention lohnen? . . . . .	150
8.3	Die Wahl eines „optimalen“ Szenarios . . . . .	154
<b>9</b>	<b>Sensitivitätsanalyse</b>	<b>158</b>
9.1	Festlegung der Monte-Carlo-Simulation . . . . .	158
9.1.1	$\lambda$ . . . . .	158
9.1.2	$\psi_t^{\text{sup}}$ . . . . .	159
9.1.3	Transformation betaverteilter Zufallsvariablen . . . . .	159
9.1.4	$\psi_i^{\text{sup}}$ . . . . .	160
9.1.5	$\omega_i$ . . . . .	160
9.1.6	$\omega_k$ und $\omega_t$ . . . . .	162
9.1.7	Die Prävalenz der KHK . . . . .	162
9.1.8	$\delta_{\text{HI}}$ , $\mathcal{Q}_1$ und $\mathcal{Q}_2$ . . . . .	163
9.1.9	Die Lebensqualität bei KHK . . . . .	163
9.1.10	Die Compliance . . . . .	163
9.1.11	Therapiekosten . . . . .	164
9.1.12	Screeningkosten . . . . .	164
9.1.13	$\text{RR}^{\text{Präv}}$ . . . . .	165
9.1.14	ROC-Kurve . . . . .	165
9.1.15	Zahlungsbereitschaft . . . . .	166
9.2	Weitere Berechnungen . . . . .	166
9.3	Ergebnisse der Sensitivitätsanalyse . . . . .	168

<b>10 Zusammenfassung</b>	<b>171</b>
<b>A Weitere Gleichungen</b>	<b>175</b>
A.1 Gleichungen des Ausgangsmodells . . . . .	175
A.2 Gleichungen zur ersten Unterteilung der Personen mit KHK . . . . .	176
A.3 Gleichungen zur Unterteilung der Gesunden nach Erkrankungsrisiko .	177
A.4 Gleichungen zur zweiten Unterteilung der Personen mit KHK . . . . .	179
A.5 Gleichungen zur Primärprävention . . . . .	181
<b>B Tabellen</b>	<b>183</b>
<b>C Programmierung</b>	<b>193</b>
<b>Literaturverzeichnis</b>	<b>194</b>

# Kapitel 1

## Einleitung

Wenn die Begriffe „Gesundheit“ und „Ökonomie“ in einem Zusammenhang erwähnt werden, so weckt dies bei vielen Menschen Emotionen. Unter ethischen Aspekten wird es als problematisch angesehen, wenn medizinisch sinnvolle Maßnahmen aus Kostengründen nicht durchgeführt werden (Breyer und Zweifel 1997). Dennoch ist das Budget, das einer Gesellschaft für Gesundheitsausgaben zur Verfügung steht, in der Regel begrenzt. Das ökonomische Prinzip fordert, die zur Verfügung stehenden Ressourcen möglichst effektiv einzusetzen. Gesundheitsökonomische Analysen können helfen zu entscheiden, welche medizinischen Maßnahmen sinnvoller Weise durchgeführt werden sollten. Daher ist die Erstellung gesundheitsökonomischer Analysen sinnvoll, auch wenn methodische oder ethische Probleme nicht vollkommen ausgeräumt werden können.

Die vorliegende Arbeit befasst sich mit der Frage, ob es ökonomisch sinnvoll ist, gewisse Screeningverfahren anzuwenden. Es handelt sich dabei um Screeningverfahren, mit denen die Wahrscheinlichkeit ermittelt wird, dass eine Person innerhalb eines gegebenen Zeitraumes an der koronaren Herzkrankheit (KHK) erkrankt. Unter der koronaren Herzkrankheit wird eine Verengung der Herzkranzgefäße verstanden, durch die es zu einer Minderversorgung von Teilen des Herzmuskels mit Sauerstoff kommt. Die KHK kann sich z. B. durch einen Herzinfarkt oder durch Schmerzen

in der Brust (Angina Pectoris) äußern, sie kann aber auch symptomlos verlaufen (Pschyrembel 1990). Sie ist die häufigste Todesursache in den Industrieländern (Andrawes *et al.* 2005). Die KHK ist eine chronische, unheilbare Krankheit. Männer sind von ihr häufiger betroffen als Frauen. In Folge einer KHK kann es zu einer Herzinsuffizienz kommen (Hoppe *et al.* 2005).

Da viele Personen im Laufe ihres Lebens an der KHK erkranken und da die durch die KHK entstehenden Gesundheitsausgaben nicht unbeachtlich sind (Klever-Deichert *et al.* 2000), wurden bereits viele Modellierungen und gesundheitsökonomische Analysen zur KHK durchgeführt. Ein methodischer Klassiker ist das Markovmodell von Weinstein *et al.* (1987). In diesem Modell wird die sogenannte Framingham-Risikogleichung benutzt, um die Erkrankungswahrscheinlichkeit an der KHK abzuschätzen. In dieses Modell ist jedoch keine Prävention integriert, durch die die KHK vermieden werden soll.

Einige Analysen beschäftigen sich mit der Frage, ob es kosteneffektiv ist, Sekundärprävention durchzuführen. Unter Sekundärprävention wird Prävention bei bereits erkrankten Personen verstanden. Zu den Analysen mit Sekundärprävention gehören der Entscheidungsbaum von Maclaine und Patel (2001) und ein Markovmodell für Diabetiker von Hayashino *et al.* (2004). Auch Analysen ohne komplexe Modellierung wurden durchgeführt (Lindgren *et al.* 2005; Lange und Szucs 2004). In diesen wurden Kosten und Nutzen direkt in den zugrundeliegenden Studien erhoben, so dass keine Modellierung notwendig war.

Analysen zur Primärprävention, d. h. Prävention bei Personen ohne KHK, haben unter anderem Hay und Sterling (2005), Delea *et al.* (2005), O'Malley *et al.* (2004) und Szucs *et al.* (2004) durchgeführt. Von diesen Modellen bezieht sich nur das von Szucs *et al.* (2004) auf Deutschland. Die Analyse von Szucs *et al.* (2004) wird aus der Perspektive des deutschen Gesundheitssystems durchgeführt.

Als Präventionsmaßnahme erfolgt die Gabe von Atorvastatin und im Szenario ohne Prävention die Gabe von Placebo. Als Kostenfaktoren werden von Szucs *et al.* (2004) die Medikamentenkosten für Atorvastatin (ein Blutfett senkendes Medikament), die Kosten eines akuten Herzinfarktes sowie die Kosten eines Schlaganfalls berücksichtigt. Nicht berücksichtigt werden u. a. Krankheitskosten, die im weiteren Verlauf des Lebens anfallen. Es wird eine mittlere Therapiedauer von 3,3 Jahren betrachtet. Es ist sinnvoll, eine derartige Analyse aus der Sicht der gesetzlichen Krankenversicherung (GKV) durchzuführen. Für den politischen Entscheidungsprozess, ob Maßnahmen durchgeführt werden sollen, können jedoch auch Auswirkungen von Bedeutung sein, die nur indirekt auf die Therapie zurückzuführen sind. In umfangreicheren Analysen könnte berücksichtigt werden, dass Personen mit KHK auch bei anderen Krankheiten eine höhere Morbidität aufweisen. Auch die Auswirkungen einer Interventionsmaßnahme auf andere Zweige der Sozialversicherungen spielen im Entscheidungsprozess eine Rolle.

Eine Kosten-Effektivitätsanalyse zur Primärprävention der KHK aus Sicht der Sozialversicherungen wurde von Wendland (2002) durchgeführt. Prävention erfolgt dort durch die lebenslange Gabe von Statinen. In seinem Modell wird davon ausgegangen, dass jeweils eine gesamte Alters-Geschlechts-Kohorte als Präventionsmaßnahme Statine erhält. In dem Modell von Wendland (2002) kommt es an unterschiedlichen Stellen zu einer Überschätzung der Inzidenz der KHK, was die Primärprävention mit Statinen begünstigt.

Die bisherigen Modelle zur Primärprävention versuchen die Frage zu klären, ob sich Primärprävention unter ökonomischen Gesichtspunkten „lohnen“ kann, d. h. ob zusätzliche Kosten in einem „angemessenen“ Verhältnis zu zusätzlichem Nutzen stehen. Einige von diesen Modellen sind verhältnismäßig grob und sind daher lediglich richtungsweisend. In der vorliegenden Arbeit wird ein Modell entwickelt, mit dem viele unterschiedliche Alternativen verglichen werden können.

Erstens kann ermittelt werden, ob ein Szenario mit Primärprävention existiert, das unter ökonomischen Gesichtspunkten dem Szenario ohne Primärprävention vorzuziehen ist. Zweitens kann geklärt werden, wie bei einem Screeningverfahren der Trennpunkt, mit dem festgelegt wird, ob eine Person Primärprävention erhält, optimaler Weise gewählt werden sollte. Drittens können unterschiedliche Screeningverfahren miteinander verglichen werden. So kann geprüft werden, ob es sich unter ökonomischen Gesichtspunkten lohnt, ein teureres, aber dafür genaueres Screeningverfahren zu verwenden.

Die Kostenerhebung erfolgt aus der Perspektive der Sozialversicherungen. Primärprävention erfolgt durch eine lebenslange Statintherapie. Als Screeningverfahren wird die Framingham-Risikogleichung nach Anderson *et al.* (1991) verwendet. Das Modell ist so ausgelegt, dass auch weitere Screeningverfahren darauf angewendet werden können. So ist es z. B. möglich, das Modell auf sich noch in Entwicklung befindende Risikoscores nach ihrer Fertigstellung zu übertragen.

In Kapitel 2 werden die für das Modell und für die Analyse benötigten statistischen und gesundheitsökonomischen Grundlagen dargelegt. In Kapitel 3 wird die Struktur des zugrundeliegenden Markovmodells entwickelt und erläutert. Wie die Übergangswahrscheinlichkeiten des Markovmodells geschätzt werden, wird in Kapitel 4 beschrieben. Dazu werden bestimmte Größen benötigt, die in Kapitel 5 quantifiziert werden. Den einzelnen Zuständen des Markovmodells werden jeweils Kosten und Nutzen zugeordnet, was in Kapitel 6 beschrieben wird. Wie die Sensitivität und die Spezifität anhand des Modells oder anhand von Risikoscores ermittelt werden kann, ist in Kapitel 7 beschrieben. Die Evaluation des Markovmodells wird in Kapitel 8 erörtert. Um die Unsicherheit der Analyse abzuschätzen erfolgt in Kapitel 9 eine Sensitivitätsanalyse. Die Arbeit wird in Kapitel 10 zusammengefasst.

# Kapitel 2

## Statistische und gesundheitsökonomische Grundlagen

### 2.1 Verallgemeinerungen des linearen Regressionsmodells

#### 2.1.1 Das verallgemeinerte lineare Modell (GLM)

Viele statistische Probleme lassen sich mit Hilfe von Regressionsmodellen lösen. Dabei wird eine beobachtete Responsevariable  $y$  in Abhängigkeit von bekannten Kovariablen  $x_1, \dots, x_P$  modelliert. Sowohl die Responsevariable als auch die Kovariablen sind Vektoren der Länge  $N$  mit  $P \leq N$ . Am weitesten verbreitet ist das lineare Regressionsmodell (vgl. Hartung und Elpelt (1999) und Wood (2006)), in dem der Vektor  $y$  als Realisation eines normalverteilten Zufallsvektors  $Y$  angesehen wird.

Die Annahme im linearen Regressionsmodell, dass  $Y$  normalverteilt ist, ist häufig nicht gerechtfertigt. Es kommt z. B. vor, dass der Wertebereich der Komponenten  $Y_i$ ,  $i = 1, \dots, N$ , von  $Y$  nur einen Teilbereich der reellen Zahlen  $\mathbb{R}$  umfasst. Besser ist

es, ein verallgemeinertes (generalisiertes) lineares Modell (GLM) zu verwenden. In diesem Modell werden auch andere Verteilungen als die Normalverteilung zugelassen. Es wird gefordert, dass die Verteilungsfunktion der Responsevariable  $Y$  zu einer Exponentialfamilie gehört und sich ihre Dichtefunktion folgendermaßen darstellen lässt:

$$f_{\theta}(y) = \exp [(y \cdot \theta - b(\theta))/a(\phi) + c(y, \phi)] , \quad (2.1)$$

mit beliebigen Funktionen  $b$ ,  $a$  und  $c$  sowie den Parametern  $\phi \in \mathbb{R}^+$  und  $\theta \in \mathbb{R}$  (Hastie und Tibshirani 1990; Wood 2006). Der Parameter  $\phi$  wird als Skalenparameter bezeichnet. Wenn  $\phi$  bekannt ist, so handelt es sich bei  $\theta$  um den kanonischen Parameter. Wenn  $\phi$  unbekannt ist, dann handelt es sich bei Dichtefunktionen der Form 2.1 nicht automatisch um Dichtefunktionen einer Exponentialfamilie (McCullagh und Nelder 1989). Zusätzlich wird gefordert, dass die Komponenten  $Y_1, \dots, Y_N$  von  $Y$  stochastisch unabhängig sind. Somit lässt sich die Dichtefunktion von  $Y_i, i = 1, \dots, N$  darstellen als

$$f_{\theta,i}(y_i) = \exp [(y_i \cdot \theta - b_i(\theta))/a_i(\phi) + c_i(y_i, \phi)] . \quad (2.2)$$

Meistens wird  $a_i$  als

$$a_i(\phi) = \frac{\phi}{\check{\omega}_i}, \quad (2.3)$$

ausgedrückt, wobei in der Regel  $\check{\omega}_i = 1$  gewählt wird. Die Varianzfunktion  $V$  wird folgendermaßen definiert:

$$V(\mu_i) := \frac{b_i''(\theta)}{\check{\omega}_i}. \quad (2.4)$$

Dabei ist  $b_i''(\theta)$  die zweite Ableitung nach  $\theta$  der in Formel 2.2 verwendeten Funktion  $b_i$ . Der Name Varianzfunktion stammt daher, dass das Produkt aus ihr mit dem Skalenparameter  $\phi$  gleich der Varianz von  $Y_i$  ist:

$$\text{Var}(Y_i) = V(\mu_i) \cdot \phi . \quad (2.5)$$

Weiter wird eine sogenannte Linkfunktion  $\mathcal{G}$  festgelegt, mit der der Erwartungswert von  $Y$  transformiert wird. Der transformierte Erwartungswert wird in linearer Abhängigkeit der Regressormatrix modelliert:

$$\mathcal{G}(\mu_i) = X_i\beta, \quad (2.6)$$

wobei  $X_i$  die  $i$ -te Zeile der Regressormatrix  $X$  bezeichnet\* und  $\beta = (\beta_1, \dots, \beta_P)^\top$  ein unbekannter Parametervektor der Länge  $P$  ist. Der Ausdruck  $X_i\beta$  wird auch als lineare Komponente bezeichnet. Liegt eine binäre Responsevariable vor, wie z. B. „Krankheit liegt vor“/„Krankheit liegt nicht vor“, so wird häufig eine Binomialverteilung unterstellt und als Linkfunktion die logit-Funktion gewählt:

$$\text{logit}(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = X_i\beta, \quad E(Y_i) = \mu_i. \quad (2.7)$$

Die logit-Funktion sorgt dafür, dass nur Erwartungswerte im Bereich  $[0, 1]$  geschätzt werden können.

Ein anderes Anwendungsbeispiel für die Verwendung eines GLM ist die Modellierung von Kosten: Durch eine log-Linkfunktion wird gewährleistet, dass nur positive Werte als Erwartungswert in Frage kommen. Durch die Annahme einer gammaverteilten Responsevariable kann eine mit dem Erwartungswert ansteigende Varianz modelliert werden. Außerdem nimmt eine gammaverteilte Zufallsvariable nur positive Werte an, eine gängige Eigenschaft von Kosten. Für Zähldaten, d. h. Variablen, die nur ganze Zahlen, inklusive 0 annehmen können, ist oft eine Poissonverteilung geeignet. Die Maximum-Likelihood-Schätzer der Parameter eines GLM werden mittels des Newton-Raphson-Algorithmus bestimmt<sup>†</sup> (vgl. Wood 2006).

---

\*Die für Vektoren unübliche Notation  $X_i$  wurde gewählt, um Verwechslungen mit den Kovariablen  $x_p$  zu vermeiden.

<sup>†</sup>Es wird das R-Paket `mgcv` verwendet, in dem dieser Algorithmus eingebettet ist.

### 2.1.2 Das verallgemeinerte additive Modell (GAM)

Generalisierte lineare Modelle können sich als zu ungenau erweisen. Sollen zum Beispiel die Ausgaben der Krankenkassen pro versicherter Person modelliert werden, so könnte das Modell

$$\log(E(\text{Ausgaben})) = \beta_0 + \beta_1 \cdot \text{weiblich} + \beta_2 \cdot \text{Alter} \quad (2.8)$$

modelliert werden, wobei die Ausgaben als gammaverteilt angenommen werden. Der natürliche Logarithmus  $\log$  wäre in diesem Fall die Linkfunktion. Die Annahme, dass der Prädiktor  $\mathcal{G}(\mu) = \log(\text{erwartete Ausgaben})$  linear im Alter ist, ist jedoch eine starke Einschränkung. Ein Ausweg wäre es, quadratische und kubische Terme des Alters mit in die Modellgleichung aufzunehmen. Polynome höheren Grades werden in der Regel nicht verwendet, da diese häufig zu sehr instabilen Schätzungen führen und die Anpassung an die Daten nicht wesentlich verbessert wird (Wood 2006).

Stattdessen können GLMs weiter verallgemeinert werden, indem auch nichtlineare „glatte“ Terme in der Regressionsgleichung zugelassen werden. Im vorliegenden Beispiel würde Formel 2.8 folgendermaßen modifiziert werden:

$$\log(E(\text{Ausgaben})) = \beta_0 + \beta_1 \cdot \text{weiblich} + \mathcal{F}(\text{Alter}), \quad (2.9)$$

wobei  $\mathcal{F}$  eine „glatte“ Funktion des Alters ist. In der Praxis werden solche glatten Funktionen häufig als Splines modelliert. Diese werden in den Abschnitten 2.1.5 und 2.1.6 eingeführt. Modelle dieser Art werden als verallgemeinerte (generalisierte) additive Modelle (GAMs) bezeichnet.

Ein verallgemeinertes additives Modell (Hastie und Tibshirani 1990) ist ein verallgemeinertes lineares Modell, bei dem im linearen Prädiktor eine Summe glatter Funktionen von Kovariablen enthalten ist. Ein GAM lässt sich z. B. folgendermaßen

darstellen:

$$\mathcal{G}(\mu_i) = X_i^* \vartheta + \mathcal{F}_1(x_{1i}) + \mathcal{F}_2(x_{2i}, x_{3i}) + \mathcal{F}_3(x_{4i}) + \dots \quad (2.10)$$

Dabei ist  $\mu_i \equiv E(Y_i)$  und die Verteilung von  $Y_i$  gehört zu einer Exponentialfamilie (Wood 2006). Weiter ist  $\mathcal{G}$  eine bekannte, monotone und zweifach differenzierbare Link-Funktion;  $X_i^*$  ist die  $i$ -te Zeile der Design-Matrix der strikt parametrischen Komponenten mit zugehörigem Parametervektor  $\vartheta$ . Die  $\mathcal{F}_j, j = 1, 2, \dots$  sind glatte Funktionen der Kovariablen  $x_j$ . Jede glatte Funktion kann sich auf eine oder auf mehrere Kovariablen beziehen. Wie GAMs geschätzt werden, wird in Abschnitt 2.1.7 erläutert.

### 2.1.3 Überstreuung („over-dispersion“)

In Modellen, in denen die Responsevariable als binomial- oder Poisson-verteilt angesehen wird, tritt häufig das Problem auf, dass die beobachteten Werte eine stärkere Streuung aufweisen, als theoretisch zu erwarten wäre. Dieses Problem wird als Überstreuung („over-dispersion“) bezeichnet. Bei Poisson-verteilten Variablen sind Erwartungswert und Varianz identisch. Bei binomialverteilten Zufallsvariablen mit bekannter Anzahl von Bernoulliexperimenten (das ist z. B. die Anzahl der Münzwürfe) ist die Varianz ebenfalls durch den Erwartungswert (in diesem Falle die Wahrscheinlichkeit eines „positiven“ Münzwurfes) festgelegt. In der Darstellung der Dichtefunktion der Form gemäß Formel 2.1 äußert sich dieser Zusammenhang darin, dass  $a(\phi) = 1$  ist. Kurz gesagt, der Skalenparameter  $\phi$ , der auch als Dispersionsparameter bezeichnet wird, ist gleich 1.

In der Praxis ist Überstreuung eher die Regel als die Ausnahme. Mögliche Ursachen von Überstreuung wurden bereits von Lexis (1879) diskutiert<sup>‡</sup> (siehe auch McCullagh und Nelder (1989) und Dobson (2002)). Eine Möglichkeit, mit Überstreuung

---

<sup>‡</sup>Wilhelm Lexis (\* 17. Juli 1837 in Eschweiler (Rheinland); † 24. August 1914 in Göttingen) war ein deutscher Volkswirt, Nationalökonom und Statistiker.

umzugehen ist es, den Skalenparameter  $\phi$  nicht als durch die Verteilungsfunktion implizit vorgegeben anzusehen, sondern ihn, wie bei Modellen mit Normalverteilung oder Gammaverteilung, mittels der Maximum-Likelihood-Methode zu schätzen.

### 2.1.4 Quasi-Likelihood

In generalisierten linearen Modellen wird unterstellt, dass die Verteilungsfunktion der Responsevariable zu einer *bekannt* Exponentialfamilie gehört. In einigen Fällen ist es jedoch nicht möglich, eine geeignete Verteilung festzulegen. Manchmal ist es besser, lediglich eine Annahme über den Zusammenhang zwischen Erwartungswert und Varianz der Responsevariable zu treffen. Dann ist es möglich, anstelle der Likelihood die sogenannte Quasi-Likelihood zu verwenden, um die Parameter zu schätzen. Die Quasi-Likelihood geht auf Wedderburn (1974) zurück. Die hier beschriebene Darstellung orientiert sich an Wood (2006).

Es sei  $y_i$  die Realisation der Zufallsvariable  $Y_i$  mit Erwartungswert  $\mu_i$ . Zusätzlich sei die Varianzfunktion  $V(\mu_i)$  gemäß Formel 2.4 bekannt. Dann wird die Log-Quasi-Likelihood für  $\mu_i$  bei gegebenem  $y_i$  definiert als:

$$q_i(\mu_i) := \int_{y_i}^{\mu_i} \frac{y_i - z}{\phi V(z)} dz. \quad (2.11)$$

Gegeben, dass es sich bei den Daten um unabhängige Zufallsvariablen handelt, kann die Log-Quasi-Likelihood für den Erwartungsvektor  $\mu$  definiert werden als

$$q(\mu) = \sum_{i=1}^N q_i(\mu_i). \quad (2.12)$$

Die Quasi-Log-Likelihood  $q$  hat viele nützliche Eigenschaften, die auch die Log-Likelihood  $l$  einer Verteilungsfunktion einer bekannten Exponentialfamilie besitzt. Sie hat aber den entscheidenden Vorteil, dass nur eine bekannte Varianzfunktion  $V$  benötigt wird. Der Maximum-Quasi-Likelihood-Schätzer für  $\beta$  erfüllt dieselben asymptotischen Eigenschaften wie ein entsprechender Maximum-Likelihood-Schätzer.

Bei Hypothesentests muss die Log-Likelihood durch die Quasi-Log-Likelihood ersetzt werden (vgl. Wood 2006).

### 2.1.5 Splines

Der Begriff „Splines“ stammt aus dem Schiffbau: Bei der Herstellung des Rumpfes werden lange dünne Latten an einzelnen Punkten durch Nägel fixiert. Die Latten biegen sich, so dass der Rumpf eine glatte Oberfläche bildet. Nach einem ähnlichen Prinzip wird versucht, die glatten Funktionen, die in GAMs verwendet werden, zu schätzen. In der numerischen Mathematik bezeichnet ein Spline ein stückweises Polynom  $n$ -ten Grades, das an Stellen, wo zwei Polynome aufeinander treffen  $(n - 1)$ -mal stetig differenzierbar ist. Diese Stellen werden als Knoten bezeichnet. Je nach Grad des Polynoms werden Splines als linear, quadratisch oder kubisch bezeichnet. Kubische Splines werden in der Praxis am häufigsten eingesetzt. Mittlerweile gibt es auch knotenfreie Spline-Basen (siehe folgender Abschnitt).

### 2.1.6 „Thin plate regression splines“

Bei vielen Spline-Varianten (Spline-Basen) muss die Lage der Knoten vorab gewählt werden. Die Wahl der Knoten unterliegt einer gewissen Subjektivität. Mit „thin plate splines“ (Duchon 1977) wird dieses Problem gelöst, indem eine knotenfreie Basis geschaffen wird. Zusätzlich ist es mit „thin plate splines“ im Gegensatz zu anderen Spline-Basen möglich, mehr als eine Kovariable gleichzeitig zu glätten. Ein Problem ist jedoch, dass die Verwendung von „thin plate splines“ einen sehr hohen Rechenaufwand verursacht. Deshalb wurden sie zu „thin plate regression splines“ weiterentwickelt, die numerisch wesentlich leichter zu handhaben sind (Wood 2003).

Gegeben sei das Problem, eine glatte Funktion  $\mathcal{F} : \mathbb{R}^P \rightarrow \mathbb{R}$  zu schätzen, wobei  $P$  die Anzahl der gemeinsam zu glättenden Kovariablen bezeichnet. Es sei  $y = (y_1, \dots, y_N)^\top$  ein Datenvektor und es seien  $x_1, \dots, x_P$  die zu glättenden Kovariablen

mit  $x_p = (x_{1p}, \dots, x_{Np})^\top, p = 1, \dots, P$ . Dabei soll gelten:

$$y_i = \mathcal{F}(x_{i1}, \dots, x_{iP}) + \varepsilon_i, \quad i = 1, \dots, N, \quad (2.13)$$

wobei  $\varepsilon_i$  einen Fehlerterm darstellt. Bei einem „thin plate spline“ handelt es sich um den Schätzer  $\hat{\mathcal{F}}$  für  $\mathcal{F}$ , der den folgenden Term bezüglich  $\tilde{\mathcal{F}}$  minimiert:

$$\| y - (\tilde{\mathcal{F}}(x_{11}, \dots, x_{1P}), \dots, \tilde{\mathcal{F}}(x_{N1}, \dots, x_{NP}))^\top \|^2 + \mathcal{L} \cdot J_{mP}(\tilde{\mathcal{F}}). \quad (2.14)$$

Dabei bezeichnet  $\| \cdot \|$  die euklidische Norm und  $J_{mP}(\tilde{\mathcal{F}})$  stellt einen Strafterm dar, der die Schwankung der Funktion  $\tilde{\mathcal{F}}$  misst. Der Parameter  $\mathcal{L}$  legt fest, wie stark geglättet werden soll.

Um den Strafterm  $J_{mP}(\tilde{\mathcal{F}})$  zu definieren, wird eine natürliche Zahl  $m \in \mathbb{N}$  frei gewählt<sup>§</sup>. Diese gibt an, wie oft die Funktion  $\tilde{\mathcal{F}}(z_1, \dots, z_P)$  nach einzelnen seiner Argumente  $z_1, \dots, z_P$  gleichzeitig abgeleitet wird. Die Definition des Strafterms lautet:

$$J_{mP}(\tilde{\mathcal{F}}) := \int \cdots \int_{\mathbb{R}^P} \sum_{\nu_1 + \dots + \nu_P = m} \frac{m!}{\nu_1! \cdots \nu_P!} \cdot \left( \frac{\partial^m \tilde{\mathcal{F}}}{\partial z_1^{\nu_1} \cdots \partial z_P^{\nu_P}} \right)^2 dz_1 \cdots dz_P. \quad (2.15)$$

Zusätzlich wird die Bedingung  $2m > P$  festgelegt, da dann die Funktion  $\hat{\mathcal{F}}$ , die Gleichung 2.14 minimiert, in die Form

$$\begin{aligned} \hat{\mathcal{F}}(z_1, \dots, z_P) &= \sum_{i=1}^N \delta_i \cdot \Psi_{mP}(\| (z_1 - x_{i1}, \dots, z_P - x_{iP})^\top \|) \\ &\quad + \sum_{j=1}^M \rho_j \cdot \varphi_j(z_1, \dots, z_P) \end{aligned} \quad (2.16)$$

gebracht werden kann, wobei  $\delta = (\delta_1, \dots, \delta_N)^\top$  und  $\rho = (\rho_1, \dots, \rho_M)^\top$  unbekannte Parametervektoren sind. Diese müssen die Nebenbedingung  $G^\top \delta = 0$  erfüllen, wobei

---

<sup>§</sup>In dieser Arbeit wird der Default-Wert des R-Pakets `mgcv` verwendet. Bei diesem handelt es sich um den kleinsten Wert, der die Gleichung  $2m > P + 1$  erfüllt.

$G_{ij} = \varphi_j(x_i)$  ist. Die Funktionen  $\varphi_i$ ,  $i = 1, \dots, M$ ,  $M = \binom{m+P-1}{P}$ , sind linear unabhängige Polynome, durch die der Raum der Polynome zum Grad kleiner  $m$  in  $\mathbb{R}^P$  aufgespannt wird. Gilt z. B.  $m = P = 2$ , so nehmen diese Funktionen die Werte  $\varphi_1(z_1, z_2) = 1$ ,  $\varphi_2(z_1, z_2) = z_1$  und  $\varphi_3(z_1, z_2) = z_2$  an, wobei die Indizes auch vertauscht werden können (Wood 2006). Die Funktion  $\Psi_{mP}$  aus Gleichung 2.16 wird folgendermaßen definiert:

$$\Psi_{mP}(r) = \begin{cases} \frac{(-1)^{m+1+P/2}}{2^{2m-1}\pi^{P/2}(m-1)!(m-P/2)!} r^{2m-P} \log(r) & \text{wenn } P \text{ gerade ist,} \\ \frac{\Gamma(P/2-m)}{2^{2m}\pi^{P/2}(m-1)!} r^{2m-P} & \text{wenn } P \text{ ungerade ist.} \end{cases} \quad (2.17)$$

Weiter wird die Matrix  $\Lambda$  mit  $\Lambda_{ij} \equiv \Psi_{mP}(\| (x_{i1} - x_{j1}, \dots, x_{iP} - x_{jP})^\top \|)$  definiert. Das Spline-Anpassungs-Problem lässt sich nun darstellen als

$$\min_{\delta, \rho} \| y - \Lambda\delta - G\rho \|^2 + \mathcal{L}\delta^\top \Lambda\delta \text{ mit } G^\top \delta = 0. \quad (2.18)$$

Eine detailliertere Darstellung der „thin plate splines“ geben Wahba (1990), Green und Silverman (1994) sowie Wood (2006).

Wie bereits erwähnt, erweist sich die Anpassung von „thin plate splines“ als sehr rechenintensiv. Es gibt genauso viele Parameter wie Zeilen in dem anzupassenden Datensatz. Der numerische Rechenaufwand ist proportional zu der dritten Potenz der Anzahl Parameter. Deshalb werden die „thin plate splines“ modifiziert, so dass sich „thin plate regression splines“ ergeben.

Für den Term  $\sum_{j=1}^M \rho_j \cdot \varphi_j(z_1, \dots, z_P)$  aus Formel 2.16 lässt sich zeigen, dass dessen Ableitungen im Strafterm (Formel 2.15) immer gleich 0 sind. Sie beeinflussen den Strafterm nicht. „Thin plate regression splines“ basieren auf der Idee, den Raum der Schwankungskomponente zu kürzen. Das bedeutet, dass in Gleichung 2.16 einige Summanden wegfallen, die ein  $\delta_i$  enthalten (Wood 2006). Gleichzeitig sollen die Komponenten ohne Schwankungen, also die Komponenten mit einem  $\rho_j$ , un-

verändert bleiben. Es sei  $\Lambda = U\Xi U^T$  die Spektralzerlegung von  $\Lambda$ , wobei  $\Xi$  eine Diagonalmatrix der Eigenwerte von  $\Lambda$  ist, für die zusätzlich  $|\Xi_{i,i}| \geq |\Xi_{i-1,i-1}|$ ,  $i = 2, \dots, N$ , gilt. Die Spalten der Matrix  $U$  bestehen aus den zugehörigen Eigenvektoren der Matrix  $\Lambda$ . Nun bezeichne  $U_k$  die Matrix, die aus den ersten  $k$  Spalten von  $U$  besteht und  $\Xi_k$  sei die obere rechte  $k \times k$  Teilmatrix von  $\Xi$ . Der Parametervektor  $\delta$  wird auf den Spaltenraum der Teilmatrix  $U_k$  beschränkt, so dass  $\delta = U_k \delta_k$  gilt. Dadurch wird 2.18 abgewandelt zu

$$\min_{\delta_k, \rho} \|y - U_k \Xi_k \delta_k - G\rho\|^2 + \mathcal{L} \delta_k^T \Xi_k \delta_k \quad (2.19)$$

mit der Nebenbedingung

$$G^T U_k \delta_k = 0. \quad (2.20)$$

Die Nebenbedingung soll nun in die Minimierungsformel integriert werden. Dazu wird erstens eine orthogonale Spaltenbasis, die Matrix  $Z_k$ , ermittelt, so dass  $G^T U_k Z_k = 0$  ist. Eine Möglichkeit, eine geeignete Matrix  $Z_k$  zu finden besteht in der Bestimmung der QR-Zerlegung von  $U_k^T G$ : Die finalen Spalten des orthogonalen Faktors bilden eine mögliche Matrix  $Z_k$ . Indem  $\delta_k$  durch die Gleichung  $\delta_k = Z_k \tilde{\delta}$  eingeschränkt wird, entsteht ein Minimierungsproblem ohne Nebenbedingung, bei dem der Rang  $k > M$  (frei) gewählt werden muss<sup>¶</sup>:

$$\min_{\tilde{\delta}, \rho} \|y - U_k \Xi_k Z_k \tilde{\delta} - G\rho\|^2 + \mathcal{L} \tilde{\delta}^T Z_k^T \Xi_k Z_k \tilde{\delta}. \quad (2.21)$$

Der Rechenaufwand von Prozeduren wird häufig durch die Schreibweise  $O(\diamond)$  angegeben. Das bedeutet, dass der Rechenaufwand proportional zu  $\diamond$  ansteigt, wobei  $\diamond$  eine natürliche Zahl ist. Der Rechenaufwand des Minimierungsproblems in Formel 2.21 beträgt  $O(k^3)$ . Nachdem das GAM angepasst ist und somit alle Parameter geschätzt sind, kann jede Stelle des Splines berechnet werden. Dazu muss lediglich  $\delta = U_k Z_k \tilde{\delta}$  berechnet und anschließend Formel 2.16 angewendet

---

<sup>¶</sup>Sofern nicht anders angegeben, wird in dieser Arbeit der Default-Wert  $10 \cdot 3^{P-1}$  ( $= 10$ , wenn nur eine Variable geglättet werden soll) des R-Paketes `mgcv` verwendet.

werden.

Das größte verbleibende Problem ist es,  $U_k$  und  $\Xi_k$  ohne viel Rechenaufwand zu bestimmen. Eine volle Eigenwertzerlegung von  $\Lambda$  benötigt  $O(N^3)$  Operationen, was weiterhin sehr rechenaufwendig wäre. Stattdessen wird die Iterations-Methode von Lanczos verwendet, die mit  $O(N^2k)$  wesentlich weniger rechenaufwendig ist und in Wood (2003) dargestellt wird.

### 2.1.7 Darstellung von GAMs als GLMs

Um ein GAM zu schätzen, muss für jede glatte Komponente  $\mathcal{F}_j$  eine Basis, genauer eine Menge von Basisfunktionen  $\mathcal{B}_{j\ell}$ , bestimmt werden. Als Basis könnten z. B. Polynome verwendet werden. Polynome sind jedoch verhältnismäßig ungeeignet und es empfiehlt sich, stattdessen Splines zu verwenden. Außerdem muss definiert werden, wie die Schwankung der Funktion  $\mathcal{F}_j$  gemessen wird. Glatte Komponenten müssen darstellbar sein als:

$$\mathcal{F}_j(z) = \sum_{\ell=1}^{q_j} \beta_{j\ell} \mathcal{B}_{j\ell}(z), \quad (2.22)$$

wobei  $q_j$  die Anzahl der in  $\mathcal{F}_j$  einfließenden Basisfunktionen bezeichnet. In der Regel entspricht  $z$  einer einzelnen Kovariablen, aber es ist möglich, dass  $z$  einen Vektor von Kovariablen darstellt. Die Parameter der glatten Funktion werden durch  $\beta_{j\ell}$  dargestellt und müssen innerhalb des GAM mit Hilfe der empirischen Daten  $y = (y_1, \dots, y_N)^\top$  sowie  $x_1, \dots, x_P$  mit  $x_p = (x_{1p}, \dots, x_{Np})^\top$ ,  $p = 1, \dots, P$ , geschätzt werden.

Wird die (unbekannte) Funktion  $\mathcal{F}_j$  auf die zugehörige(n) Kovariablen(n) der  $N$  Realisationen angewendet, so ergibt sich ein (unbekannter) Vektor  $\mathcal{F}_j^x = (\mathcal{F}_j^x[1], \dots, \mathcal{F}_j^x[N])^\top$ . Dieser lässt sich mit Hilfe der gegebenen Basis als Produkt einer Modellmatrix  $\tilde{X}_j \sim N \times q_j$  und dem (unbekannten) Parametervektor  $\tilde{\beta}_j :=$

$(\beta_{j1}, \beta_{j2}, \dots, \beta_{jq_j})^\top$  darstellen:

$$\mathcal{F}_j^x = \tilde{X}_j \tilde{\beta}_j. \quad (2.23)$$

Dabei entspricht das Element in der  $i$ -ten Zeile und  $\ell$ -ten Spalte von  $\tilde{X}_j$  dem Ergebnis der  $\ell$ -ten (bekannten) Basisfunktion angewendet auf die zugehörige(n) Kovariable(n) der  $i$ -ten Beobachtung.

Die Matrizen der Form  $\tilde{X}_j$  müssen noch weiter verändert werden bevor sie mit der parametrischen Matrix  $X_i^*$  und den Modellmatrizen der anderen glatten Terme zur Regressormatrix des gesamten Modells vereint werden können.

Das Gesamtmodell ist noch nicht identifizierbar: Es könnte zu jedem glatten Term eine Konstante hinzugefügt werden, die durch die Konstanten der anderen Terme wieder ausgeglichen werden. Daher wird eine Nebenbedingung gewählt und zwar soll die Summe der Elemente von  $\mathcal{F}_j^x$  insgesamt 0 ergeben:

$$1^\top \tilde{X}_j \tilde{\beta}_j = 0. \quad (2.24)$$

Durch eine Reparametrisierung wird diese Nebenbedingung in das Modell integriert. Dazu wird eine Matrix  $Z$  gesucht, die aus  $q_j - 1$  orthogonalen Spalten besteht und die Gleichung

$$1^\top \tilde{X}_j Z = 0 \quad (2.25)$$

erfüllt. Es werden  $q_j - 1$  neue Parameter  $\beta_j$  geschaffen, welche die Bedingung  $\hat{\beta}_j = Z\beta_j$  erfüllen müssen. Die neue Modell-Matrix des glatten Terms wird durch  $X_j = \tilde{X}_j Z$  bestimmt und die Modellgleichung 2.10 des GAM kann somit geschrieben werden als:

$$\mathcal{G}(\mu) = X\beta \quad (2.26)$$

wobei  $X = [X^* : X_1 : X_2 : \dots]$  und  $\beta^\top = [\vartheta^\top, \beta_1^\top, \beta_2^\top, \dots]$ . Gleichung 2.26 stellt ein GLM dar, welches jedoch nicht mit der herkömmlichen Maximum-Likelihood-Methode geschätzt wird. Diese würde zu einem zu stark angepassten Modell führen,

in dem die glatten Terme zu stark schwanken. Deshalb wird von der Likelihood ein Term abgezogen, der zu starkes Schwanken bestraft. Es bezeichne  $\mathcal{D}_j$  den Strafterm des  $j$ -ten glatten Terms, dessen Einfluss durch einen Glättungsparameter  $\mathcal{L}_j$  reguliert wird. Die bestrafte (penalized) Likelihood  $l_p$  ist folgendermaßen definiert:

$$l_p(\beta) = l(\beta) - \frac{1}{2} \sum_j \mathcal{L}_j \mathcal{D}_j. \quad (2.27)$$

Prinzipiell ist es möglich, die Glättungsparameter  $\mathcal{L}_j$  vorzugeben. Um gute Glättungsparameter automatisch zu bestimmen, kann z. B. das Verfahren der verallgemeinerten Kreuzvalidierung verwendet werden.

### 2.1.8 Die Wahl des Glättungsparameters bei GAMs - Kreuzvalidierung

Um den Glättungsparameter  $\mathcal{L}_j$  nicht subjektiv wählen zu müssen, kann das Verfahren der Kreuzvalidierung („cross validation“) verwendet werden. Es gibt unterschiedliche Formen der Kreuzvalidierung. Alle haben gemein, dass ein Kreuzvalidierungsscore bezüglich der Wahl der Glättungsparameter minimiert wird. Liegt ein minimaler Kreuzvalidierungsscore vor, so ist ein gewisses Optimalitätskriterium erfüllt. Bei der gewöhnlichen Kreuzvalidierung (ordinary cross validation, OCV) wird der entsprechende Score  $\mathcal{V}_o$  eines Modells bei gegebenen Glättungsparametern folgendermaßen ermittelt: Der Reihe nach wird das Modell erneut angepasst, wobei die  $i$ -te Beobachtung  $y_i$  ausgelassen wird. Mit Hilfe dieses gekürzten Modells wird die  $i$ -te Beobachtung vorhergesagt, der Vorhersagewert wird mit  $\hat{\mu}_i^{[-i]}$  bezeichnet.  $\mathcal{V}_o$  wird definiert als

$$\mathcal{V}_o := \frac{1}{N} \sum_{i=1}^N (y_i - \hat{\mu}_i^{[-i]})^2. \quad (2.28)$$

Es ist möglich,  $\mathcal{V}_o$  zu bestimmen, ohne dass  $N$  vollständige Modelle angepasst werden müssen (vgl. Wood (2006)). Trotzdem kann die Verwendung von OCV zu einem sehr hohen Rechenaufwand führen, wenn mehrere Glättungsparameter bestimmt werden

müssen. Weiter hat OCV den Nachteil, dass die Scores  $\mathcal{V}_o$  nicht invariant bezüglich gewisser Transformationen, wie z. B. Drehungen des Koordinatensystems sind (Golub *et al.* 1979, Wahba 1990). Diese Probleme werden mit der verallgemeinerten Kreuzvalidierung (generalized cross validation, GCV) behoben, bei der eine andere Gewichtung als bei OCV verwendet wird. Eine explizite Erklärung der Gewichtung befindet sich in Wood (2006). Zwei wichtige Größen, die in die Berechnung des GCV-Scores  $\mathcal{V}_g$  einfließen, sind die Devianz  $D$  und die Einflussmatrix (influence matrix)  $\mathcal{E}$ . Die Devianz ist folgendermaßen definiert: Es sei  $l(\hat{\beta})$  die Likelihood des durch das Maximum-Likelihood-Verfahren geschätzten Parameters  $\hat{\beta}$ . Weiter sei  $l(\hat{\beta}_{\max})$  die maximierte Likelihood des gesättigten Modells, d. h. eines Modells mit einem Parameter für jede Beobachtung, so dass  $\hat{\mu} = y$  ist. Die Devianz ist definiert als

$$D = 2[l(\hat{\beta}_{\max}) - l(\hat{\beta})]\phi . \quad (2.29)$$

Der GCV-Score lässt sich darstellen als

$$\mathcal{V}_g := \frac{ND(\hat{\beta})}{(N - \text{tr}(\mathcal{E}))^2} . \quad (2.30)$$

Für die Einflussmatrix  $\mathcal{E}$  gilt:

$$\mathcal{E} = X(X^\top W X)^{-1} X^\top W . \quad (2.31)$$

Dabei ist  $W = (W_{ii})$  eine Diagonalmatrix, die iterativ im Zuge der (bestraften) Maximum-Likelihood-Schätzung der Parameter bestimmt wird. Es gilt:

$$W_{ii} = \frac{1}{V(\hat{\mu}_i)\mathcal{G}'(\hat{\mu}_i)^2} . \quad (2.32)$$

Bei GLMs ist die Spur der Einflussmatrix gleich der Anzahl der zu identifizierenden Parameter  $o$ :

$$\text{tr}(\mathcal{E}) = o . \quad (2.33)$$

Bei GAMs übersteigt die Zahl der tatsächlich geschätzten Parameter die Spur der Einflussmatrix. Da nicht die Likelihood, sondern die bestrafte Likelihood maximiert

wird, sind GAMs wesentlich schwächer an die Daten angepasst, als mit der verwendeten Parameterzahl denkbar wäre. In approximativen Signifikanztests für die glatten Terme werden daher nicht die tatsächlichen Freiheitsgrade verwendet, sondern effektive Freiheitsgrade geschätzt. Diese entsprechen einer Verallgemeinerung von Formel 2.33 auf GAMs. In der Regel werden nicht die Freiheitsgrade des gesamten Modells, sondern nur die einzelner glatter Terme benötigt. Wie die Freiheitsgrade auf die einzelnen Terme aufgeteilt werden, kann in Wood (2006) nachgelesen werden.

### 2.1.9 Überprüfung der Modellvoraussetzungen

Es ist wichtig, ein angepasstes Modell daraufhin zu überprüfen, ob die Modellvoraussetzungen erfüllt sind. Gelten die Modellvoraussetzungen nicht, so sollte (falls möglich) ein adäquateres Modell gewählt werden. Oft sind bestimmte Arten von Residuen bei Gültigkeit der Modellvoraussetzungen approximativ normalverteilt. Ob die Residuen einer Normalverteilung folgen, lässt sich graphisch überprüfen.

In linearen Modellen der Form  $y \sim \mathcal{N}(\mu, \sigma^2 I)$  können die rohen Residuen  $\hat{\varepsilon}_i := y_i - \hat{\mu}_i$  betrachtet werden. In verallgemeinerten linearen Modellen hängt die Varianz der Zufallsvariable  $Y_i$  häufig von ihrem Erwartungswert  $\mu_i$  ab. Der Zusammenhang lässt sich durch eine Funktion  $V$  charakterisieren:

$$\text{Var}(Y_i) = V(\mu_i) \cdot \phi. \quad (2.34)$$

Die Varianz der rohen Residuen  $\hat{\varepsilon}_i$  hängt dadurch ebenfalls von  $\mu_i$  ab. Zur Überprüfung der Modellvoraussetzungen wäre es daher besser, anstelle der rohen Residuen die intuitiv leicht verständlichen Pearson Residuen  $\hat{\varepsilon}_i^p$  zu betrachten. Diese sind folgendermaßen definiert:

$$\hat{\varepsilon}_i^p := \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}. \quad (2.35)$$

Die Pearson Residuen haben einen approximativen Erwartungswert von 0 und Varianz  $\phi$ , falls die Modellvoraussetzungen stimmen. In der Praxis können die Pearson

Residuen jedoch etwas unsymmetrisch um die 0 verteilt sein, so dass ihr Verhalten der Normalverteilung nicht stark genug entspricht. Es ist hilfreich, stattdessen die Devianz-Residuen zu verwenden. In den Termen der Dichtefunktion einer Exponentialfamilie lässt sich die Devianz folgendermaßen schreiben:

$$D = \sum_{i=1}^N 2\check{\omega}_i \left[ y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right]. \quad (2.36)$$

Bei  $\tilde{\theta}_i$  und  $\hat{\theta}_i$  handelt es sich um die jeweiligen Maximum-Likelihood-Schätzer des kanonischen Parameters  $\theta_i$  für das gesättigte bzw. das zu überprüfende Modell. Es sei

$$d_i := 2\check{\omega}_i \left[ y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right]. \quad (2.37)$$

Dann sind die Devianz-Residuen definiert als:

$$\hat{\varepsilon}_i^d := \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i}. \quad (2.38)$$

Die Devianz-Residuen  $\hat{\varepsilon}_i^d$  sind approximativ standardnormalverteilt, d. h.  $\hat{\varepsilon}_i^d \stackrel{\text{appr.}}{\sim} \mathcal{N}(0, 1)$ .

Die Normalverteilungsannahme lässt sich mit Hilfe eines Quantil-Quantil-Diagramms oder eines Histogramms graphisch überprüfen. Ob die Varianzfunktion  $V(\mu)$  angemessen ist, lässt sich prüfen, indem die Devianz-Residuen gegen die geschätzten Erwartungswerte abgetragen werden. Es ist auch empfehlenswert, die Responsevariable gegen die durch das Modell vorhergesagten Werte zu plotten, um eventuelle Unregelmäßigkeiten festzustellen (Wood 2006).

## 2.2 Delaunay Triangulation

Um Punkte im dreidimensionalen Raum zu interpolieren, wird in dieser Arbeit die sogenannte „Delaunay Triangulation“ als Hilfsmittel verwendet. In einer Triangulation werden Punkte im zweidimensionalen Raum als Eckpunkte aneinander angrenzender Dreiecke aufgefasst. Oberhalb dieser Dreiecke können Punkte durch eine

Ebene linear interpoliert werden. Die Delaunay Triangulation ist definiert als eine Menge von Dreiecken mit den folgenden fünf Eigenschaften (Gebhardt 2005):

1. Die Eckpunkte der Dreiecke sind Knoten.
2. Kein Dreieck enthält andere Knoten, als seine eigenen Eckpunkte.
3. Das Innere der Dreiecke ist paarweise disjunkt.
4. Die Vereinigungsmenge der Dreiecke ist die konvexe Hülle der Punktmenge. Das ist die kleinste konvexe Punktmenge, die alle Knoten enthält.
5. Das Innere des Umkreises jedes Dreiecks enthält keinen Knoten.

Die ersten vier Eigenschaften geben die Definition einer Triangulation. Durch die letzte Eigenschaft wird eine Triangulation zu einer „Delaunay Triangulation“, die auf Delaunay (1934) zurück geht. Eine Triangulation, die die Delaunay-Eigenschaft erfüllt, kommt in einem gewissen Sinne so nah wie möglich an gleichwinklige Dreiecke heran. Die Delaunay Triangulation einer Punktmenge ist eindeutig, es sei denn vier oder mehr Knoten liegen genau auf einem Kreis. In der vorliegenden Arbeit wird die Delaunay Triangulation mittels des R-Pakets „tripack“ ermittelt (Gebhardt 2005), das auf dem Algorithmus von Renka (1996) basiert.

## 2.3 Markov-Modellierung

Wenn in einem Entscheidungsproblem Ereignisse zu unterschiedlichen Zeitpunkten stattfinden und sich Wahrscheinlichkeiten im Laufe der Zeit verändern, so ist es empfehlenswert, das Entscheidungsproblem mit Hilfe eines Markovmodells zu lösen (Sonnenberg und Beck 1993). In Markovmodellen wird die Zeit in gleichgroße Abschnitte, sogenannte Zyklen unterteilt. Den Zyklen werden Zeitpunkte  $t_1, t_2, t_3, \dots$  zugeordnet, die in der Regel in der Intervallmitte liegen. Außerdem wird

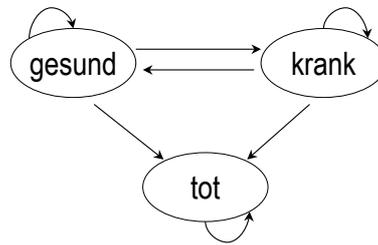


Abbildung 2.1: Erreichbarkeitsgraph eines Markovmodells

eine endliche Anzahl von (z. B. Gesundheits-)Zuständen festgelegt. Die Zustände sind disjunkt und erschöpfend. Dies bedeutet, dass sich die einzelnen Zustände gegenseitig ausschließen und alle möglichen Zustände, in denen sich eine Person befinden kann, durch das Markovmodell abgedeckt werden. Von Zyklus zu Zyklus sind Übergänge möglich. Jedem Übergang wird eine Übergangswahrscheinlichkeit zugeordnet. Des Weiteren gibt eine Startverteilung an, wie sich die betrachtete Population auf die Zustände zum Zeitpunkt  $t_0$  verteilt. Ein Beispiel eines Markovmodells befindet sich in Abbildung 2.1. Dabei handelt es sich um ein klassisches Markovmodell mit drei Zuständen: „gesund“, „krank“ und „tot“. Bei dem Zustand „tot“ handelt es sich um einen absorbierenden Zustand, da von ihm aus keine anderen Übergänge als zu sich selbst möglich sind. Die Pfeile von „gesund“ auf „gesund“ und von „krank“ auf „krank“ deuten an, dass Personen in aufeinanderfolgenden Zyklen im jeweiligen Zustand verbleiben können. Die Pfeile zwischen „gesund“ und „krank“ bedeuten, dass von Zyklus zu Zyklus eine Person erkranken bzw. gesund werden kann. Die Darstellung des Markovmodells in Abbildung 2.1 wird als Erreichbarkeitsgraph bezeichnet.

Jedem Zustand werden sogenannte „Auszahlungen“ zugeordnet. Das sind in gesundheitsökonomischen Analysen Kosten und/oder Nutzen. Pro Zyklus, den sich eine Person in einem Zustand befindet, wird eine entsprechende Auszahlung ausgeschüttet. Auch einzelnen Übergängen können Auszahlungen zugeordnet werden. Soll z. B. berücksichtigt werden, dass im Jahr vor dem Tod in der Regel

Zeitpunkte/Zyklen	$t_0$	$t_1$	$\dots$	$t_{n-1}$	$t_n$
Gewerteter Nutzen:					
- ohne Halbzykluskorrektur	0	$u_1$	$\dots$	$u_{n-1}$	$u_n$
- mit Halbzykluskorrektur	$u_0/2$	$u_1$	$\dots$	$u_{n-1}$	$u_n/2$

Tabelle 2.1: In der Evaluation des Markovmodells gewerteter Nutzen mit und ohne Durchführung der Halbzykluskorrektur;  $u_i$  bezeichnet den durchschnittlichen im Zyklus  $i$  anfallenden Nutzen

höhere Gesundheitsausgaben entstehen, so können allen Übergängen zum Zustand „tot“ entsprechende Kosten zugewiesen werden.

Markovmodelle sind Momentaufnahmen, in der Realität finden dagegen kontinuierlich Übergänge statt. Manchmal ist es notwendig, eine Halbzykluskorrektur durchzuführen. Ohne diese wird implizit angenommen, dass die Übergänge am Anfang eines jeweiligen Zyklus stattfinden. In der Regel ist es realistischer, von Übergängen in der Intervallmitte auszugehen, was Modellen mit Halbzykluskorrektur entspricht. Wird auf eine angemessene Korrektur verzichtet, so können unter Umständen relevante Berechnungsungenauigkeiten entstehen. In Modellen ohne Halbzykluskorrektur finden zum Zeitpunkt der Initialisierung keine Ausschüttungen statt. Für die übrigen Zyklen werden pro Zyklus die vollen Auszahlungen ausgeschüttet. Tabelle 2.1 zeigt die Auszahlungen mit und ohne Halbzykluskorrektur. Bei der Halbzykluskorrektur werden für den Ausgangszeitpunkt sowie für den letzten Zyklus die halben Auszahlungen ausgeschüttet.

Markovmodelle haben die sogenannte Markov-Eigenschaft. Diese besagt, dass alle für die Zukunft relevante Information im jeweiligen Zyklus enthalten ist. Für den weiteren Verlauf ist es irrelevant zu wissen, in welchen Zuständen sich die Personen

in den vergangenen Zyklen befanden. Die Übergangswahrscheinlichkeiten und Auszahlungen des Markovmodells können im Laufe der Zeit variieren. Im Markovmodell dieser Arbeit hängen die Übergangswahrscheinlichkeiten u. a. vom Alter ab. (Vgl. Sonnenberg und Beck (1993))

## 2.4 Gesundheitsökonomische Evaluationsverfahren

Es gibt unterschiedliche gesundheitsökonomische Evaluationsverfahren. Die gängigsten sind die Kosten-Minimierungs-Analyse („cost minimization analysis“, CMA), die Kosten-Effektivitätsanalyse („cost effectiveness analysis“, CEA), die Kosten-Nutzwert-Analyse („cost-utility analysis“, CUA) sowie die Kosten-Nutzen-Analyse im engeren Sinne („cost benefit analysis“, CBA) (Drummond *et al.* 2005). Die Evaluationsverfahren CEA, CUA und CBA werden auch als Kosten-Nutzen-Analysen im weiteren Sinne bezeichnet. Bei einer gesundheitsökonomischen Evaluation ist es erforderlich, eine Perspektive festzulegen. Diese gibt an, aus wessen Sicht die Analyse durchgeführt wird. Am gängigsten sind Analysen aus der Sicht der Gesellschaft. Häufig werden auch Analysen aus der Sicht eines Krankenhauses oder aus der Sicht einer Krankenversicherung durchgeführt. In der vorliegenden Arbeit wird die Perspektive der Sozialversicherungen gewählt, da diese auf der politischen Entscheidungsebene am relevantesten zu sein scheint (Drummond *et al.* 2005; Gold *et al.* 1996).

In der Kosten-Minimierungs-Analyse werden zwei Maßnahmen gleicher Wirksamkeit (nachzuweisen mittels einer Äquivalenzstudie) miteinander verglichen. Für beide Maßnahmen werden die Kosten in Geldeinheiten ermittelt. Die Maßnahme mit den niedrigeren Kosten wird präferiert.

Die Analysetypen Kosten-Effektivitätsanalyse und Kosten-Nutzwert-Analyse sind sich methodisch sehr ähnlich. Ihr einziger Unterschied besteht darin, dass bei der Kosten-Effektivitätsanalyse der Nutzen in einer klinisch messbaren Größe gemessen wird und in der Kosten-Nutzwert-Analyse verschiedene Ergebnisdimensionen zu einem Index zusammengefasst werden. Bei Kosten-Effektivitätsanalysen kann der Nutzen beispielsweise als Anzahl geretteter Leben, geheilter Patienten oder gewonnener Lebensjahre gemessen werden. In Kosten-Nutzwert-Analysen wird der Nutzen meist in Form von gewonnenen qualitätsadjustierten Lebensjahren (QALYs) gemessen. Die Kosten werden in Geldeinheiten gemessen (Drummond *et al.* 2005; Gold *et al.* 1996).

In Kosten-Effektivitätsanalysen sowie in Kosten-Nutzwert-Analysen werden paarweise zwei Szenarien miteinander verglichen. In der Regel wird eine „innovative Therapie“ (Maßnahme 2) mit der Standardbehandlung (Maßnahme 1) verglichen. Der Kostenzuwachs wird durch den Nutzenzuwachs geteilt. Dieser Quotient wird als inkrementelle Kosten-Nutzen-Relation (IKNR) bezeichnet und lautet:

$$\text{IKNR} = \frac{\text{Kosten der Maßnahme 2} - \text{Kosten der Maßnahme 1}}{\text{Nutzen der Maßnahme 2} - \text{Nutzen der Maßnahme 1}} \quad (2.39)$$

Bei einer negativen IKNR ist die Entscheidung trivial: Ist eine Maßnahme billiger und bringt gleichzeitig einen höheren Nutzen mit sich, so wird sie bevorzugt. Wenn hingegen ein höherer Nutzen einer Maßnahme auch mit höheren Kosten einhergeht, muss ermittelt werden, ob die zusätzlichen Kosten einer Maßnahme in einem günstigen Verhältnis zu dem zusätzlichen Nutzen stehen. Gemäß der ökonomischen Theorie lässt sich eine maximale Zahlungsbereitschaft ableiten (zur ökonomischen Begründung siehe Breyer und Zweifel 1997). Dabei wird angenommen, dass der Nutzen bei gegebener Budgetbeschränkung maximiert wird. Wird eine Analyse beispielsweise aus der Sicht der Sozialversicherungen durchgeführt, so gibt die maximale Zahlungsbereitschaft an, wieviele zusätzliche Geldeinheiten die Sozialversicherungen bereit sind, für eine zusätzliche Nutzenein-

heit zu zahlen. Ist die inkrementelle Kosten-Nutzen-Relation kleiner oder gleich der maximalen Zahlungsbereitschaft, so wird das Szenario mit dem höheren Nutzen bevorzugt, andernfalls wird das Szenario mit den geringeren Kosten bevorzugt. Die „wahre“ maximale Zahlungsbereitschaft ist zumeist unbekannt. Sie kann jedoch approximiert werden, indem ermittelt wird, wieviel für die bereits praktizierten Maßnahmen pro gewonnener Nutzeneinheit maximal ausgegeben wird.

Bei Kosten-Nutzen-Analysen im engeren Sinne („cost benefit analysis“) wird keine maximale Zahlungsbereitschaft benötigt. Stattdessen werden sowohl Kosten als auch Nutzen in Geldeinheiten gemessen. Implizit stellt sich dabei die Frage, wie der gewonnene Nutzen monetär bewertet werden soll. Es wird die Alternative mit den geringeren Kosten bevorzugt.

## 2.5 Diskontierung

Diskontierung ist das Gegenteil einer Verzinsung. Sowohl Kosten als auch Nutzen können über viele Jahre verstreut anfallen. Um Kosten und Nutzen unterschiedlicher medizinischer Interventionsmaßnahmen miteinander vergleichen zu können, müssen sie auf den gleichen Zeitpunkt bezogen werden. Andernfalls würde das Verhältnis von Kosten und Nutzen verzerrt dargestellt. Das Hauptprinzip der Diskontierung ist es, dass Kosten und Nutzen, die zu einem früheren Zeitpunkt anfallen, stärker gewertet werden, als Kosten und Nutzen, die zu einem späteren Zeitpunkt auftreten. Kosten und Nutzen späterer Jahre werden abgezinst. Es ist gängige Praxis, sowohl Kosten als auch Nutzen mit derselben Rate zu diskontieren. In den meisten Richtlinien wird diese Vorgehensweise empfohlen. Laut Drummond und Jefferson (1996) wird üblicherweise eine Diskontrate von 5 % pro Jahr gewählt.

Es bezeichne  $\mathcal{R}$  die Diskontrate und  $\mathcal{C}_0, \dots, \mathcal{C}_n$  die in den Zyklen 0 bis  $n$  gewerteten Kosten. Mit Formel 2.40 werden die Kosten in Kosteneinheiten diskontiert. Es

bezeichne  $\mathcal{C}_{\text{heute}}$  die Gesamtkosten einer Maßnahme zum Zeitpunkt  $t_0$ . Es gilt:

$$\mathcal{C}_{\text{heute}} = \mathcal{C}_0 + \frac{\mathcal{C}_1}{1 + \mathcal{R}} + \frac{\mathcal{C}_2}{(1 + \mathcal{R})^2} + \dots + \frac{\mathcal{C}_n}{(1 + \mathcal{R})^n} = \sum_{i=0}^n \frac{\mathcal{C}_i}{(1 + \mathcal{R})^i}. \quad (2.40)$$

Die Diskontierung der gewerteten Nutzen  $\mathcal{U}_0, \dots, \mathcal{U}_n$  erfolgt analog:

$$\mathcal{U}_{\text{heute}} = \sum_{i=0}^n \frac{\mathcal{U}_i}{(1 + \mathcal{R})^i}. \quad (2.41)$$

## 2.6 Sensitivitätsanalyse

Ergebnisse gesundheitsökonomischer Analysen unterliegen Unsicherheiten. Mit einer Sensitivitätsanalyse können die Robustheit der Ergebnisse überprüft und kritische Annahmen identifiziert werden. Der erste Schritt einer Sensitivitätsanalyse ist die Auswahl der Variablen, die im Zuge der Sensitivitätsanalyse variiert werden sollen. In der Regel werden alle Variablen ausgewählt, die Schwankungen unterliegen bzw. deren Werte unter Unsicherheit geschätzt wurden. Als zweiter Schritt muss für jede der ausgewählten Variablen festgelegt werden, in welchem Bereich oder unter welcher Verteilungsannahme sie variiert wird. Bei einer klassischen Sensitivitätsanalyse werden einzelne oder mehrere Variablen gleichzeitig innerhalb vorgegebener Intervalle variiert. Bei einer probabilistischen Monte-Carlo-Simulation werden unterschiedliche Ausprägungen der Variablen simuliert, so dass Prognoseintervalle einer vorgegebenen Zielgröße ermittelt werden können. Es handelt sich um eine univariate Sensitivitätsanalyse, wenn jeweils nur eine Variable verändert wird, und um eine multivariate Sensitivitätsanalyse, wenn mehrere Variablen gleichzeitig verändert werden. In der vorliegenden Arbeit wird eine multivariate Monte-Carlo-Simulation durchgeführt, in der alle einfließenden Variablen gleichzeitig variiert werden. (Vgl. Petitti 1994)

# Kapitel 3

## Entwicklung des Markovmodells

Die Entwicklung des Markovmodells, mit dem die Kosten-Nutzen-Relation primärer Prävention zur Vermeidung der koronaren Herzkrankheit berechnet wird, erfolgt schrittweise. Ausgehend von einem einfachen Markovmodell, das im Folgenden als Ausgangsmodell bezeichnet wird, entstehen immer feiner untergliederte Modelle. Dabei werden einige Zustände in disjunkte Subzustände unterteilt. Die schrittweise Entwicklung des Modells ist notwendig, um die Übergangswahrscheinlichkeiten der feineren Modelle zu schätzen. Es werden Ein-Jahreszyklen gewählt, und die Übergänge finden jeweils in der Mitte eines Zyklus statt.

### 3.1 Das Ausgangsmodell

Als Ausgangsmodell dient ein einfaches Modell mit drei Zuständen (siehe Abbildung 3.1 a)). Die Abkürzungen werden in Anlehnung an das Modell von Wendland (2002) gewählt. Der Zustand  $g$  repräsentiert alle Personen, die nicht an der koronaren Herzkrankheit erkrankt sind ( $g$  wie „gesund“, wobei andere Erkrankungen als die KHK *nicht* ausgeschlossen werden). Personen ohne KHK können an der KHK erkranken und würden somit in den Zustand  $h$  („an der KHK erkrankt“) wechseln. Die Zustände  $g$  und  $h$  werden auch als Lebendzustände bezeichnet. Bei der Berechnung der Übergangswahrscheinlichkeiten ist es an einigen Stellen hilfreich, diese

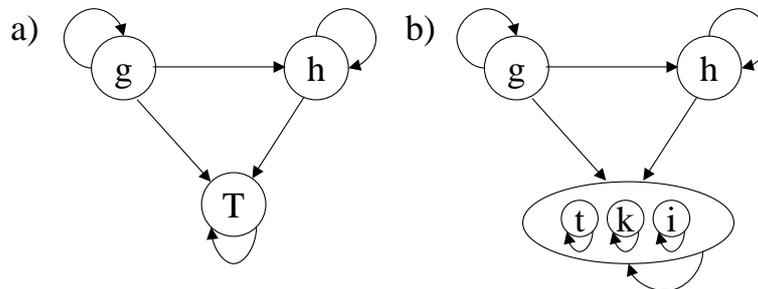


Abbildung 3.1: Das Ausgangsmodell mit den Zuständen „ohne KHK“ ( $g$ ), „KHK“ ( $h$ ) und „Tod“ ( $T$ ); Es werden drei unterschiedliche Todesursachen unterschieden

beiden Zustände zu einem einzigen Zustand  $L$  zusammenzufassen. Falls Personen aus den Lebenszuständen  $g$  und  $h$  sterben, wechseln sie in den Zustand  $T$  („Tod“) über. Der Zustand  $T$  ist ein absorbierender Zustand. Wenn Personen aus dem Zustand  $g$  innerhalb eines Zyklus weder an der KHK erkranken noch versterben, so verbleiben sie im Zustand  $g$ . Dieser Fall wird durch einen Pfeil von  $g$  nach  $g$  als Übergang zu sich selbst repräsentiert. Analog deutet der Pfeil von  $h$  nach  $h$  an, dass eine Person im Folgezyklus immer noch an der koronaren Herzkrankheit erkrankt sein kann. Bei der KHK handelt es sich um eine chronische Krankheit. Eine Heilung ist derzeit nicht möglich. Aus diesem Grund ist der Übergang von  $h$  nach  $g$  nicht möglich. Personen aus dem Zustand  $h$  können in diesem Zustand verbleiben, oder sie sterben und wechseln in den Zustand  $T$ . Jedem Übergang wird eine Übergangswahrscheinlichkeit zugeordnet. Diese Wahrscheinlichkeiten sind in den vorliegenden Markovmodellen nicht konstant, sondern variieren in Abhängigkeit des Alters und des Geschlechts. Dieser Aspekt wird in das Modell integriert, da bei zunehmendem Alter die Inzidenz (Neuerkrankungsrate) und die Mortalität steigen. Generell haben Männer ein höheres Risiko, an der KHK zu erkranken, und sterben im Mittel früher als Frauen.

In dem vorliegenden Modell werden drei verschiedene Todesursachen unterschieden (vgl. Abbildung 3.1) b): Der Tod durch die koronare Herzkrankheit ( $k$ ), der

Tod durch Herzinsuffizienz ( $i$ ) und sonstige Todesursachen ( $t$ ). Somit lässt sich der Zustand  $T$  gemäß der Todesursache weiter unterteilen. Diese Unterteilung ist im finalen Modell nicht mehr notwendig, sie hilft aber bei der Modellierung der Übergangswahrscheinlichkeiten.

### 3.2 Unterteilung des Ausgangsmodells bezüglich des Erkrankungsrisikos

Da das Risikoscreening im Ausgangsmodell nicht integriert ist, wird das Modell erweitert. Es wird zwischen Personen mit hohem und geringem KHK-Risiko unterschieden. Das betrachtete Risiko bezieht sich auf einen begrenzten Zeitraum. Im vorliegenden Modell wird ein Zeithorizont von 10 Jahren gewählt. Das Risiko wird als Wahrscheinlichkeit definiert, im vorgegebenen Zeitraum an der KHK zu erkranken. Im Modell wird davon ausgegangen, dass nur ein einmaliges Screening stattfindet. Nicht gescreent werden Personen, die zum Ausgangszeitpunkt  $t_0$  bereits an der KHK erkrankt sind, da sie nicht erneut erkranken können. Personen mit höherer Erkrankungswahrscheinlichkeit werden als Personen mit positivem Score oder als positiv gescreent bezeichnet und werden durch ein „+“ gekennzeichnet. Im Ausgangsmodell befinden sie sich zum Zeitpunkt  $t_0$  im Zustand  $g$  („ohne KHK“). Die übrigen Personen, die sich zum Ausgangszeitpunkt  $t_0$  im Zustand  $g$  befinden sind die Personen mit geringem KHK-Risiko. Diese werden als negativ gescreente Personen oder als Personen mit negativem Score bezeichnet und werden mit einem „-“ gekennzeichnet. Der Zustand  $g$  wird somit in die Zustände  $g^+$  und  $g^-$  unterteilt. Im Laufe der Zeit können sowohl Personen mit positivem Score als auch Personen mit negativem Score an der KHK erkranken. Daher wird der Zustand  $h$  in drei Gruppen aufgeteilt: Die erste Gruppe wird mit  $h^+$  bezeichnet. Sie umfasst Personen mit KHK, die zum Screeningzeitpunkt zwar frei von der KHK sind, aber ein hohes KHK-Erkrankungsrisiko aufweisen. Zum Screeningzeitpunkt befindet sich keine Person in  $h^+$ , aber Personen

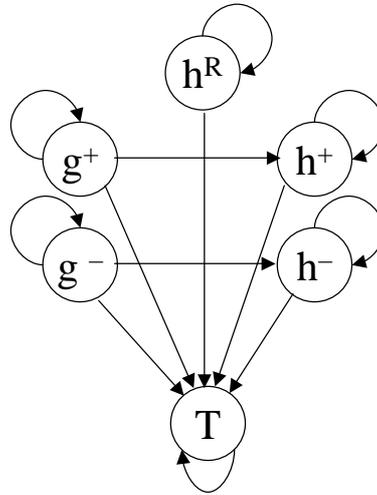


Abbildung 3.2: Das nach Erkrankungsrisiko unterteilte Modell

können im Laufe der Zeit vom Zustand  $g^+$  nach  $h^+$  wechseln. Die zweite Gruppe,  $h^-$ , sind Personen mit KHK, die zum Screeningzeitpunkt noch nicht an der KHK erkrankt sind und denen beim Screening ein geringes KHK-Erkrankungsrisiko zugeordnet wird. Diejenigen, die sich bereits zum Screeningzeitpunkt in  $h$  befinden, werden dem Subzustand  $h^R$  ( $R$  wie Rest) zugeordnet. Das unterteilte Modell ist in Abbildung 3.2 graphisch dargestellt.

### 3.3 Das Präventionsmodell für Personen mit höherem Risiko

Die Durchführung eines Screenings verursacht zusätzliche Kosten; ein höherer Nutzen (gemessen in qualitätsadjustierten gewonnenen Lebensjahren) wird durch das Screening an sich zunächst nicht erzielt. Daher ist es notwendig, dass bei Personen mit hohem Risiko eine Interventions- bzw. Präventionsmaßnahme stattfindet, durch die eine Erkrankung an der KHK unwahrscheinlicher wird. Da Prävention nur bei Personen mit hohem Score stattfindet, wird nur das Submodell von Personen mit positivem Score (siehe Abbildung 3.3) durch Prävention beeinflusst. Findet Prävention

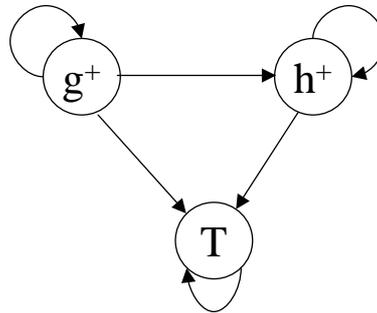


Abbildung 3.3: Submodell für Personen mit hohem Score im Szenario ohne Prävention

statt, so verringern sich die individuellen Erkrankungs- und Sterbewahrscheinlichkeiten. Personen, die ohne Prävention an der KHK erkrankt wären, besitzen im Mittel eine höhere Komorbidität als die Personen, die auch ohne Prävention (noch) nicht an der KHK erkrankt wären. Personen, die früh von  $g^+$  nach  $h^+$  wechseln, verursachen im Vergleich zu denen, die später oder gar nicht von  $g^+$  nach  $h^+$  wechseln, im Mittel höhere Kosten und weisen eine erhöhte Mortalitätsrate auf. Wenn nun Prävention stattfindet, so bleiben auf der Ebene einer einzelnen Person die Krankheiten neben der KHK größtenteils unberührt. Der Anteil der Herzinsuffizienz hingegen, der als Folgeerkrankung auf die KHK zurückzuführen ist, fällt weg. Da die Personengruppe, bei der die KHK durch Prävention vermieden wird, sich strukturell von den übrigen Personen ohne KHK unterscheidet, wird für sie ein eigener Zustand  $g^{V+}$  eingerichtet ( $V$  wie „vermieden“). Es sei angemerkt, dass sich in der Realität nicht unterscheiden lässt, ob sich eine Person im Präventionsszenario im Zustand  $g^+$  oder im Zustand  $g^{V+}$  befindet. Die Modellierung der Übergangswahrscheinlichkeiten wird jedoch durch Erstellung dieses Zustandes erleichtert. Das Präventionsmodell für Personen mit hohem Score ist in Abbildung 3.4 dargestellt.

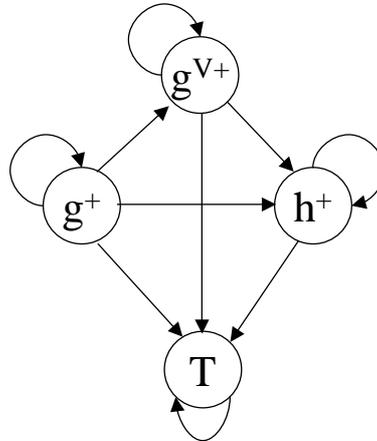


Abbildung 3.4: Präventionsmodell für Personen mit hohem Score

### 3.4 Das finale Markovmodell

Das vollständige (finale) Markovmodell ist in Abbildung 3.5 dargestellt. Es wird für Szenarien mit und ohne Prävention bestimmt. Im Falle, dass keine Prävention stattfindet, existiert der Zustand  $g^{V+}$  nicht. Dies ist gleichbedeutend damit, dass die Übergangswahrscheinlichkeiten zu ihm auf 0 gesetzt werden. Alle weiteren Übergangswahrscheinlichkeiten im Modell verändern sich je nachdem, ob Prävention durchgeführt wird. Im Zuge der Auswertung müssen unterschiedliche Varianten des finalen Modells miteinander verglichen werden. Dazu werden die Auszahlungen, d. h. die Kosten und Nutzen unterschiedlicher Modelle miteinander verglichen.

### 3.5 Die Wahl der Präventionsmaßnahme

Um zusätzlichen Nutzen in Form von qualitätsadjustierten gewonnenen Lebensjahren zu erzielen, wird im Modell Primärprävention durchgeführt. Als präventive Maßnahme zur Vermeidung der KHK kommen mehrere Optionen in Betracht. Prävention kann durch Verhaltensveränderungen oder durch medikamentöse Prävention erfolgen.

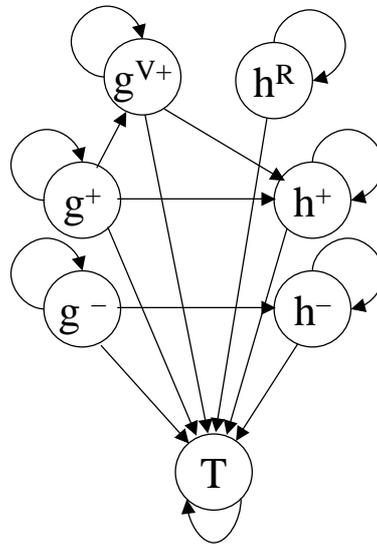


Abbildung 3.5: Das vollständige Markovmodell inklusive Prävention

Die möglichen Verhaltensveränderungen können u. a. in „mehr Bewegung“, „gesündere Ernährung“ (mediterrane Kost) und „Verzicht auf Zigaretten“ weiter unterteilt werden. Dass durch ein derart gesünderes Verhalten die Wahrscheinlichkeit eines koronaren Ereignisses reduziert wird, gilt als hinreichend gesichert (Hu *et al.* 2003, Jousilahti *et al.* 1999, Löllgen 2003). Welche Kosten mit einer verhaltenspräventiven Maßnahme verbunden sind, lässt sich anhand des derzeitigen Stands der Wissenschaft kaum quantifizieren. Es gibt Hinweise darauf, dass bei in Frage kommenden Erwachsenen eine kontinuierliche Motivation durch Außenstehende notwendig ist, damit die Erstgenannten nicht in ihre gewohnten Verhaltensweisen zurückfallen. Da eine Modellierung von Verhaltensveränderungen als Präventionsmaßnahme mit vielen ungewissen Annahmen verbunden ist, wird von dieser abgesehen.

Als medikamentöse Maßnahmen kommen je nach Indikation u. a. Blutdruck senkende Mittel sowie Lipidsenker (Cholesterin senkende Mittel) in Betracht (Vrečer *et al.* 2003; Dahlof *et al.* 2002). Exemplarisch wird eine Primärprävention mit Statinen

modelliert. Statine gehören zu den Lipidsenkern und sind anderen Lipidsenkern im Bereich der Prävention überlegen (Schwabe und Paffrath 2004).

## 3.6 Einmaliges Screening versus multiples Screening

In dem hier vorgestellten Präventionsmodell findet nur ein Screening pro Person statt. Modelle mit multiplem Screening sind wesentlich komplexer. Eine Schwierigkeit bei multiplem Screening besteht darin, dass bei Personen, die nach einem ersten Screening an einer Präventionsmaßnahme teilnehmen, das Ergebnis eines erneuten Screenings durch die Präventionsmaßnahme beeinflusst werden kann. So kann beim zweiten Screening aufgrund der besseren Blutwerte ein geringes KHK-Risiko zugeordnet werden. Brechen Personen die Prävention daraufhin ab, so verschlechtern sich ihre Blutwerte und das KHK-Risiko erhöht sich. Modelle bei denen Personen ohne höheres Risiko wiederholt gescreent werden, sind denkbar, werden hier jedoch nicht behandelt.

# Kapitel 4

## Schätzung der Übergangswahrscheinlichkeiten

### 4.1 Notationen und Vorbemerkungen

Übergangswahrscheinlichkeiten von einem Zustand  $A$  zu einem Zustand  $B$  werden mit  $p_{AB}$  gekennzeichnet. In den vorliegenden Modellen werden Übergangswahrscheinlichkeiten in Abhängigkeit des Alters und des Geschlechts spezifiziert. Diese Abhängigkeit wird durch die Notation

$$p_{AB} = p_{AB}(\text{Alter, Geschlecht}) \quad (4.1)$$

verdeutlicht. Es werden Gleichungssysteme aufgestellt, die für jede Alters-Geschlechtsklasse (das sind in dieser Arbeit Personen gleichen (Einjahres-)Alters und gleichen Geschlechts) gelöst werden. Damit die Formeln übersichtlich bleiben, wird in der Regel auf die Angabe der Abhängigkeit verzichtet, es sei denn, sie ist für das Verständnis der Formeln erforderlich. Das kann vorkommen, wenn z. B. innerhalb einer Formel Übergangswahrscheinlichkeiten, Prävalenzen etc. aus unterschiedlichen Altersgruppen verwendet werden.

Oft ist es notwendig, Kombinationen von Zuständen zu betrachten. Das hängt

u. a. damit zusammen, dass Markovmodelle unterschiedlicher Feinheitgrade betrachtet werden. In feineren Markovmodellen werden Zustände des zugrundeliegenden gröberen Markovmodells weiter unterteilt. Somit kann ein Zustand in Subzustände unterteilt werden. Häufig ist der Anteil von Interesse, den ein Subzustand an einem übergeordneten Zustand einnimmt. Es sei  $A$  ein Subzustand und  $B$  ein übergeordneter Zustand, d. h. es gilt  $A \subseteq B$ . Dann bezeichnet  $a_B^A$  den (erwarteten) Anteil von Personen im Zustand  $A$  an den Personen im Zustand  $B$ .

Häufig werden bedingte Wahrscheinlichkeiten verwendet. In der Darstellung der bedingten Wahrscheinlichkeiten werden Abkürzungen verwendet. So steht „n. KHK-M.“ für die Bedingung, dass nur KHK-Todesfälle vorkommen können, „n. i. M.“ dafür, dass nur Todesfälle durch die Herzinsuffizienz vorkommen können und „n. s. M.“ steht für eine Beschränkung auf sonstige Mortalität, d. h. die Todesursachen KHK und Herzinsuffizienz werden ausgeschlossen.

Beispiel:

Die Wahrscheinlichkeit  $p_{gk}|_{n. \text{KHK-M.}}$  ist die Wahrscheinlichkeit vom Zustand  $g$  zum Zustand  $k$  überzuwechseln, d. h. als bisher Gesunder an der KHK zu sterben, *gegeben*, dass die Person im entsprechenden Jahr nicht an einer anderen Ursache als der KHK verstirbt.

## 4.2 Übergangswahrscheinlichkeiten des Ausgangsmodells

Im ersten Schritt werden die Übergangswahrscheinlichkeiten des Ausgangsmodells geschätzt. Dazu wird ein Gleichungssystem aufgestellt, mit Hilfe dessen die

Übergangswahrscheinlichkeiten alters- und geschlechtsspezifisch geschätzt werden. Für alle Übergangswahrscheinlichkeiten wird die Annahme getroffen, dass sie zwar vom Alter und Geschlecht abhängen, sich aber nicht im Laufe der Jahrzehnte verändern. Das Problem beispielsweise, dass sich Kinder in den kommenden Jahrzehnten anders verhalten und ernähren als es zum Zeitpunkt der Datenerhebung der Fall war, wird in der Modellierung vernachlässigt.

Es wird angenommen, dass die Prävalenz  $\pi$  der koronaren Herzkrankheit in Deutschland eine Funktion in Abhängigkeit des Alters und des Geschlechts ist. Aufgrund des demographischen Wandels verändert sich somit zwar die Prävalenz der KHK in ganz Deutschland, alters- und geschlechtsspezifisch bleibt die Prävalenz jedoch konstant. Aus der Prävalenz und den Übergangswahrscheinlichkeiten einer Kohorte lässt sich die Prävalenz derselben Kohorte einen Zyklus später, also ein Jahr später, berechnen. Da diese mit der Prävalenz einer um ein Jahr gealterten Kohorte übereinstimmen muss, ergibt sich die folgende Gleichung:

$$\pi(\text{Alter} + 1) = \frac{\pi(\text{Alter}) \cdot p_{hh} + (1 - \pi(\text{Alter})) \cdot p_{gh}}{(1 - \pi(\text{Alter})) \cdot (p_{gg} + p_{gh}) + \pi(\text{Alter}) \cdot p_{hh}}. \quad (4.2)$$

Es wird gefordert, dass die Übergangswahrscheinlichkeiten des Ausgangsmodells nicht im Widerspruch zu den Sterbetafeln des Statistischen Bundesamtes stehen. Die Todesfälle werden vom Statistischen Bundesamt ursachenspezifisch dokumentiert. In der Modellierung der vorliegenden Arbeit wird zwischen drei Todesarten unterschieden (siehe Kapitel 3). Es sei  $p_{Lk}$  die jährliche (alters- und geschlechtsspezifische) Sterberate der Bevölkerung an der koronaren Herzkrankheit, d. h. der Übergangswahrscheinlichkeit des zusammengesetzten Zustandes  $L = g \cup h$  aller Lebenden zum Zustand  $k$  (Tod durch KHK). Damit die exogen vorgegebene Wahrscheinlichkeit  $p_{Lk}$  eingehalten wird, muss die folgende Gleichung erfüllt werden:

$$p_{Lk} = (1 - \pi(\text{Alter})) \cdot p_{gk} + \pi(\text{Alter}) \cdot p_{hk}. \quad (4.3)$$

Weiter seien  $p_{Li}$  die Sterberate mit Todesursache Herzinsuffizienz und  $p_{Lt}$  die Sterberate an sonstigen Ursachen. Mit ihnen wird analog verfahren (siehe Gleichungen

A.1 und A.2; Anhang A.1).

Es werden gewisse Odds Ratios als bekannt angenommen, die ursachenspezifisch angeben, wie stark bei Personen mit KHK ( $h$ ) die Mortalität gegenüber Personen ohne KHK ( $g$ ) erhöht ist. Diese Odds Ratios beziehen sich auf die bedingten Wahrscheinlichkeiten, gegeben dass keine sonstige Todesursache vorliegt. So beziehe sich  $\omega_k$  auf die bedingte KHK-spezifische Mortalität:

$$\omega_k = \frac{p_{gk}|_{n. \text{ KHK-M.}}}{1 - p_{gk}|_{n. \text{ KHK-M.}}} \cdot \frac{1 - p_{hk}|_{n. \text{ KHK-M.}}}{p_{hk}|_{n. \text{ KHK-M.}}} \quad (4.4)$$

Analog seien die Odds Ratios  $\omega_i$  und  $\omega_t$  für Mortalität durch Herzinsuffizienz und für sonstige Mortalität definiert (Gleichungen A.3 und A.4 in Anhang A.1). Die bedingte Wahrscheinlichkeit  $p_{gk}|_{n. \text{ KHK-M.}}$  wird dabei folgendermaßen berechnet:

$$p_{gk}|_{n. \text{ KHK-M.}} = \frac{p_{gk}}{1 - p_{gi} - p_{gt}} \quad (4.5)$$

Die Berechnung weiterer bedingter Übergangswahrscheinlichkeiten erfolgt analog (siehe Anhang A). Die Übergangswahrscheinlichkeiten eines jeweiligen Zustandes müssen 1 ergeben. Die Gesamtsterbewahrscheinlichkeit für Personen mit und ohne KHK ergeben sich aus den ursachenspezifischen Sterbewahrscheinlichkeiten (Gleichungen siehe Anhang A.1).

Das Gleichungssystem (Gleichungen 4.2 bis 4.5 und A.1 bis A.13) besteht für jede Alters-Geschlechts-Kohorte aus 17 Gleichungen und 17 Unbekannten. Neben den Gleichungen muss weiterhin die Bedingung gelten, dass alle Wahrscheinlichkeiten im kompakten Intervall von 0 bis 1 liegen. Dieses Gleichungssystem sowie die weiteren Gleichungssysteme dieses Kapitels werden mit Maple (Version 9.01, vgl. Monagan *et al.* 2003) für jede Werte-Kombination gelöst.

Eine Darstellung der Lösungen dieses Gleichungssystems sowie der in den nächsten Abschnitten behandelten Gleichungssysteme ist nicht möglich. Für die einfacheren

der Gleichungssysteme kann zwar mit Maple eine parametrische Lösung ermittelt werden, der Maple-Output ist jedoch sehr umfangreich und umfasst mehr als 100 Seiten. Hinzu kommt, dass dieser Output sehr unhandlich ist, da er den Computer sehr stark auslastet. Bei den meisten der Gleichungssysteme ist es aufgrund der Rechenleistung der zur Verfügung stehenden Computer nicht möglich, das Gleichungssystem mit Variablen zu lösen. Für explizit vorgegebene Wertekombinationen kann jedoch eine Lösung ermittelt werden.

### 4.3 Erste Unterteilung der Personen mit KHK

Die Population ohne KHK, d. h. alle Personen in  $g$ , soll zum Zeitpunkt  $t_0$  gescreent werden. Personen mit KHK ( $h$ ) werden nicht gescreent, da sie nicht erneut an der KHK erkranken können.

Es muss zwischen Personen unterschieden werden, die zum Screeningzeitpunkt bereits an der KHK erkrankt sind (Subzustand  $h^R$ ) und Personen, die erst nach dem Screening an der KHK erkranken (Subzustand  $h^N$ ). Um den Zustand  $h$  in  $h^N$  und  $h^R$  zu unterteilen, wird als Hilfsmittel der Zustand  $g^0 := g \cup h^N$  definiert. Er umfasst alle zum Zeitpunkt  $t$ ,  $t \geq t_0$ , lebenden Personen, die sich zum Ausgangszeitpunkt  $t_0$  in  $g$  befinden.

Um die Übergangswahrscheinlichkeiten für die neu geschaffenen Zustände  $h^N$  und  $h^R$  zu berechnen, müssen Annahmen getroffen werden. Dazu wird im Folgenden unterstellt, dass sich gewisse Odds Ratios zwischen den Subpopulationen  $g^0$  und  $h^R$  im weiteren Zeitverlauf nicht verändern. Es sei angemerkt, dass die Personenkollektive  $g^0$  und  $h^R$  sich im Laufe der Zeit verändern, da aus beiden Kollektiven Personen versterben. Daher kommt es zu strukturellen Veränderungen, wodurch die Annahme konstanter Odds Ratios in der Plausibilität leicht abgeschwächt wird.

### 414.3 Erste Unterteilung der Personen mit koronarer Herzkrankheit (KHK)

Das erste Odds Ratio  $OR_{\text{KHK-M.}}^{g^0 h^R}$ , das als konstant angenommen wird, ist folgendermaßen definiert:

$$OR_{\text{KHK-M.}}^{g^0 h^R} := \frac{p_{g^0 k} |_{\text{n. KHK-M.}}}{1 - p_{g^0 k} |_{\text{n. KHK-M.}}} \cdot \frac{1 - p_{h^R k} |_{\text{n. KHK-M.}}}{p_{h^R k} |_{\text{n. KHK-M.}}} . \quad (4.6)$$

Es bezieht sich auf die bedingte KHK-spezifische Sterbewahrscheinlichkeit zwischen den Populationen der Zustände  $g^0$  und  $h^R$ . Die Definition der weiteren beiden als konstant angenommenen Odds Ratios  $OR_{\text{i. M.}}^{g^0 h^R}$  und  $OR_{\text{s. M.}}^{g^0 h^R}$  bezüglich der Mortalität durch Herzinsuffizienz und der sonstigen Mortalität erfolgt analog (Gleichungen A.15 und A.14). Zum Screeningzeitpunkt  $t_0$  gilt:

$$g^0 = g \quad \text{und} \quad h^R = h.$$

Daher können die Odds Ratios  $OR_{\text{i. M.}}^{g^0 h^R}$  und  $OR_{\text{s. M.}}^{g^0 h^R}$  aus den Übergangswahrscheinlichkeiten des Ausgangsmodells direkt berechnet werden.

Die Übergangswahrscheinlichkeiten des Zustands  $h$  müssen sich anteilmäßig aus den Übergangswahrscheinlichkeiten der Zustände  $h^N$  und  $h^R$  ergeben. Beispielsweise muss für die KHK-spezifische Sterbewahrscheinlichkeit die folgende Gleichung gelten:

$$p_{hk} = a_h^{h^N} \cdot p_{h^N k} + a_h^{h^R} \cdot p_{h^R k} , \quad (4.7)$$

wobei  $a_h^{h^N}$  und  $a_h^{h^R}$ , wie bereits in Abschnitt 4.1 angedeutet, die Anteile der nach bzw. vor dem Screeningzeitpunkt Erkrankten an allen Erkrankten bezeichnet. Die Übergangswahrscheinlichkeiten des zusammengesetzten Zustands  $g^0$  müssen aus den Übergangswahrscheinlichkeiten seiner Subzustände  $g$  und  $h^N$  resultieren. Für die KHK-spezifische Sterbewahrscheinlichkeit gilt:

$$p_{g^0 k} = a_{g^0}^g \cdot p_{gk} + a_{g^0}^{h^N} \cdot p_{h^N k} . \quad (4.8)$$

Analog müssen die Beziehungen 4.7 und 4.8 auch für die übrigen ursachenspezifischen Sterbewahrscheinlichkeiten gelten (Formeln A.16 bis A.19). Weiter gelten

für die neuen Zustände  $h^N$  und  $h^R$  die trivialen Bedingungen, dass die Summe der Übergangswahrscheinlichkeiten eines Zustands 1 ergibt und dass die Gesamtsterberate gleich der Summe der ursachenspezifischen Sterbewahrscheinlichkeiten ist (Formeln A.20 bis A.23).

Personen in  $g$ , die an der KHK erkranken, gehen in den neu geschaffenen Zustand  $h^N$  über. Die Übergangswahrscheinlichkeit  $p_{gh}$  ist im neuen Modell nicht mehr enthalten und wird durch  $p_{gh^N}$  ersetzt. Es gilt:

$$p_{gh^N} := p_{gh}. \quad (4.9)$$

Die Bevölkerungsanteile  $a_{g^0}^g$ ,  $a_{g^0}^{h^N}$ ,  $a_h^{h^N}$  und  $a_h^{h^R}$  müssen induktiv bestimmt werden. Die Rekursionsformeln des Induktionsschritts hängen u. a. von den Übergangswahrscheinlichkeiten des Zustands  $h^N$  ab. Da diese für den Screeningzeitpunkt  $t_0$  nicht definiert sind, erfolgt die Initialisierung für den ersten Zyklus nach dem Screeningzeitpunkt, d. h. für  $t_1$ .

Initialisierung:

$$a_{g^0}^g(t_1) = \frac{p_{gg}(t_0)}{p_{gg}(t_0) + p_{gh}(t_0)}, \quad (4.10)$$

$$a_{g^0}^{h^N}(t_1) = \frac{p_{gh}(t_0)}{p_{gg}(t_0) + p_{gh}(t_0)}, \quad (4.11)$$

$$a_h^{h^N}(t_1) = \frac{(1 - \pi(t_0)) \cdot p_{gh}(t_0)}{\pi(t_0) \cdot p_{hh}(t_0) + (1 - \pi(t_0)) \cdot p_{gh}(t_0)}, \quad (4.12)$$

$$a_h^{h^R}(t_1) = \frac{\pi(t_0) \cdot p_{hh}(t_0)}{\pi(t_0) \cdot p_{hh}(t_0) + (1 - \pi(t_0)) \cdot p_{gh}(t_0)}, \quad (4.13)$$

$$a_L^g(t_1) = \frac{(1 - \pi(t_0)) \cdot p_{gg}(t_0)}{(1 - \pi(t_0)) \cdot (p_{gg}(t_0) + p_{gh^N}(t_0)) + \pi(t_0) \cdot p_{h^R h^R}(t_0)}, \quad (4.14)$$

$$a_L^{h^N}(t_1) = \frac{(1 - \pi(t_0)) \cdot p_{gh^N}(t_0)}{(1 - \pi(t_0)) \cdot (p_{gg}(t_0) + p_{gh^N}(t_0)) + \pi(t_0) \cdot p_{h^R h^R}(t_0)}, \quad (4.15)$$

### 434.3 Erste Unterteilung der Personen mit koronarer Herzkrankheit (KHK)

$$a_L^{h^R}(t_1) = \frac{\pi(t_0) \cdot p_{h^R h^R}(t_0)}{(1 - \pi(t_0)) \cdot (p_{gg}(t_0) + p_{gh^N}(t_0)) + \pi(t_0) \cdot p_{h^R h^R}(t_0)}. \quad (4.16)$$

Induktionsschritt:

$$a_{g^0}^g(t_{i+1}) = \frac{a_{g^0}^g(t_i) \cdot p_{gg}(t_i)}{a_{g^0}^g(t_i) \cdot (p_{gg}(t_i) + p_{gh^N}(t_i)) + a_{g^0}^{h^N} \cdot p_{h^N h^N}(t_i)}, \quad (4.17)$$

$$a_{g^0}^{h^N}(t_{i+1}) = \frac{a_{g^0}^g(t_i) \cdot p_{gh^N}(t_i) + a_{g^0}^{h^N} \cdot p_{h^N h^N}(t_i)}{a_{g^0}^g(t_i) \cdot (p_{gg}(t_i) + p_{gh^N}(t_i)) + a_{g^0}^{h^N} \cdot p_{h^N h^N}(t_i)}, \quad (4.18)$$

$$a_h^{h^N}(t_{i+1}) = \frac{a_L^g(t_i) \cdot p_{gh^N}(t_i) + a_L^{h^N}(t_i) \cdot p_{h^N h^N}(t_i)}{a_L^g(t_i) \cdot p_{gh^N}(t_i) + a_L^{h^N}(t_i) \cdot p_{h^N h^N}(t_i) + a_L^{h^R}(t_i) \cdot p_{h^R h^R}(t_i)}, \quad (4.19)$$

$$a_h^{h^R}(t_{i+1}) = \frac{a_L^{h^R}(t_i) \cdot p_{h^R h^R}(t_i)}{a_L^g(t_i) \cdot p_{gh^N}(t_i) + a_L^{h^N}(t_i) \cdot p_{h^N h^N}(t_i) + a_L^{h^R}(t_i) \cdot p_{h^R h^R}(t_i)}, \quad (4.20)$$

$$a_L^g(t_{i+1}) = \frac{a_L^g(t_i) \cdot p_{gg}(t_i)}{a_L^g(t_i) \cdot (p_{gg}(t_i) + p_{gh^N}(t_i)) + a_L^{h^N}(t_i) \cdot p_{h^N h^N}(t_i) + a_L^{h^R}(t_i) \cdot p_{h^R h^R}(t_i)}, \quad (4.21)$$

$$a_L^{h^N}(t_{i+1}) = \frac{a_L^g(t_i) \cdot p_{gh^N}(t_i) + a_L^{h^N}(t_i) \cdot p_{h^N h^N}(t_i)}{a_L^g(t_i) \cdot (p_{gg}(t_i) + p_{gh^N}(t_i)) + a_L^{h^N}(t_i) \cdot p_{h^N h^N}(t_i) + a_L^{h^R}(t_i) \cdot p_{h^R h^R}(t_i)}, \quad (4.22)$$

$$a_L^{h^R}(t_{i+1}) = \frac{a_L^{h^R}(t_i) \cdot p_{h^R h^R}(t_i)}{a_L^g(t_i) \cdot (p_{gg}(t_i) + p_{gh^N}(t_i)) + a_L^{h^N}(t_i) \cdot p_{h^N h^N}(t_i) + a_L^{h^R}(t_i) \cdot p_{h^R h^R}(t_i)}. \quad (4.23)$$

## 4.4 Unterteilung der Gesunden nach KHK-Erkrankungsrisiko

In einem weiteren Schritt werden die Personen ohne KHK ( $g$ ) in Personen mit positivem Score ( $g^+$ ) und Personen mit negativem Score ( $g^-$ ) unterteilt. Dadurch entstehen zwei neue Zustände, die den alten Zustand  $g$  ersetzen. Es wird unterstellt, dass der Anteil  $a_g^{g^+}(t_0)$  der Personen, die zum Zeitpunkt  $t_0$  beim Screening einen positiven Score erhalten, bekannt ist. Weiterhin werden zur Unterteilung gewisse Odds Ratios verwendet. Es ist davon auszugehen, dass die Aussagekraft des Screenings im Laufe der Zeit nachlässt. Dies lässt sich dadurch modellieren, dass die im Folgenden betrachteten Odds Ratios sich im Laufe der Zeit der 1 annähern (siehe Abschnitte 4.5 und 4.6). Vorerst wird unterstellt, dass diese Odds Ratios für jeden Zeitpunkt bekannt sind.

Das erste Odds Ratio  $\gamma$  ist ein Maß dafür, wie stark sich die KHK-Eintrittswahrscheinlichkeit zwischen Personen mit positivem und negativem Score unterscheidet, gegeben dass kein Tod einer anderen Ursache als der KHK auftritt. Das bedeutet:

$$\gamma = \frac{(p_{g^+h^N} + p_{g^+k})|_{\text{n. KHK-M.}}}{1 - (p_{g^+h^N} + p_{g^+k})|_{\text{n. KHK-M.}}} \cdot \frac{1 - (p_{g^-h^N} + p_{g^-k})|_{\text{n. KHK-M.}}}{(p_{g^-h^N} + p_{g^-k})|_{\text{n. KHK-M.}}} \quad (4.24)$$

Die auf Herzinsuffizienz basierende Mortalität wird mit dem Odds Ratio  $\psi_i$  und die sonstige Mortalität mit dem Odds Ratio  $\psi_t$  aufgeteilt. Sinkt  $\gamma$ , so sinken auch  $\psi_i$  und  $\psi_t$  (vgl. Abschnitt 4.6). Es soll gelten:

$$\psi_i = \frac{p_{g^+i}|_{\text{n. i. M.}}}{1 - p_{g^+i}|_{\text{n. i. M.}}} \cdot \frac{1 - p_{g^-i}|_{\text{n. i. M.}}}{p_{g^-i}|_{\text{n. i. M.}}}, \quad (4.25)$$

$$\psi_t = \frac{p_{g^+t}|_{\text{n. s. T.}}}{1 - p_{g^+t}|_{\text{n. s. M.}}} \cdot \frac{1 - p_{g^-t}|_{\text{n. s. M.}}}{p_{g^-t}|_{\text{n. s. M.}}}. \quad (4.26)$$

Damit das Gleichungssystem lösbar ist, muss eine weitere Annahme getroffen werden. Und zwar wird gefordert, dass die Fatalitätsrate (Anteil der tödlich verlaufenden

KHK-Erstereignisse) bei Personen mit hohem Score und niedrigem Score in einer Alters-Geschlechtsklasse gleich ist. Es gilt:

$$\frac{p_{g^+k}}{p_{g^+k} + p_{g^+h^N}} = \frac{p_{g^-k}}{p_{g^-k} + p_{g^-h^N}}. \quad (4.27)$$

Die Übergangswahrscheinlichkeiten der neuen Zustände  $g^+$  und  $g^-$  müssen, gewichtet mit den Anteilen der Personen, die sich in den Zuständen befinden, die Übergangswahrscheinlichkeiten von  $g$  ergeben (Formeln A.36 bis A.39). Die Gesamtmortalität ergibt sich aus den Teilmortalitäten und die Wahrscheinlichkeiten eines Zustandes addieren sich zu 1 (Formeln A.40 bis A.43). Eine weitere Bedingung ist es, dass alle Wahrscheinlichkeiten im kompakten Intervall von 0 bis 1 liegen müssen. Die Anteile  $a_g^{g^+}$  und  $a_g^{g^-}$  werden induktiv bestimmt.

Initialisierung zum Zeitpunkt  $t_0$ :

Der Anteil  $a_g^{g^+}(t_0)$  ist gegeben, d. h. er fließt als exogene Variable in das Gleichungssystem ein. Es werden im Laufe der Analyse für  $a_g^{g^+}(t_0)$  unterschiedliche Werte eingesetzt. Bei diesem Anteil handelt es sich um einen Parameter, mit dessen Hilfe die Sensitivität und die Spezifität des Modells gesteuert werden (vgl. Abschnitt 7.2).

$$a_g^{g^-}(t_0) = 1 - a_g^{g^+}(t_0). \quad (4.28)$$

Induktionsschritt:

$$a_g^{g^+}(t_{i+1}) = \frac{a_g^{g^+}(t_i) \cdot p_{g^+g^+}(t_i)}{a_g^{g^+}(t_i) \cdot p_{g^+g^+}(t_i) + a_g^{g^-}(t_i) \cdot p_{g^-g^-}(t_i)}, \quad (4.29)$$

$$a_g^{g^-}(t_{i+1}) = \frac{a_g^{g^-}(t_i) \cdot p_{g^-g^-}(t_i)}{a_g^{g^+}(t_i) \cdot p_{g^+g^+}(t_i) + a_g^{g^-}(t_i) \cdot p_{g^-g^-}(t_i)}. \quad (4.30)$$

## 4.5 Zeitlicher Verlauf der KHK-spezifischen Mortalität

Je länger das Screening zurück liegt, desto geringer wird die Aussagekraft des Screenings. Das heißt, dass das Odds Ratio  $\gamma$  sich im Laufe der Zeit der 1 annähert. Das Odds Ratio  $\gamma$  ist eine Funktion, die von der Zeit abhängt. Dieser Zusammenhang wird durch folgende Notation ausgedrückt:

$$\gamma = \gamma(t_i), \quad i = 0, 1, 2, \dots \quad (4.31)$$

Der folgende funktionaler Zusammenhang wird gewählt (vgl. Abbildung 4.1):

$$\gamma(t_i) := \gamma_0 \cdot \lambda^i + (1 - \lambda^i). \quad (4.32)$$

Dabei sind  $\gamma_0$  und  $\lambda$  Konstanten, die prinzipiell vom Startalter und vom Geschlecht abhängen können. In dieser Arbeit wird jedoch von einem konstanten  $\lambda$  ausgegangen. Im Zuge der Analyse müssen unterschiedliche Werte für  $\gamma_0$  eingesetzt werden (siehe Abschnitt 7.2). Der Parameter  $\lambda$  wird in Abschnitt 5.4 geschätzt.

## 4.6 Zeitlicher Verlauf nicht KHK-bedingter Mortalität

Es ist davon auszugehen, dass Personen mit höherem KHK-Risiko auch einer erhöhten nicht KHK-spezifischen Mortalität ausgesetzt sind. Ähnlich wie bei der Unterteilung der KHK-spezifischen Mortalität, bei der das Odds Ratio  $\gamma$  festgelegt wird, werden nun zwei Odds Ratios festgesetzt, mit der die nicht KHK-spezifische Mortalität bei Personen ohne KHK aufgeteilt wird. Bei diesen Odds Ratios handelt es sich um  $\psi_i$  und  $\psi_t$  (siehe Formeln 4.25 und 4.26). Es ist sinnvoll, beide dieser Odds Ratios als Funktionen in Abhängigkeit von  $\gamma$  zu modellieren. Gibt es beim Screeningverfahren eine bessere Trennschärfe der KHK-Mortalität, so steigt auch

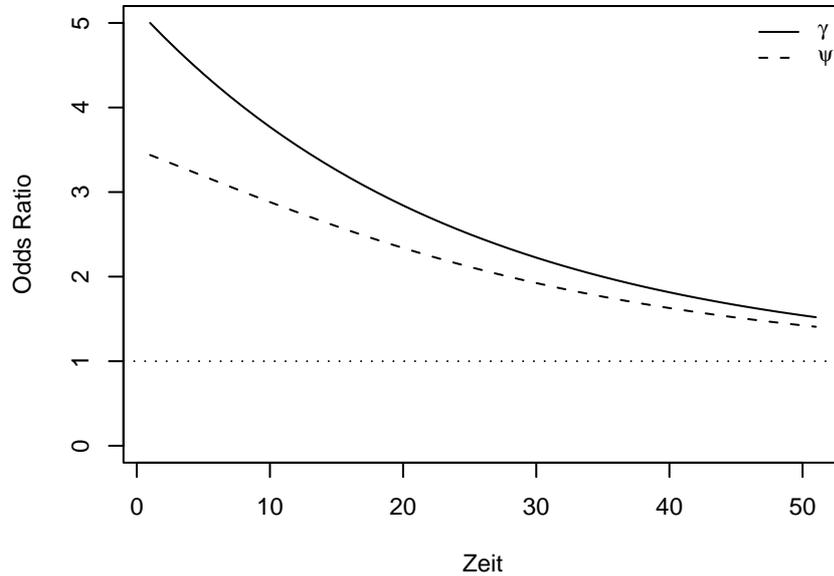


Abbildung 4.1: Zeitlicher Verlauf gewisser Odds Ratios zwischen Personen mit positivem und negativem Score;  $\lambda = 0,96$ ,  $\gamma_0 = 5$  (vgl. Formel 4.32) und  $\psi^{\text{sup}} = 5$  (vgl. Formel 4.45)

die Trennschärfe bezüglich der sonstigen Mortalität:

$$\gamma \uparrow \Rightarrow \psi_i(\gamma) \uparrow, \psi_t(\gamma) \uparrow. \quad (4.33)$$

Da es weitere Eigenschaften gibt, die  $\psi_i(\gamma)$  und  $\psi_t(\gamma)$  gleichermaßen erfüllen sollen, wird im folgenden für beide die gemeinsame Notation  $\psi(\gamma)$  verwendet.

Wenn das Screeningverfahren bezüglich einer Erkrankung an der KHK keine Aussagekraft besitzt, so soll es auch bezüglich der nicht KHK-bedingten Mortalität keine Aussagekraft besitzen. Beide Odds Ratios ( $\gamma$  und  $\psi$ ) nehmen in diesem Fall den Wert 1 an:

$$\psi(1) = 1. \quad (4.34)$$

Kommen durch zwei verschiedene Screeningverfahren dieselben Personengruppen zustande, werden jedoch die Personen, die beim ersten Verfahren als positiv einge-

stuft werden, beim zweiten Verfahren als negativ eingestuft und umgekehrt, so muss sich für  $\psi$ , gemäß der Rechenregeln für Odds Ratios, beim zweiten Verfahren der Kehrwert des ersten Verfahrens ergeben. D. h.

$$\psi\left(\frac{1}{\bar{\gamma}}\right) = \frac{1}{\psi(\bar{\gamma})} \quad \forall \quad \bar{\gamma} \in \mathbb{R}^+ . \quad (4.35)$$

Das Odds Ratio  $\psi$  soll nach oben beschränkt sein. Genauer soll gelten:

$$\lim_{\gamma \rightarrow \infty} \psi(\gamma) = \psi^{\text{sup}} < \infty . \quad (4.36)$$

Der Grund dafür ist folgender: Wenn das Odds Ratio  $\gamma$  bereits sehr hoch ist, so sind die Unterschiede in der KHK-Erkrankungsrate und Mortalität fast ausgereizt. Nahezu alle Erst-KHK-Todesfälle\* finden bei Personen mit positivem Score statt. Eine weitere Erhöhung des Odds Ratios  $\gamma$  ist zwar möglich, hat aber nur noch eine geringe Auswirkung. Wäre  $\psi(\gamma)$  nicht nach oben beschränkt, so wäre es unter Umständen möglich, dass *alle* sonstigen Todesfälle bei Personen mit positivem Score eintreten oder sogar nahezu alle Personen mit positivem Score an einem sonstigen Tod im entsprechenden Zyklus versterben. Dieser Effekt entspräche einer extremen Abweichung von der Realität. Dadurch, dass  $\psi(\gamma)$  nach oben begrenzt ist, wird dieser Effekt vermieden.

Weiterhin ist es sinnvoll anzunehmen, dass das KHK-spezifische Odds Ratio größer als das nicht KHK-spezifische Odds Ratio ist. Daher wird gefordert, dass

$$\psi(\gamma) < \gamma \quad \text{für} \quad \gamma > 1 \quad (4.37)$$

ist. Graphisch bedeutet dies, dass die Funktion  $\psi(\gamma)$  im Bereich  $\gamma > 1$  unterhalb der Winkelhalbierenden verläuft. Weiterhin sollte  $\psi$  in  $\gamma$  stetig, monoton steigend und möglichst glatt sein.

---

\*Das sind Übergänge von  $g$  (ohne KHK) nach  $k$  (Tod durch KHK), ohne über den Zustand  $h$  (KHK) zu gehen.

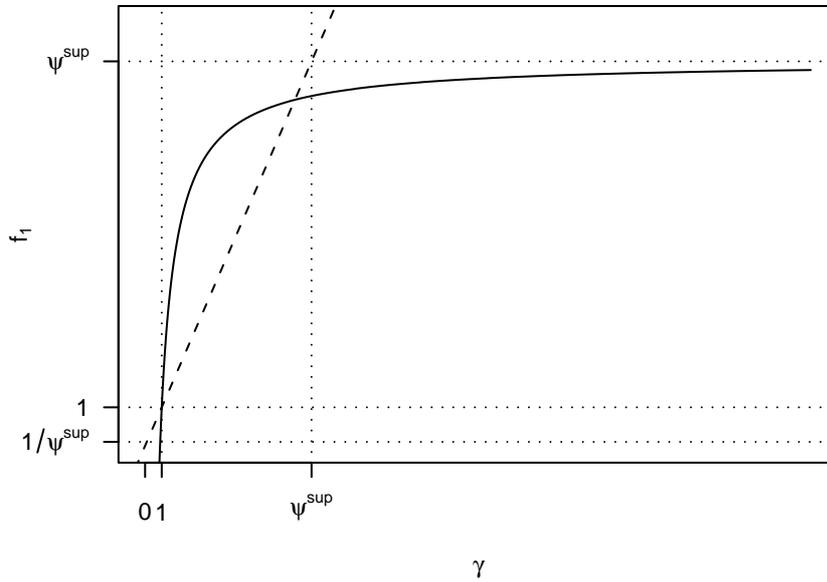


Abbildung 4.2: Verlauf der Funktion  $f_1(\gamma)$ : Sie geht durch den Punkt  $(1, 1)$ , konvergiert gegen  $\psi^{\text{sup}}$ , verläuft im Bereich  $(1, \infty)$  aber nicht überall unterhalb der Winkelhalbierenden

Es ist nicht trivial, eine Funktion  $\psi(\gamma)$  zu konstruieren, die Bedingungen 4.33 bis 4.37 erfüllt. Terme der Form  $\gamma^{1/\kappa}$ ,  $\kappa > 1$ , wie zum Beispiel die Quadratwurzel, verlaufen zwar im Bereich von 1 bis unendlich unterhalb der Winkelhalbierenden, sind aber nicht beschränkt. Die einfachsten beschränkten Funktionen mit einem maximal vorgegebenen  $\psi^{\text{sup}}$  bleiben nicht unterhalb der Winkelhalbierenden. Ein Beispiel einer solchen Funktion ist die Funktion  $f_1$ :

$$f_1(\gamma) := \psi^{\text{sup}} - \frac{\psi^{\text{sup}} - 1}{\gamma}, \quad (4.38)$$

$$f_1(1) = \psi^{\text{sup}} - \psi^{\text{sup}} + 1 = 1, \quad (4.39)$$

$$\lim_{\gamma \rightarrow \infty} f_1(\gamma) = \psi^{\text{sup}}, \quad (4.40)$$

die in Abbildung 4.2 dargestellt ist. Aus  $f_1$  kann eine Funktion hergeleitet werden, die die vorgegebenen Kriterien erfüllt. Dabei wird  $f_1$  horizontal (nach rechts) verschoben, wobei  $f_1$  um so stärker verschoben wird, je höher ihre Funktionswerte

sind. Somit bliebe  $f_1(1)$  unverändert auf der 1.

Um die gesuchte Funktion zu ermitteln wird zunächst die Umkehrfunktion  $f_1^{-1} : [1, \psi^{\text{sup}}] \rightarrow [1, \infty]$  von  $f_1 : [1, \infty] \rightarrow [1, \psi^{\text{sup}}]$  bestimmt:

$$f_1^{-1}(y) = \frac{\psi^{\text{sup}} - 1}{\psi^{\text{sup}} - y}. \quad (4.41)$$

Wird zu  $f_1^{-1}(y)$  der Term  $(y - 1)$  addiert, so resultiert eine Umkehrfunktion einer Funktion, die die zuvor aufgestellten Kriterien erfüllt. Diese Umkehrfunktion  $f_2^{-1} : [1, \psi^{\text{sup}}] \rightarrow [1, \infty]$  lautet:

$$\begin{aligned} f_2^{-1}(y) &:= f_1^{-1}(y) + (y - 1) \\ &= \frac{\psi^{\text{sup}} - 1}{\psi^{\text{sup}} - y} + (y - 1). \end{aligned} \quad (4.42)$$

Wird Gleichung 4.42 gleich  $\gamma$  gesetzt und nach  $y$  gelöst, so ergibt sich

$$y = \frac{\psi^{\text{sup}} + \gamma + 1}{2} \pm \sqrt{\frac{(\psi^{\text{sup}} + \gamma + 1)^2}{4} - \gamma \cdot \psi^{\text{sup}} - 1}. \quad (4.43)$$

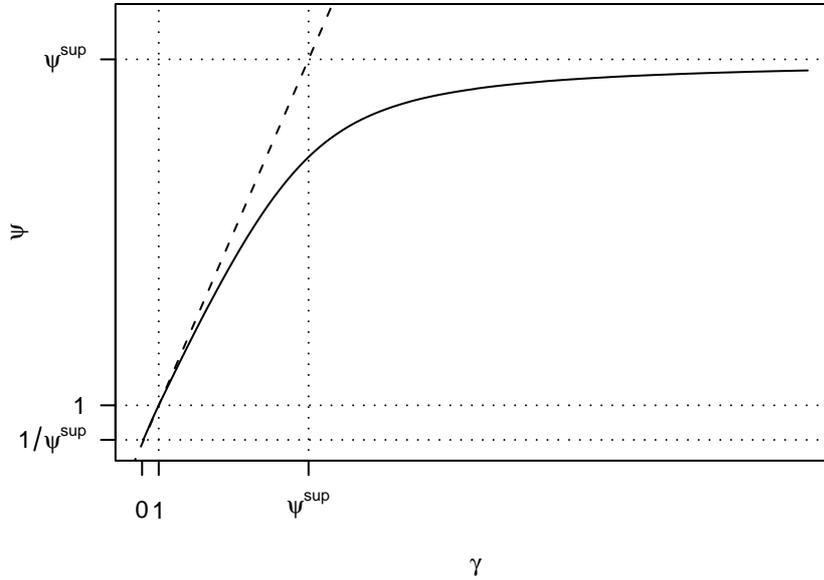
Nur die Lösung mit der negativen Wurzel kommt in Frage, da die Funktion  $f_2$  sonst gegen unendlich ginge. Daher gilt

$$f_2(\gamma) = \frac{\psi^{\text{sup}} + \gamma + 1}{2} - \sqrt{\frac{(\psi^{\text{sup}} + \gamma + 1)^2}{4} - \gamma \cdot \psi^{\text{sup}} - 1}. \quad (4.44)$$

Die Funktion  $\psi(\gamma)$  wird auch für Werte zwischen 0 und 1 definiert:

$$\psi(\gamma) := \begin{cases} f_2(\gamma) & \text{für } \gamma \geq 1 \\ \frac{1}{f_2(1/\gamma)} & \text{für } \gamma \in (0, 1) \\ 1/\psi^{\text{sup}} & \text{für } \gamma = 0. \end{cases} \quad (4.45)$$

Die Funktion  $\psi(\gamma)$  ist in Abbildung 4.3 dargestellt. Die sich implizit ergebende Funktion  $\psi(\text{Zyklus}) = \psi(\gamma(\text{Zyklus}))$  wird in Abbildung 4.1 der Funktion  $\gamma(\text{Zyklus})$  gegenübergestellt.

Abbildung 4.3: Verlauf der Funktion  $\psi(\gamma)$ 

## 4.7 Zweite Unterteilung der Personen mit KHK

Der Zustand  $h^N$  wird weiter unterteilt. Bisher umfasst er alle Personen mit koronarer Herzkrankheit, die zum Screeningzeitpunkt, d. h. zum Zeitpunkt  $t_0$ , nicht an der KHK erkrankt waren. Er wird durch die Zustände  $h^+$  und  $h^-$  ersetzt. Der Zustand  $h^+$  umfasst KHK-Erkrankte mit positivem Score zum Screeningzeitpunkt, der Zustand  $h^-$  umfasst die KHK-Erkrankten, denen zum Zeitpunkt  $t_0$  ein geringes KHK-Erkrankungsrisiko zugeordnet wurde. Es wird unterstellt, dass die Odds Ratios der bedingten ursachenspezifischen Mortalitätsraten zwischen  $g^+$  zu  $g^-$  und  $h^+$  zu  $h^-$  identisch sind (Formeln siehe Abschnitt A.4).

Im neu entstehenden Modell werden die Übergänge zu  $h^N$  durch die Übergänge zu  $h^+$  und zu  $h^-$  ersetzt. Es gilt:

$$p_{g^+h^+} := p_{g^+h^N}, \quad (4.46)$$

$$p_{g^-h^-} := p_{g^-h^N}. \quad (4.47)$$

Das zu lösende Gleichungssystem besteht aus 30 Gleichungen mit 30 Unbekannten: Gleichungen durch die Annahme konstanter Odds Ratios (Formeln A.47 bis A.49), Gleichungen durch die Definition der Odds Ratios (A.50 bis A.55), Gleichungen bedingter Wahrscheinlichkeiten (Formeln A.32 bis A.35 und A.56 bis A.63), Bedingungen, dass untergeordnete Zustände sich zu ihren übergeordneten Zuständen ergeben (Formeln A.44 bis A.46) sowie trivialer Gleichungen (Formeln A.64 bis A.67 sowie 4.46 und 4.47). Zusätzlich müssen alle Wahrscheinlichkeiten im kompakten Intervall von 0 bis 1 liegen.

Die im Gleichungssystem vorkommenden Anteile  $a_{h^N}^{h^+}$  und  $a_{h^N}^{h^-}$  müssen induktiv bestimmt werden. Sie sind nicht für  $t_0$  definiert, da sich zu diesem Zeitpunkt keine Personen in den zugehörigen Zuständen  $h^N$ ,  $h^+$  und  $h^-$  befinden. Zu späteren Zeitpunkten können sie mit folgenden Formeln berechnet werden:

$$a_{h^N}^{h^+} = \frac{a_{g^0}^{h^+}}{a_{g^0}^{h^+} + a_{g^0}^{h^-}} \quad \text{für } t > t_0, \quad (4.48)$$

$$a_{h^N}^{h^-} = \frac{a_{g^0}^{h^-}}{a_{g^0}^{h^+} + a_{g^0}^{h^-}} \quad \text{für } t > t_0. \quad (4.49)$$

Im eigentlichen Induktionsverfahren werden die Anteile  $a_{g^0}^{g^+}$ ,  $a_{g^0}^{g^-}$ ,  $a_{g^0}^{h^+}$  und  $a_{g^0}^{h^-}$  bestimmt. Danach können die gesuchten Anteile mit Hilfe der Formeln 4.48 und 4.49 bestimmt werden.

Initialisierung (zum Zeitpunkt  $t_0$ ):

$$a_{g^0}^{g^+}(t_0) = a_g^{g^+}(t_0), \quad (4.50)$$

$$a_{g^0}^{g^-}(t_0) = a_g^{g^-}(t_0), \quad (4.51)$$

$$a_{g^0}^{h^+}(t_0) = 0 \quad \text{und} \quad (4.52)$$

$$a_{g^0}^{h^-}(t_0) = 0. \quad (4.53)$$

Induktionsschritt:

Es wird der Anteil  $a_{g^0}^{g^0(t_{i+1})}$  der gescreenten Population bestimmt, der von  $t_i$  nach  $t_{i+1}$  überlebt hat.

$$\begin{aligned} a_{g^0}^{g^0(t_{i+1})} &= a_{g^0}^{g^+}(t_i) \cdot (p_{g^+g^+}(t_i) + p_{g^+h^+}(t_i)) \\ &\quad + a_{g^0}^{g^-}(t_i) \cdot (p_{g^-g^-}(t_i) + p_{g^-h^-}(t_i)) \\ &\quad + a_{g^0}^{h^+}(t_i) \cdot p_{h^+h^+}(t_i) \\ &\quad + a_{g^0}^{h^-}(t_i) \cdot p_{h^-h^-}(t_i). \end{aligned} \quad (4.54)$$

Dieser Anteil wird zwecks kürzerer Schreibweise in den folgenden Formeln als Nenner verwendet.

$$a_{g^0}^{g^+}(t_{i+1}) = \frac{a_{g^0}^{g^+}(t_i) \cdot p_{g^+g^+}(t_i)}{a_{g^0}^{g^0(t_{i+1})}}, \quad (4.55)$$

$$a_{g^0}^{g^-}(t_{i+1}) = \frac{a_{g^0}^{g^-}(t_i) \cdot p_{g^-g^-}(t_i)}{a_{g^0}^{g^0(t_{i+1})}}, \quad (4.56)$$

$$a_{g^0}^{h^+}(t_{i+1}) = \frac{a_{g^0}^{g^+}(t_i) \cdot p_{g^+h^+}(t_i) + a_{g^0}^{h^+}(t_i) \cdot p_{h^+h^+}(t_i)}{a_{g^0}^{g^0(t_{i+1})}} \quad \text{und} \quad (4.57)$$

$$a_{g^0}^{h^-}(t_{i+1}) = \frac{a_{g^0}^{g^-}(t_i) \cdot p_{g^-h^-}(t_i) + a_{g^0}^{h^-}(t_i) \cdot p_{h^-h^-}(t_i)}{a_{g^0}^{g^0(t_{i+1})}}. \quad (4.58)$$

## 4.8 Übergangswahrscheinlichkeiten bei Primärprävention

Es wird eine Primärprävention mit Statinen modelliert. Zusätzliche Statine erhalten nur Personen mit positivem Score, die noch nicht an der KHK erkrankt sind. Per-

sonen mit positivem Score, die bereits an der KHK erkrankt sind, bekommen keine *zusätzlichen* Statine verabreicht, da Statine derzeit bereits als Behandlungsmethode der KHK etabliert sind.

Um die Wirkungsweise der Statine zu modellieren, wird ein künstlicher Zustand  $g^{V+}$  gebildet. Dieser umfasst alle Personen, die im Szenario ohne Primärprävention bereits an der KHK erkrankt oder verstorben sind, bei denen im Szenario mit Primärprävention jedoch keine KHK vorliegt. Die Personen, die sich im Zustand  $g^+$  befinden, sind in den Szenarien mit und ohne Primärprävention identisch. In der Realität lässt sich nicht unterscheiden, ob eine Person sich im Zustand  $g^+$  oder  $g^{V+}$  befindet. Bei der Modellierung der Übergangswahrscheinlichkeiten ist die Trennung jedoch hilfreich.

Die „sonstige“ Mortalität wird in der Primärprävention durch die Statintherapie in der Modellierung nicht beeinflusst. Zwar senkt die Statintherapie in der Sekundärprävention signifikant die Wahrscheinlichkeit eines Schlaganfalls, in der Primärprävention konnten derartige Effekte jedoch nicht nachgewiesen werden (Vreccer *et al.* 2003). Bei Personen im Zustand  $g^{V+}$  ist eine höhere sonstige Mortalität zu erwarten als bei Personen in  $g^+$ . Das durch stärkeres Rauchen verursachte Lungenkrebsrisiko beispielsweise wird nicht durch Statine reduziert. Daher wird die (bedingte) sonstige Mortalität für  $g^{V+}$  vom Zustand  $h^+$  übernommen:

$$p_{g^{V+t}}|_{n. s. M.} = p_{h+t}|_{n. s. M.} \cdot \quad (4.59)$$

Herzinsuffizienz ist zum Teil als Folgeerkrankung auf die KHK zurückzuführen: Durch einen Herzinfarkt kann Herzmuskelgewebe absterben. Wird die Leistung des Herzens dadurch stark beeinflusst, so liegt eine Herzinsuffizienz vor. Es sei  $\delta_{HI}$  der Anteil, um den die Mortalität der Herzinsuffizienz bei Wegfall der KHK reduziert wird. Der Anteil  $\delta_{HI}$  wird in Abschnitt 5.6 alters- und geschlechtsspezifisch anhand von Prävalenzdaten geschätzt. Es wird die Annahme getroffen, dass er auf die Mor-

talität der Herzinsuffizienz übertragbar ist. Die Mortalität durch Herzinsuffizienz des Zustands  $g^{V+}$  wird vom Zustand  $h^+$  übernommen, aber um den Anteil  $\delta_{\text{HI}}$  reduziert:

$$p_{g^{V+i}}|_{\text{n. i. M.}} = (1 - \delta_{\text{HI}}) \cdot p_{h+i}|_{\text{n. i. M.}} \cdot \quad (4.60)$$

In der Meta-Analyse von Vreecer *et al.* (2003) wird der Effekt der Statine in der Primärprävention bezüglich koronarer Ereignisse zusammengefasst. Die Endpunkte der klinischen Studien entsprechen den Übergängen von den Zuständen  $g^+$  und  $g^{V+}$  nach  $h$  und  $k$ . Die Interventionsgruppe entspricht dem Personenkollektiv  $g^+ \cup g^{V+}$  unter Primärprävention, die Kontrollgruppe entspricht den Personen des Zustands  $g^+$  im Szenario ohne Primärprävention. Der Effekt der Statine wird als relatives Risiko  $\text{RR}^{\text{Präv}}$  quantifiziert. Es beziehe sich der Anteil  $a_{g^+ \cup g^{V+}}^{g^+}$  auf das Präventionsszenario (Im Szenario ohne Primärprävention befinden sich keine Personen in  $g^{V+}$ ; ein analog definierter Anteil wäre immer gleich 1). Die Übergangswahrscheinlichkeiten des Zustands  $g^+$  unter Primärprävention seien zwecks Unterscheidung zum Szenario ohne Prävention mit einem hochgestellten „Präv“ gekennzeichnet. Um mit der Modellierung möglichst nahe an der in den klinischen Studien gemessenen Größe zu bleiben, wird der Effekt durch die folgende Gleichung integriert:

$$\begin{aligned} & \text{RR}^{\text{Präv}} \cdot (p_{g^+h^+} + p_{g^+k})|_{\text{n. KHK-M.}} \\ &= a_{g^+ \cup g^{V+}}^{g^+} \cdot \left( p_{g^+h^+}^{\text{Präv}} + p_{g^+k}^{\text{Präv}} \right)|_{\text{n. KHK-M.}} \\ &+ \left( 1 - a_{g^+ \cup g^{V+}}^{g^+} \right) \cdot (p_{g^{V+h^+}} + p_{g^{V+k}})|_{\text{n. KHK-M.}} \cdot \quad (4.61) \end{aligned}$$

Formel 4.61 lässt sich folgendermaßen interpretieren: Die linke Seite stellt die um das relative Risiko reduzierte bedingte Erkrankungswahrscheinlichkeit dar. Diese wird mit der (gewichteten) gemeinsamen Erkrankungswahrscheinlichkeit der Zustände  $g^+$  und  $g^{V+}$  gleichgesetzt.

Aus den klinischen Studien geht nicht hervor, wie sich im Präventionsszenario die Erkrankungswahrscheinlichkeit zwischen den Personenkollektiven  $g^+$  und  $g^{V+}$  un-

terscheidet. Es ist ferner unmöglich, diesen Unterschied zu quantifizieren, da der Unterschied zwischen  $g^+$  und  $g^{V+}$  nicht beobachtbar ist. Der Einfachheit halber wird angenommen, dass die bedingten Erkrankungswahrscheinlichkeiten sowohl für fatale (tödliche Ereignisse, Übergänge nach  $k$ ) als auch nicht-fatale Ereignisse (Übergänge nach  $h$ ) identisch sind:

$$p_{g^{+h^+}}^{\text{Präv}} \Big|_{\text{n. KHK-M.}} = p_{g^{V+h^+}} \Big|_{\text{n. KHK-M.}} \quad \text{und} \quad (4.62)$$

$$p_{g^{+k}}^{\text{Präv}} \Big|_{\text{n. KHK-M.}} = p_{g^{V+k}} \Big|_{\text{n. KHK-M.}} . \quad (4.63)$$

Damit das Gleichungssystem lösbar ist, muss zusätzlich festgelegt werden, in welchem Verhältnis die fatalen zu den nicht-fatalen KHK-Ereignissen im Präventionsszenario stehen. Es wird angenommen, dass die Fatalitätsrate (der Anteil der tödlichen KHK-Ereignisse an allen KHK-Erstereignissen) im Präventionsszenario mit der Fatalitätsrate des Szenarios ohne Prävention übereinstimmt:

$$\frac{p_{g^{+k}}^{\text{Präv}}}{p_{g^{+h^+}}^{\text{Präv}} + p_{g^{+k}}^{\text{Präv}}} = \frac{p_{g^{+k}}}{p_{g^{+h^+}} + p_{g^{+k}}} . \quad (4.64)$$

Wegen

$$\begin{aligned} \left( p_{g^{+h^+}}^{\text{Präv}} + p_{g^{+k}}^{\text{Präv}} \right) \Big|_{\text{n. KHK-M.}} &= \frac{p_{g^{+h^+}}^{\text{Präv}} + p_{g^{+k}}^{\text{Präv}}}{1 - p_{g^{+i}}^{\text{Präv}} - p_{g^{+t}}^{\text{Präv}}} \\ &= p_{g^{+h^+}}^{\text{Präv}} \Big|_{\text{n. KHK-M.}} + p_{g^{+k}}^{\text{Präv}} \Big|_{\text{n. KHK-M.}} \end{aligned} \quad (4.65)$$

und wegen Gleichungen 4.62 und 4.63 gilt:

$$\left( p_{g^{+h^+}}^{\text{Präv}} + p_{g^{+k}}^{\text{Präv}} \right) \Big|_{\text{n. KHK-M.}} = \left( p_{g^{V+h^+}} + p_{g^{V+k}} \right) \Big|_{\text{n. KHK-M.}} , \quad (4.66)$$

so dass Gleichung 4.61 geschrieben werden kann als:

$$\text{RR}^{\text{Präv}} \cdot \left( p_{g^{+h^+}} + p_{g^{+k}} \right) \Big|_{\text{n. KHK-M.}} = \left( p_{g^{+h^+}}^{\text{Präv}} + p_{g^{+k}}^{\text{Präv}} \right) \Big|_{\text{n. KHK-M.}} . \quad (4.67)$$

Es ist demnach nicht nötig, den Anteil  $a_{g^+ \cup g^{V+}}^{g^+}$  zu bestimmen.

Da das Personenkollektiv im Zustand  $g^+$  in den Szenarien mit und ohne Prävention identisch ist, gilt:

$$p_{g^+g^+}^{\text{Präv}} = p_{g^+g^+} . \quad (4.68)$$

Die (bedingte) herzensuffizienzspezifische Mortalität sowie die (bedingte) sonstige Mortalität werden durch Primärprävention bei den Personen in  $g^+$  nicht verändert. Das heißt, es gilt:

$$p_{g^+t}^{\text{Präv}} \Big|_{\text{n. s. M.}} = p_{g^+t} \Big|_{\text{n. s. M.}} \quad \text{sowie} \quad (4.69)$$

$$p_{g^+i}^{\text{Präv}} \Big|_{\text{n. i. M.}} = p_{g^+i} \Big|_{\text{n. i. M.}} . \quad (4.70)$$

Somit fließen aus diesem Abschnitt die Gleichungen der Formeln 4.59, 4.60, 4.62 bis 4.64 und 4.67 bis 4.70 in das Gleichungssystem ein. Hinzu kommen die bedingten Wahrscheinlichkeiten (Formeln A.68 bis A.81) sowie die trivialen Bedingungen, dass die Übergangswahrscheinlichkeiten jedes Zustands sich zu 1 addieren und dass sich die Gesamtsterbewahrscheinlichkeit aus den ursachenspezifischen Sterbewahrscheinlichkeiten zusammen setzt (Formeln A.83 bis A.86). Insgesamt umfasst das Gleichungssystem 27 Gleichungen.

# Kapitel 5

## Quantifizierung der Input-Größen zur Bestimmung der Übergangswahrscheinlichkeiten

### 5.1 Verwendete Datensätze

Um die Input-Größen des Markovmodells zu quantifizieren, werden eine Vielzahl von Daten benötigt. Dazu gehören u. a. die Sterbetafeln des Statistischen Bundesamtes und viele Literaturquellen, auf die an den entsprechenden Stellen im Einzelnen verwiesen wird. Eine besondere Bedeutung haben die in diesem Abschnitt beschriebenen Datensätze, da sie an vielen Stellen für Berechnungen benötigt werden.

#### 5.1.1 Die Daten des Bundes-Gesundheitssurveys 1998

Das „Public Use File“ des Bundes-Gesundheitssurveys 1998 (BGS 98) ist ein für Deutschland repräsentativer Datensatz, der 1998 vom Robert-Koch-Institut (RKI) erhoben wurde. Er umfasst 7124 Personen und 582 Variablen. Befragt wurden Personen im Alter von 18 bis 79 Jahren. Neben der Befragung zu gesundheitsrelevanten Themen fand eine medizinische Untersuchung statt. Der Datensatz ist ausführlich

von Stolzenberg (2000, 2002) beschrieben. Zusätzlich zum „Public Use File“ hat das RKI im Rahmen eines Kooperationsvertrages vier weitere Variablen übermittelt: Den systolischen Blutdruck, das Gesamtcholesterin, das HDL-Cholesterin („High Density Lipoprotein“) sowie ob eine Erkrankung an Diabetes Mellitus vorliegt oder nicht.

### 5.1.2 Der SOEP-Datensatz

Bei den Sozio-ökonomischen Panel (SOEP) Daten des Deutschen Instituts für Wirtschaftsforschung (2006) handelt es sich um eine repräsentative Stichprobe mit jährlich mehr als 21000 Befragten. Befragt werden Personen, die das 16. Lebensjahr bereits vollendet haben. Es werden die Daten des Erhebungsjahres 2003 verwendet. Die herangezogenen Einkünfte beziehen sich auf das Jahr 2002, der Versichertenstatus auf das Jahr 2003. Die Einkommen werden differenziert abgefragt (Lohn, Einkommen aus selbständiger Tätigkeit, Renten jeder Art, Arbeitslosengeld/Arbeitslosenhilfe, Weihnachtsgeld etc.). Der Datensatz enthält Hochrechnungsfaktoren. Diese geben an, wieviele Personen in Deutschland durch die jeweilige befragte Person repräsentiert werden.

### 5.1.3 Der RSA-Datensatz von sechs großen Krankenkassen

Im Rahmen eines Gutachtens von Lauterbach *et al.* (2001) zum Risikostrukturausgleich (RSA) sind Daten von mehreren Krankenkassen erhoben worden. Ein Teil dieser Daten ist dazu geeignet, die Prävalenz der KHK zu quantifizieren, den Zusammenhang zwischen KHK und Herzinsuffizienz zu spezifizieren und die attributablen KHK-Ausgaben der Krankenkassen abzuschätzen. Es werden die Daten von sechs großen Krankenkassen (BARMER, DAK, KKH, AOK Hessen, AOK Sachsen-Anhalt, AOK Baden-Württemberg) verwendet, die jeweils aggregierte Daten über ihr gesamtes Versichertenkollektiv und personenspezifische Daten über chronisch kranke Versicherte bereitgestellt haben. In allen Datensätzen wird nach

Alter und Geschlecht stratifiziert, im personenspezifischen Datensatz ist zusätzlich dokumentiert, an welchen der betrachteten Krankheiten der jeweilige Versicherte erkrankt ist. Alle sechs Krankenkassen machen sowohl im aggregierten als auch im personenspezifischen Datensatz Angaben über die Versichertenzeiten in Tagen; bei vier der Krankenkassen liegen für beide Datensätze auch Ausgaben in drei Leistungsbereichen vor: Arzneimittel, Krankenhaus und Krankengeld. Die Angaben zu den Versichertenzeiten und zu den Kosten beziehen sich bei fünf Krankenkassen auf das Jahr 1999, bei einer Krankenkasse (nur Versichertenzeiten) auf das Jahr 2000.

Insgesamt liegen personenspezifische Angaben von 7,3 Mio. Versicherten vor. Diese beziehen sich auf alle Versicherten der teilnehmenden Krankenkassen, die mindestens eine von sieben chronischen Krankheiten haben: Hypertonus, Diabetes Mellitus, Asthma, Herzinsuffizienz, KHK, Brustkrebs und Schlaganfall. Die Personen mit den Krankheiten Hypertonus, Diabetes Mellitus, Asthma, Herzinsuffizienz und KHK wurden anhand von Medikationen identifiziert. Um z. B. als chronisch KHK-Kranker zu gelten, musste ein Patient im Bezugsjahr mindestens einmal Nitrate oder Molsidomin erhalten haben (Wendland 2002). Die Krankheiten Brustkrebs und Schlaganfall wurden anhand von Krankenhausentlassungsdiagnosen identifiziert. Die Identifizierung der chronisch Kranken erfolgte durch die Kassen.

Für weitere Analysen wurden die personenspezifischen Daten mit den aggregierten Daten zusammengeführt. Dabei wurden zwei Datensätze erstellt: Der erste Datensatz umfasst die Versichertenzeiten aller sechs Krankenkassen, enthält aber keine Kosten. Der zweite Datensatz enthält zwar die Kosten, bezieht sich aber auf die vier Krankenkassen, die vollständige Angaben zu den Kosten gemacht haben. Von den Versichertenzeiten und Kosten aller Versicherten einer Alters-Geschlechtsklasse wurden dabei die Versichertenzeiten bzw. Kosten aller chronisch Kranken derselben Alters-Geschlechtsklasse abgezogen. Somit wurden die Versichertenzeiten und Kosten aller Personen ohne eine der genannten chronischen Krankheiten bestimmt.

Für jede vorkommende Krankheitskombination einer Alters-Geschlechtsklasse wurden die Versichertenzeiten und ggf. die Kosten aggregiert. Die resultierenden Datensätze enthalten pro vorkommender Alters-/Geschlechts-/Krankheitskombination genau eine Zeile. Der Datensatz der vier Krankenkassen umfasst 9517 Zeilen und 5120 Mio. Versichertentage; der Datensatz der sechs Krankenkassen umfasst 9945 Zeilen und 9501 Mio. Versichertentage.

#### 5.1.4 Der RSA-Datensatz des Bundesversicherungsamtes (BVA)

Es werden in dieser Arbeit Risikostrukturausgleichsdaten (RSA-Daten) des Bundesversicherungsamtes (BVA) verwendet. Beim Risikostrukturausgleich handelt es sich um ein Transfersystem zwischen den Krankenkassen der Gesetzlichen Krankenversicherung (Lauterbach *et al.* 2001). Die verwendeten Daten beziehen sich auf den Jahresausgleich 2003 und sind auf der Homepage des BVA abrufbar. Für jede Kombination aus Versichertengruppe, Alter und Geschlecht werden Versichertenzeiten sowie Ausgaben für unterschiedliche Leistungsarten angegeben. Es handelt sich um aggregierte Daten.

## 5.2 Die Prävalenz der koronaren Herzkrankheit

Ein wichtiger Einflussfaktor bei der Schätzung der Übergangswahrscheinlichkeiten ist die Prävalenz der KHK. Für Deutschland gibt es bisher keine veröffentlichten Angaben über die alters- und geschlechtsspezifische Prävalenz. Es gibt zwei umfassende Datenquellen, anhand derer die Prävalenz in dieser Arbeit geschätzt wird: Erstens die Daten des Bundesgesundheits surveys 1998 (siehe Abschnitt 5.1.1) und zweitens die oben beschriebenen Krankenkassendaten (siehe Abschnitt 5.1.3).

### 5.2.1 Prävalenz der KHK aus dem Bundesgesundheits-survey

Die Teilnehmer wurden zu 43 Krankheiten befragt, ob sie die entsprechende Krankheit jemals gehabt haben. Die Antwortmöglichkeiten waren „Ja“, „Nein“ und „Weiß nicht“. Als Person mit KHK zählen in der folgenden Analyse diejenigen, die bei mindestens einer der Krankheiten\* „Durchblutungsstörungen am Herzen, Verengung der Herzkranzgefäße, Angina pectoris“ und „Herzinfarkt“ diese Frage mit „Ja“ beantwortet haben. Personen, die beide Fragen verneint haben zählen als Personen ohne KHK. Die übrigen Kombinationen zählen als fehlende Werte und werden für die Analyse ausgeschlossen. Insgesamt gibt es 484 Personen mit und 5903 Personen ohne KHK; 737 Antworten werden als fehlend eingestuft.

Es wird eine logistische Regression mit der Responsevariable KHK (ja/nein) und den Einflussvariablen Alter und Geschlecht durchgeführt. Gemäß den Empfehlungen von McCullagh und Nelder (1989) wird von Überstreuung („over-dispersion“) ausgegangen. Der Skalenparameter  $\phi$  wird mit 1,1514 geschätzt. Die Prävalenz wird alters- und geschlechtsspezifisch in Abbildung 5.1 dargestellt. Um die Plausibilität des Schätzers zu überprüfen, wird die beobachtete alters-/geschlechtsspezifische Prävalenz des BGS eingezeichnet. Der Schätzer erscheint plausibel. Alle im Modell vorkommenden Parameter haben signifikanten Einfluss, wie Tabelle 5.1 zu entnehmen ist. Außerdem sind in Tabelle 5.1 die Parameterschätzer mit zugehörigen Standardfehlern angegeben. Die Prävalenz, die sich aus der logistischen Regression ergibt, wird als  $\pi_{\text{BGS}}(\text{Alter}, \text{Geschlecht})$  bezeichnet.

Die Modellvoraussetzungen wurden mittels der vier Graphiken in Abbildung 5.2 überprüft. Da der Verlauf der einzelnen Kurven deutlich von dem üblichen Erschei-

---

\*Die in den Anführungszeichen angegebenen Krankheiten entsprechen den abgefragten Krankheiten des BGS 1998.

	Parameterschätzer	Std. Error	Teststatistik $z$	p-Wert
Achsenabschnitt	-7,2639	0,2824	-25,72	<0,001
Alter	0,0929	0,004584	20,27	<0,001
weibliches Geschlecht	-0,4993	0,1101	-4,535	<0,001

Tabelle 5.1: Parameterschätzer der logistischen Regression, um die Prävalenz der KHK in Deutschland zu bestimmen

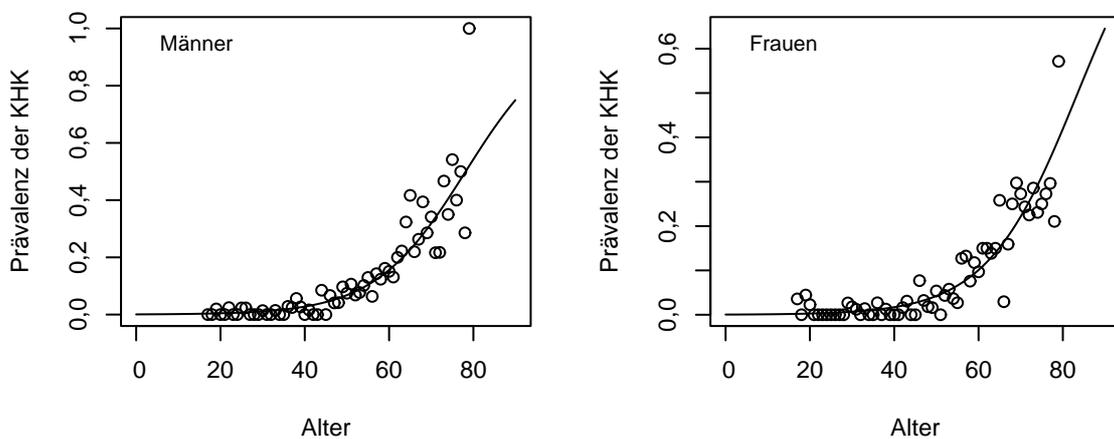


Abbildung 5.1: Prävalenz der KHK in Deutschland: Altersklassenmittel und durch logistischer Regression erhaltener Schätzer; Daten des BGS 98

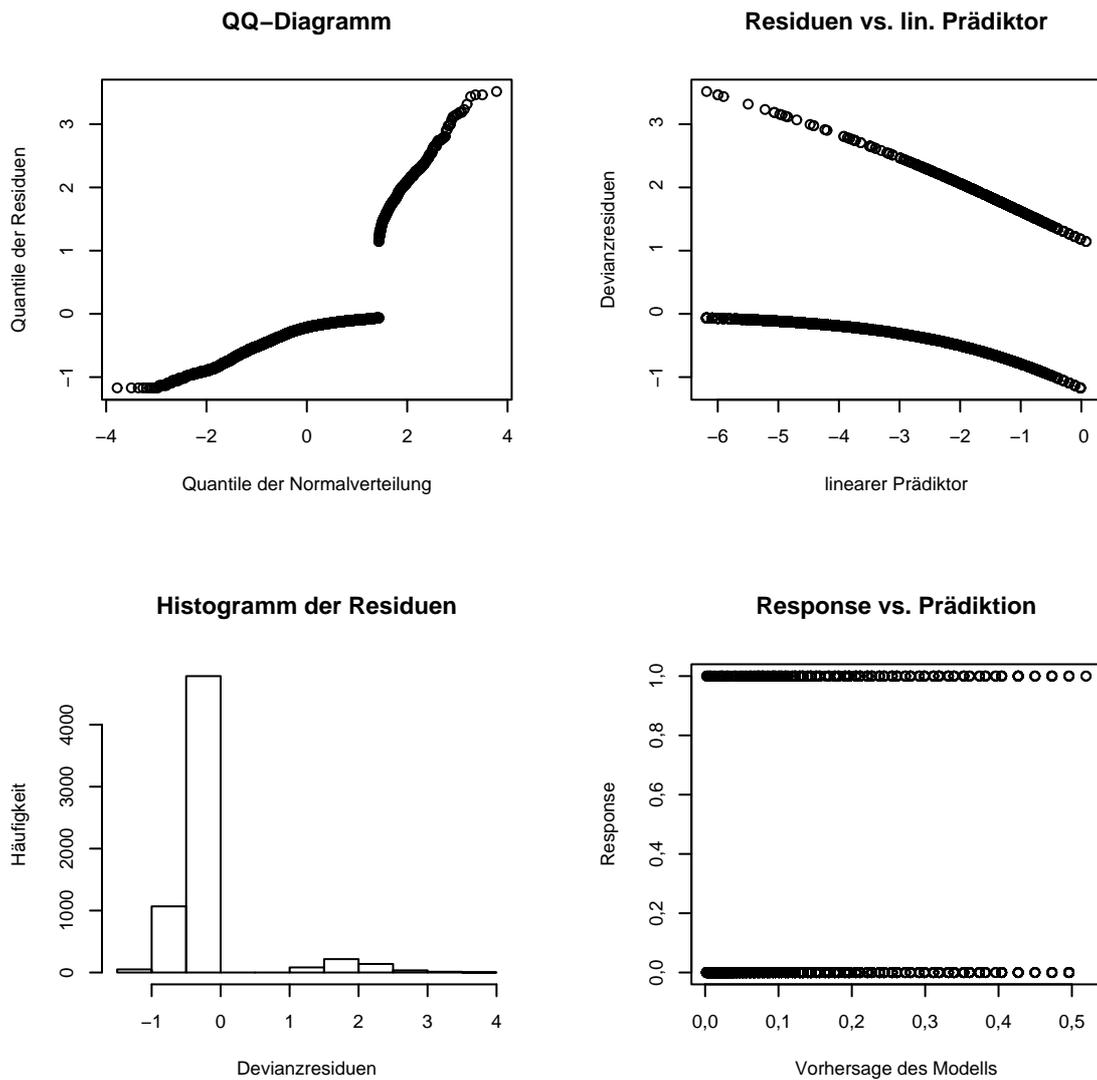


Abbildung 5.2: Graphische Überprüfung der Modellvoraussetzungen des Regressionsmodells der Prävalenz der KHK

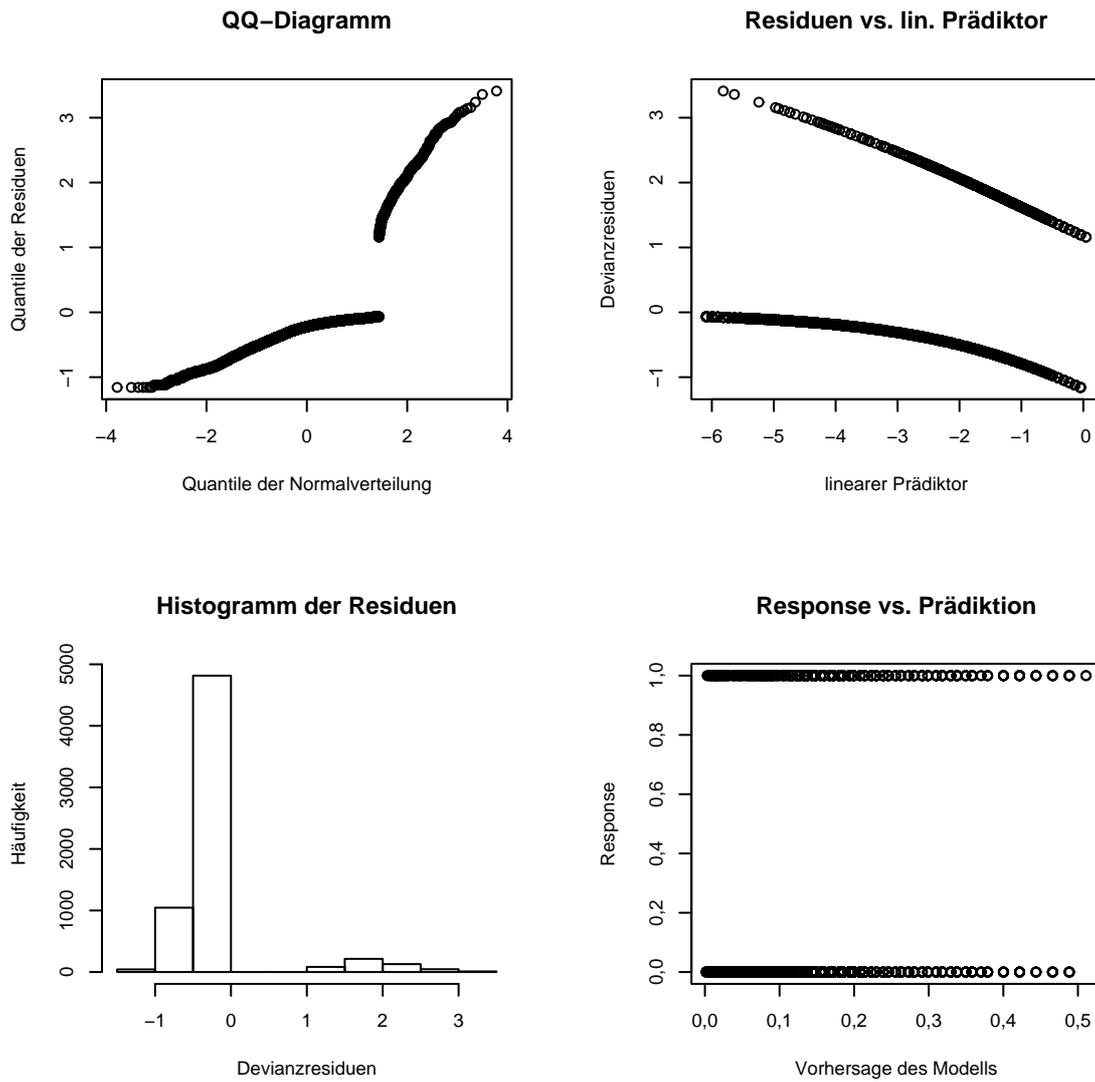


Abbildung 5.3: Simulierte KHK-Prävalenzen: Überprüfung der Modellvoraussetzungen

nungsbild bei Modellen mit normalverteilten Devianzresiduen abweicht, wurde für die gegebenen Kovariablen mit Hilfe der geschätzten Parameter die Responsevariable simuliert. Bei der Simulation wurde der Einfachheit halber nicht von einer Überstreuung ausgegangen. Für jede Person wurde mittels des Modells die Wahrscheinlichkeit bestimmt, an der KHK erkrankt zu sein. Zusätzlich wurde pro Person eine  $[0; 1]$ -rechteckverteilte Zufallsvariable bestimmt. Lag diese unter dem simulierten Erwartungswert, so galt die entsprechende Person als prävalent, d. h. als Person mit KHK. Für die so simulierten Daten wurde das Modell erneut geschätzt und Graphiken zur Überprüfung der Modellvoraussetzungen erstellt, siehe Abbildung 5.3. Diese unterscheiden sich nur unwesentlich vom tatsächlichen Modell. Daher können die Modellvoraussetzungen als gerechtfertigt angesehen werden.

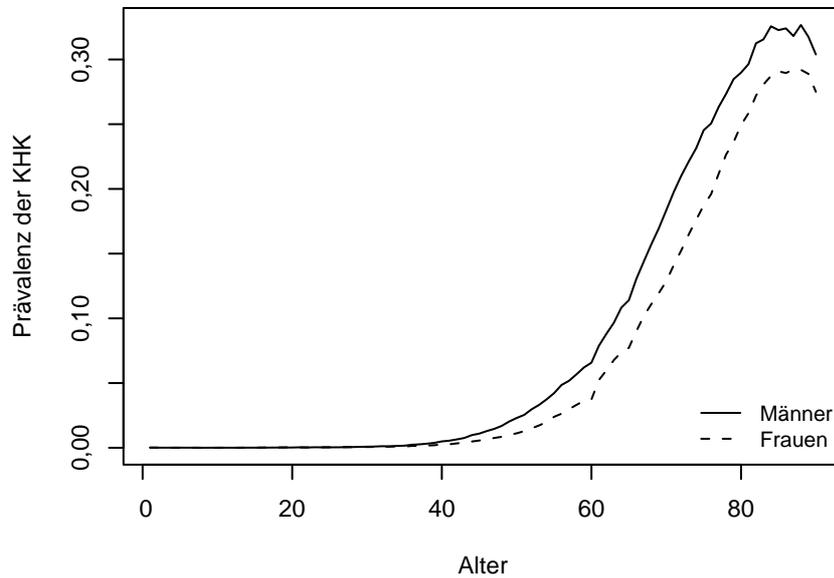


Abbildung 5.4: Prävalenz der KHK in Deutschland: Quotient aus Versichertenzeiten von Personen mit und ohne entsprechender Medikation; Kassendaten

### 5.2.2 Prävalenz der KHK aus Kassendaten

Die Prävalenz der KHK wird aus den Daten von sechs großen Krankenkassen mit der von Wendland (2002) beschriebenen Methode berechnet: Die Versichertenzeiten der Personen mit KHK werden kohortenspezifisch durch die Versichertenzeiten aller Personen der Kohorte geteilt. Als Kohorten werden Einjahresgruppen differenziert nach dem Geschlecht gewählt. Die Prävalenz aus den Kassendaten für das RSA-Gutachten wird als  $\pi_{\text{RSA}}(\text{Alter}, \text{Geschlecht})$  bezeichnet. Sie wird graphisch in Abbildung 5.4 veranschaulicht. Entsprechende Tabellen befinden sich in Anhang B.

### 5.2.3 Die Prävalenz der KHK im Markovmodell

In das Markovmodell fließt eine Gewichtung beider Prävalenzschätzer ein. Es sei  $\Omega \in [0; 1]$  das Gewicht. In der Basisanalyse wird  $\Omega = 0,5$  gewählt (siehe Abbildung

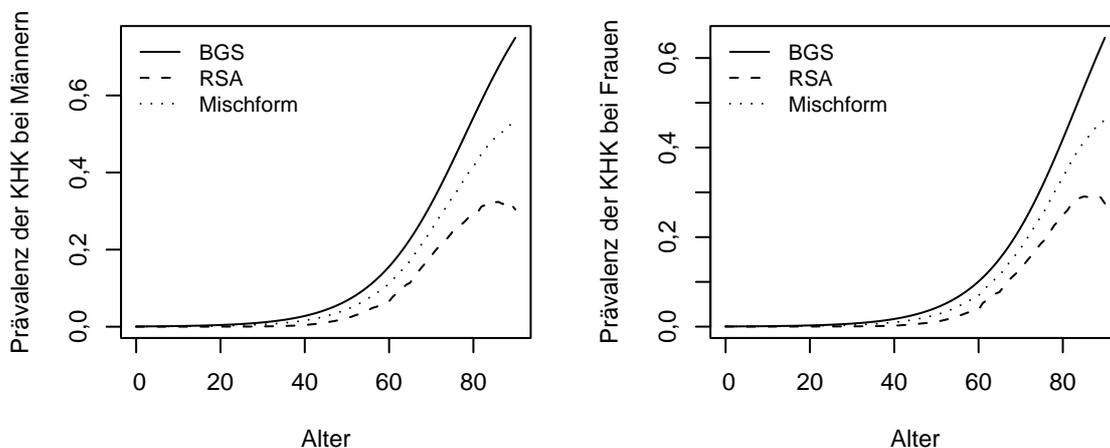


Abbildung 5.5: Prävalenz der KHK in Deutschland im Modell; Im Basisszenario wird die Mischform gewählt, in der beide Schätzer zu 50 % einfließen

5.5). In der Sensitivitätsanalyse wird  $\Omega$  im Intervall  $[0; 1]$  variiert.

$$\pi := \Omega \cdot \pi_{\text{BGS}} + (1 - \Omega) \cdot \pi_{\text{RSA}}. \quad (5.1)$$

### 5.3 Ursachenspezifische Mortalität

Die Mortalitätsraten werden aufgrund der Sterbetafeln des Statistischen Bundesamtes (2005) bestimmt. Die Anzahl der Todesfälle liegt nach Todesursachen getrennt vor. Die Todesursachen sind gemäß der „International Statistical Classification of Diseases“ Richtlinien (ICD-10) teilweise gruppiert klassifiziert. In der vorliegenden Analyse werden drei unterschiedliche Todesursachen unterschieden. Als KHK-Tod zählen die Codes I20-I25 (Ischämische Herzkrankheiten) sowie I46 (Herzstillstand). Als Herzinsuffizienz wird der Code I50 gewertet (Herzinsuffizienz). Alle übrigen Todesursachen fallen in die Kategorie „sonstiger Tod“.

Die Sterbefälle werden separat nach Alter und Geschlecht größtenteils in

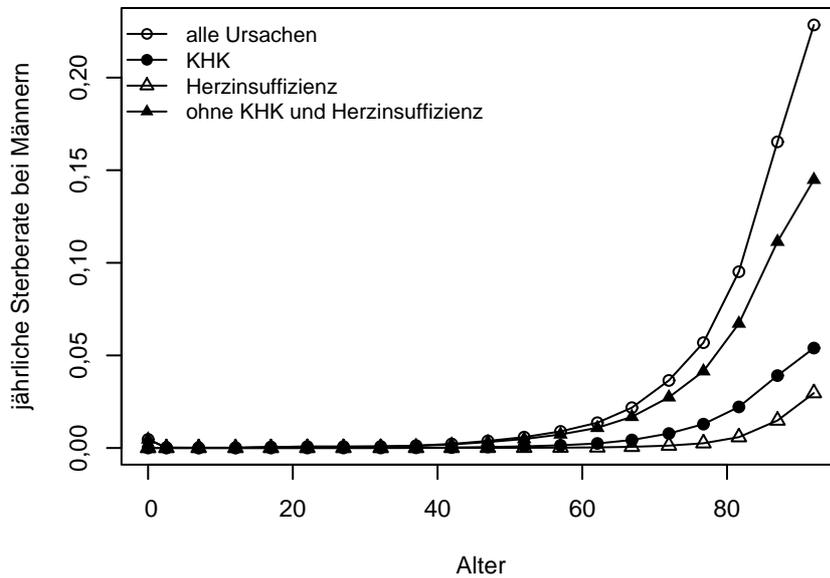


Abbildung 5.6: Ursachenspezifische jährliche Sterberaten für Männer

Fünffjahresgruppen angegeben (genauer: 0, 1-4, 5-9, 10-14, ..., 85-89, 90+). Die Bevölkerungsgröße liegt in der Regel für Einjahresgruppen vor (genauer: 0 bis 94, dann 95+ als eine Gruppe).

Für jede der Altersgruppen der Todesursache werden erstens die Personenzahl und zweitens das Durchschnittsalter bestimmt. Bei den über 95-Jährigen wird das Alter zwecks Durchschnittsbildung auf 96 Jahre gesetzt. Es werden gruppenspezifische Sterberaten berechnet.

Anschließend erfolgt eine lineare Interpolation, wobei die Sterberaten über dem zugehörigen Durchschnittsalter abgetragen werden. Die ursachenspezifischen Sterberaten werden beispielhaft für Männer in Abbildung 5.6 gezeigt.

## 5.4 Schätzung des Parameters $\lambda$

In diesem Abschnitt wird untersucht, wie stark die Aussagekraft des Screeningergebnisses im Laufe der Zeit abnimmt, d. h. es wird untersucht, wie der Parameter

$\lambda$  gewählt werden soll (vgl. Abschnitt 4.5 und Abbildung 4.1). Es ist denkbar, dass der Parameter  $\lambda$  bei unterschiedlichen Screeningmaßnahmen unterschiedliche Werte annimmt. Nach welchen Regeln und in welcher Größenordnung  $\lambda$  variiert werden müsste, ist nicht bekannt. Es wird die Annahme getroffen, dass  $\lambda$  bei allen Screeningmaßnahmen denselben Wert annimmt. Eine Variation von  $\lambda$  ist nicht erforderlich, um unterschiedliche Sensitivität-Spezifität-Kombinationen zu simulieren (siehe Abschnitt 7.2). Würde  $\lambda$  zusätzlich variiert werden, so wären Kosten und Nutzen nicht eindeutig in Abhängigkeit einer vorgegebenen Sensitivität und Spezifität des Screeningverfahrens bestimmbar.

Der Parameter  $\lambda$  wird mit Hilfe der Framingham-Risikogleichung nach Anderson *et al.* (1990, 1991) geschätzt, die auf die Population des BGS98-Datensatzes angewendet wird. Die Variable Linksventrikuläre Hypertrophie liegt in dem Datensatz nicht vor und wird gemäß Empfehlungen des Robert-Koch-Instituts auf 0 gesetzt, da sie „keinen wesentlichen Beitrag zur Risikoabschätzung leistet“. Die Anderson-Formel ist nur für Personen bis 74 Jahre geeignet. Es wird empfohlen einen Zeithorizont von 4 bis 12 Jahren zu wählen. Nur für diesen Bereich ist die Formel zulässig (Anderson *et al.* 1990). Für unterschiedliche Alters-/Geschlechts-Kohorten (nur Personen ohne KHK) werden die 4 bis 12-Jahres-Risiken berechnet. Die jeweilige Kohorte wird mittels eines Trennpunktes in Gruppen „+“ und „-“ unterteilt. Für jede Gruppe wird das durchschnittliche 10-Jahres-Risiko berechnet.

Aus den Mehr-Jahres-Risiken sollen Ein-Jahres-Risiken berechnet werden. Es sei  $p_i$  das Risiko im  $i$ -ten Jahr. Dann ist  $(1 - p_i)$  die Wahrscheinlichkeit, im  $i$ -ten Jahr gesund zu bleiben. Die Wahrscheinlichkeit, beispielsweise die ersten  $z$  Jahre gesund zu bleiben, beträgt

$$\prod_{i=1}^z (1 - p_i)$$

und das  $z$ -Jahres-Risiko beträgt

$$\text{JR}_z := \left\{ 1 - \prod_{i=1}^z (1 - p_i) \right\}. \quad (5.2)$$

Wegen

$$\frac{1 - \text{JR}_z}{1 - \text{JR}_{z-1}} = \frac{\prod_{i=1}^z (1 - p_i)}{\prod_{i=1}^{z-1} (1 - p_i)} = 1 - p_z \quad (5.3)$$

lässt sich das Erkrankungsrisiko im  $z$ -ten Jahr folgendermaßen berechnen:

$$p_z = 1 - \frac{1 - \text{JR}_z}{1 - \text{JR}_{z-1}}. \quad (5.4)$$

Für jede Person ohne KHK wird das Risiko bestimmt, im  $i$ -ten Jahr zu erkranken,  $i = 5, \dots, 12$ . Für jede Alters-/Geschlechtskohorte werden die Dezile des 10-Jahres-Risikos der entsprechenden Kohorte als Trennpunkt verwendet. Durch Festlegung des Trennpunktes entstehen zwei Gruppen mit positivem bzw. negativem Score. Für beide dieser Gruppen wird das durchschnittliche Risiko berechnet, im  $i$ -ten Jahr zu erkranken. Aus diesen Risiken wird das Odds Ratio zwischen den beiden Gruppen berechnet. Mittels nichtlinearer Regression (Venables und Ripley (2002) und Bates und Watts (1988)) werden für die Kohorte bei gegebenem Trennpunkt die Parameter  $\gamma_0$  und  $\lambda$  geschätzt. Als Startwert für  $\gamma_0$  wird das erste zur Verfügung stehende Odds Ratio gewählt, d. h. das Odds Ratio für das 5. Jahr. Für den Parameter  $\lambda$  wird der Startwert 0,96 gewählt. Für den Bereich, in dem die Risikoformel nach Anderson angewendet werden darf, kommt es augenscheinlich zu sehr guten Anpassungen. Ein Beispiel ist in Abbildung 5.7 dargestellt. Der Median aller geschätzten  $\lambda$  beträgt 0,9575368 und wird im Basisszenario der Analyse verwendet. Als Maximum wurde der Wert 0,9780753 und als Minimum der Wert 0,8865839 realisiert. Die Standardabweichung aller Schätzer für  $\lambda$  beträgt 0,008706904.

## 5.5 Mortalitätsspezifische Parameter

Es gibt im Wesentlichen fünf Parameter, durch die der Zusammenhang von KHK und erhöhter Mortalität im Modell beschrieben wird. Diese lauten  $\omega_k$ ,  $\omega_i$ ,  $\omega_t$ ,  $\psi_i^{\text{sup}}$  und

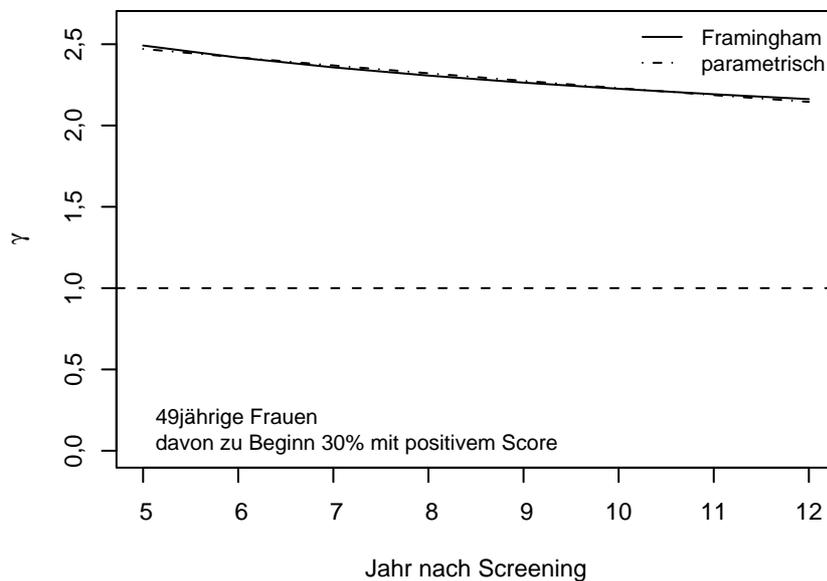


Abbildung 5.7: Der Verlauf der Funktion  $\gamma$  im Laufe der Zeit im Vergleich: Schätzer aus den Framingham Scores im Vergleich mit der parametrischen Variante in Abhängigkeit von  $\lambda$

$\psi_t^{\text{sup}}$  (vgl. Kapitel 4). Um sie zu quantifizieren, wurde im Rahmen einer Projektarbeit am Institut für Gesundheitsökonomie und Klinische Epidemiologie der Universität zu Köln (IGKE) von Herrn Markus Feger ein systematisches Review durchgeführt. Es sollten Studien identifiziert werden, in denen der Zusammenhang zwischen KHK und nicht-KHK-spezifischer Mortalität quantifiziert wird. Dabei sollte sowohl ein Vergleich zwischen Personen mit KHK und Personen ohne KHK sowie zwischen Personen mit hohem und mit niedrigem KHK-Erkrankungsrisiko berücksichtigt werden. Es wurden zwar keine Studien gefunden, die das Ziel hatten, die Fragestellung des systematischen Reviews zu beantworten, aber es wurden Studien identifiziert, aus denen Rückschlüsse auf den quantitativen Zusammenhang zwischen KHK und sonstiger Mortalität möglich sind. Insgesamt wurden acht Publikationen identifiziert: Daviglius *et al.* 2004, Kattainen *et al.* 2004, Lowe *et al.* 1998, Robbins *et al.* 2003, Rosengren *et al.* 1998, Rosengren *et al.* 2001, Stamler *et al.* 1999, Wannamethee

*et al.* 2004. Weiterhin sind die Studien dazu geeignet, den Zusammenhang zwischen der KHK und KHK-spezifischer Mortalität zu quantifizieren. Differenzierungen, in denen auf den Zusammenhang zwischen KHK und Mortalität durch Herzinsuffizienz geschlossen werden könnte, lag in den identifizierten Studien nicht vor.

### 5.5.1 Die Wahl von $\psi_t^{\text{sup}}$

Der Parameter  $\psi_t$  ist das in Abschnitt 4.4 eingeführte Odds Ratio, welches besagt, wie sich die bedingte sonstige Mortalität zwischen Personen mit positivem und negativem Score unterscheidet. Weiter hängt  $\psi_t$  funktional von dem Odds Ratio  $\gamma$  ab. Um die Funktion  $\psi_t$  eindeutig festzulegen, muss lediglich der Parameter  $\psi^{\text{sup}}$  spezifiziert werden. Aus den durch das systematische Review identifizierten Studien werden Realisationen bestimmt, mit denen Parameterkombinationen der Form  $(\gamma, \psi_t)$  approximiert werden.

Aus den folgenden Studien des systematischen Reviews können Parameterkombinationen der Form  $(\gamma, \psi_t)$  approximiert werden: Daviglus *et al.* (2004), Lowe *et al.* (1998), Rosengren *et al.* (2001), Stamler *et al.* (1999). In diesen Kohortenstudien werden jeweils zwei oder mehrere Gruppen mit unterschiedlichem KHK-Erkrankungsrisiko betrachtet. Werden mehrere Gruppen betrachtet, so sind unterschiedliche Aufteilungen in Personen mit positivem und negativem Score möglich. Jede dieser Aufteilungsmöglichkeiten wird im Folgenden betrachtet. Separat ausgewiesene Angaben zu Männern und Frauen werden nicht zusammengefasst. Um die Odds Ratios  $\gamma$  und  $\psi_t$  zu approximieren, werden Sterberaten bestimmt.

Dazu wird die Anzahl der KHK-Todesfälle und der sonstigen Todesfälle durch die Anzahl der beobachteten Personenjahre der entsprechenden Gruppe geteilt. Teilweise sind die Personenjahre angegeben, teilweise werden sie aus den Angaben „KHK-Tote pro Personenjahr“ und „Anzahl KHK-Todesfälle“ berechnet. Die

verwendeten Sterberaten in den Studien von Stamler *et al.* (1999) und Lowe *et al.* (1998) sind altersadjustiert. Dies ist hilfreich, da im Markovmodell dieser Arbeit sich diese Odds Ratios auf Kohorten gleichen Alters beziehen. Die Rate von Lowe *et al.* (1998) ist zusätzlich nach der ethnischen Herkunft adjustiert. Die Anzahl der an einem sonstigen Tod verstorbenen Personen im Beobachtungszeitraum ist in den Studien nicht dokumentiert und muss berechnet werden. Dazu wird von der Anzahl aller Todesfälle die Anzahl der KHK-bedingten Todesfälle abgezogen. Es sei angemerkt, dass im Markovmodell die Bezeichnung „sonstige Todesfälle“ den Tod durch Herzinsuffizienz ausschließt. Da die durch Herzinsuffizienz bedingten Todesfälle in den identifizierten Studien nicht separat ausgewiesen werden, kommt es an dieser Stelle zu Verzerrungen. Die zu  $\gamma$  zugehörigen  $\psi_t$  werden somit tendenziell leicht überschätzt.

Weiterhin ist anzumerken, dass die aus den Studien ermittelten Sterberaten keine Sterbewahrscheinlichkeiten sind. Denn es wäre theoretisch denkbar, dass die Anzahl der Todesfälle pro Personenjahr so groß wird, dass Raten größer 1 zustande kommen. Eine Berechnung von Odds Ratios anhand dieser Raten ist daher nicht sinnvoll. Statt dessen wird das relative Risiko, d. h. der Quotient dieser beiden Raten bestimmt. Relative Risiken können bei niedrigen Sterbewahrscheinlichkeiten approximativ als Odds Ratios verwendet werden (Schumacher und Schulgen 2002). Von dieser Eigenschaft wird hier Gebrauch gemacht<sup>†</sup>. Die geschätzten Odds Ratios werden verwendet, um Rückschlüsse auf den Parameter  $\psi_t^{\text{sup}}$ , durch den die Funktion  $\psi_t(\gamma)$  festgelegt wird, zu ziehen. In Tabelle 5.2 wird angegeben, wie die Personenaufteilung der Studien auf die Gruppen „+“ und „-“ des Markovmodells übertragen werden. In Tabelle 5.3 werden die Personenjahre und die Todesfälle, separat nach KHK-Tod und sonstigem Tod, aufgelistet. Die Schätzwerte, die sich aufgrund dessen für  $\gamma$  und für  $\psi_t$  ergeben, befinden sich in Tabelle 5.4.

---

<sup>†</sup>In der Praxis wird häufig die entgegengesetzte Vorgehensweise gewählt, d. h. relative Risiken werden durch Odds Ratios approximiert.

S1	Stamler, MRFIT	Männer, 35–39 J.	Risiko: hoch vs. gering
S2	Stamler, CHA	Männer, 18–39 J.	Risiko: hoch vs. gering
S3	Stamler, MRFIT	Männer, 40–57 J.	Risiko: hoch vs. gering
S4	Stamler, CHA	Männer, 40–59 J.	Risiko: hoch vs. gering
S5	Stamler, CHA	Frauen, 40–59 J.	Risiko: hoch vs. gering
R1	Rosengren		Risiko: hoch, mittel, leicht vs. optimal
R2	Rosengren		Risiko: hoch, mittel vs. leicht, optimal
R3	Rosengren		Risiko: hoch vs. mittel, leicht, optimal
D1	Daviglus		Risiko: hoch, mittel, gering vs. optimal
D2	Daviglus		Risiko: hoch, mittel vs. gering, optimal
D3	Daviglus		Risiko: hoch vs. mittel, gering, optimal
L1	Lowe	Frauen	Blutdruck: hoch, mittel, niedrig vs. optimal
L2	Lowe	Frauen	Blutdruck: hoch, mittel vs. niedrig, optimal
L3	Lowe	Frauen	Blutdruck: hoch vs. mittel, niedrig, optimal
L4	Lowe	Frauen	Cholesterin: hoch, mittel vs. niedrig
L5	Lowe	Frauen	Cholesterin: hoch vs. mittel, niedrig
L6	Lowe	Frauen	Rauchen: ja vs. nein
L7	Lowe	Männer	Blutdruck: hoch, mittel, niedrig vs. optimal
L8	Lowe	Männer	Blutdruck: hoch, mittel vs. niedrig, optimal
L9	Lowe	Männer	Blutdruck: hoch vs. mittel, niedrig, optimal
L10	Lowe	Männer	Cholesterin: hoch, mittel vs. niedrig
L11	Lowe	Männer	Cholesterin: hoch vs. mittel, niedrig
L12	Lowe	Männer	Rauchen: ja vs. nein

Tabelle 5.2: Beschreibung der Studienkürzel: Publikation, aus der Angaben entnommen wurden, ggf. Teilstudie, Einschränkungen der verwendeten Population, Definition von Personen mit positivem und negativem Score zwecks Übertragung auf das Markovmodell

	Personenjahre		Anzahl KHK-Tote		Anzahl sonstige Tote	
	„+“	„-“	„+“	„-“	„+“	„-“
S1	4900000	550000	735	11	1839	128
S2	213559	16667	126	1	353	19
S3	4813065	286364	9578	126	21456	722
S4	135433	6818	516	6	1168	30
S5	124828	5714	181	2	662	28
R1	14608	126995	149	528	228	1246
R2	120353	21250	646	31	1362	112
R3	139103	2500	676	1	1456	18
D1	170954	43168	45	2	373	49
D2	124918	89204	42	5	319	103
D3	30803	183319	19	28	91	331
L1	163399	15663	392	13	1080	92
L2	137779	41282	361	44	937	235
L3	105882	73179	324	81	759	413
L4	139328	43415	316	89	852	20
L5	65614	117129	187	218	400	472
L6	56032	130667	209	196	537	635
L7	187214	8974	1084	21	1973	72
L8	163076	33112	1007	98	1773	272
L9	129104	67084	865	240	1475	570
L10	124825	71202	791	314	1279	766
L11	41842	154184	318	787	434	1611
L12	70812	126742	541	564	1110	935

Tabelle 5.3: Personenjahre und Anzahl Todesfälle, unterteilt nach KHK-Tod und sonstigem Tod für unterschiedliche Definitionen von Personen mit positivem und negativem Score, gemäß der Einteilungen aus Tabelle 5.2

	KHK-Sterberate		Sterberate sonst. Tod		Schätzer für	
	„+“	„-“	„+“	„-“	$\gamma$	$\psi_t$
S1	0,15	0,02	0,38	0,23	7,50	1,61
S2	0,59	0,06	1,65	1,14	9,83	1,45
S3	1,99	0,44	4,46	2,52	4,52	1,77
S4	3,81	0,88	8,62	4,40	4,33	1,96
S5	1,45	0,35	5,30	4,90	4,14	1,08
R1	10,20	4,16	15,61	9,81	2,45	1,59
R2	5,37	1,46	11,32	5,27	3,68	2,15
R3	4,86	0,40	10,47	7,20	12,15	1,45
D1	0,26	0,05	2,18	1,14	5,68	1,92
D2	0,34	0,06	2,55	1,15	6,00	2,21
D3	0,62	0,15	2,95	1,81	4,04	1,64
L1	2,40	0,83	6,61	5,87	2,89	1,13
L2	2,62	1,07	6,80	5,69	2,46	1,19
L3	3,06	1,11	7,17	5,64	2,76	1,27
L4	2,27	2,05	6,12	0,46	1,11	13,27
L5	2,85	1,86	6,10	4,03	1,53	1,51
L6	3,73	1,50	9,58	4,86	2,49	1,97
L7	5,79	2,34	10,54	8,02	2,47	1,31
L8	6,18	2,96	10,87	8,21	2,09	1,32
L9	6,70	3,58	11,42	8,50	1,87	1,34
L10	6,34	4,41	10,25	10,76	1,44	0,95
L11	7,60	5,10	10,37	10,45	1,49	0,99
L12	7,64	4,45	15,68	7,38	1,72	2,12

Tabelle 5.4: Sterberaten (Todesfälle pro 1000 Personenjahren) bezüglich der Ursache KHK bzw. sonstiger Todesursachen sowie daraus gewonnene Schätzer für die Parameter  $\gamma$  und  $\psi_t$

	Schätzer	Standardfehler
$\psi_t^{\text{sup}}$	2,056	0,171

Tabelle 5.5: Durch nichtlineare Regression ermittelter Schätzwert für  $\psi_t^{\text{sup}}$  sowie Standardfehler

Ziel dieses Abschnitts ist es nicht, unterschiedliche Parameterkombinationen von  $\gamma$  und  $\psi_t$  zu bestimmen, sondern die Funktion zu schätzen, die  $\psi_t$  in Abhängigkeit von  $\gamma$  beschreibt. Dazu muss der Parameter  $\psi_t^{\text{sup}}$  geschätzt werden, durch den die Funktion festgelegt wird (vgl. Abschnitt 4.6). Der Parameter  $\psi_t^{\text{sup}}$  gibt das höchstmögliche Odds Ratio sonstiger Todesursachen zwischen Personen mit positivem und negativem Screeningergebnis an, das sich ergibt, wenn das Odds Ratio  $\gamma$  der Todesursache KHK gegen unendlich strebt.

Um  $\psi_t^{\text{sup}}$  zu schätzen, wird eine nichtlineare Regression (vgl. Venables und Ripley (2002) und Bates und Watts (1988)) durchgeführt. Dazu werden die Wertepaare  $(\gamma, \psi_t)$  aus Tabelle 5.4 gegeneinander abgetragen (siehe Abbildung 5.8, links oben).

Es wird der Zusammenhang

$$\tilde{\psi}_t = \psi_t(\tilde{\gamma}, \psi_t^{\text{sup}}) + \varepsilon \quad (5.5)$$

unterstellt, wobei  $\tilde{\psi}_t$  und  $\tilde{\gamma}$  die Realisationen von  $\psi_t$  und  $\gamma$  gemäß Tabelle 5.4 bezeichnen und  $\varepsilon$  ein normalverteilter Fehlerterm mit Erwartungswert 0 ist. Als Startwert für  $\psi_t^{\text{sup}}$  wird das Maximum der in die nichtlineare Regression einfließenden  $\tilde{\psi}_t$  gewählt. Es werden vier Ausreißer entfernt, da sie den funktionalen Verlauf von  $\psi_t(\gamma)$  zu stark verzerren würden (siehe Abbildung 5.8, links oben). Die angepasste Funktion ist in Abbildung 5.9 graphisch dargestellt. Die Modellvoraussetzung, d. h. die Annahme, dass der Fehler normalverteilt ist, scheint gegeben zu sein (siehe Abbildung 5.8). Der realisierte Schätzer  $\hat{\psi}_t^{\text{sup}}$  für  $\psi_t^{\text{sup}}$  ist in Tabelle 5.5 dargestellt.

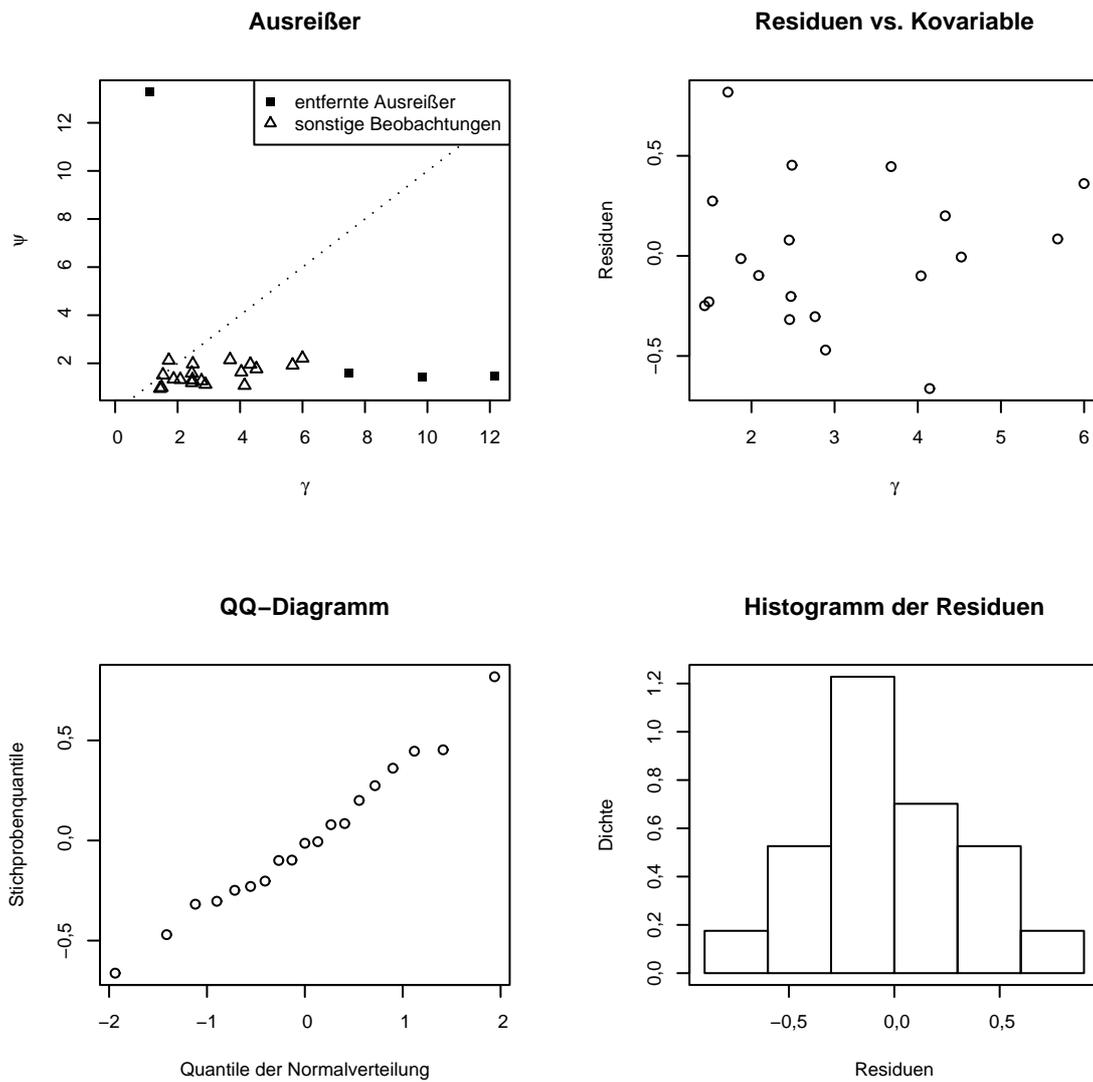


Abbildung 5.8: Residualanalyse des nichtlinearen Regressionsmodells; Ausreißer (oben links) flossen nicht in die nichtlineare Regression mit ein

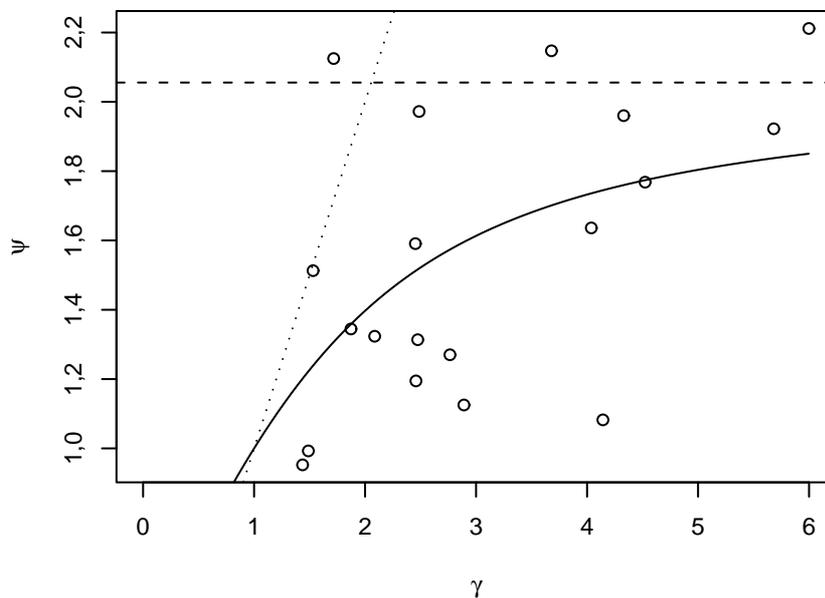


Abbildung 5.9: Die durch nichtlineare Regression angepasste Funktion  $\psi(\gamma)$ ; Für  $\psi^{\text{sup}}$  ergab sich ein Schätzwert von 2,056

### 5.5.2 Die Wahl von $\psi_i^{\text{sup}}$

Die Wahl des Parameters  $\psi_i^{\text{sup}}$  erweist sich als schwieriger als die Wahl des Parameters  $\psi_t^{\text{sup}}$  der sonstigen Mortalität, da keine entsprechenden Literaturquellen vorliegen. Die Mortalität durch Herzinsuffizienz ist stärker mit der KHK-Mortalität verknüpft als die sonstige Mortalität. Daher sollte für  $\psi_i^{\text{sup}}$  ein größerer Wert als  $\psi_t^{\text{sup}}$  gewählt werden. Für die Basisanalyse wird  $\psi_i^{\text{sup}}$  doppelt so groß wie  $\psi_t^{\text{sup}}$  gewählt. Eine Variation erfolgt im Rahmen der Sensitivitätsanalyse (Kapitel 9).

### 5.5.3 Die Wahl von $\omega_k$ und $\omega_t$

Die Parameter  $\omega_k$  und  $\omega_t$  sind Odds Ratios, welche die bedingten Sterbewahrscheinlichkeiten eines koronaren bzw. sonstigen Todes zwischen Personen mit und ohne KHK miteinander vergleichen. Zwei der im systematischen Review identifizierten Studien sind prinzipiell dazu geeignet, Rückschlüsse auf die Parameter  $\omega_k$  und  $\omega_t$

zu ziehen (Wannamethee *et al.* 2004 und Rosengren *et al.* 1998). Die Studie von Rosengren *et al.* (1998) hat gegenüber der Studie von Wannamethee *et al.* (2004) den entscheidenden Vorteil, dass die Sterbefälle und Sterberaten (Sterbefälle pro Personenjahr) für einen kürzeren Beobachtungszeitraum separat ausgewiesen sind. So können statt der gesamten Follow-up Zeit auch lediglich die ersten vier Jahre des Follow-up verwendet werden. Es ist besser, die Parameter  $\omega_k$  und  $\omega_t$  aus den Daten dieses kürzeren Zeitraums zu schätzen, da sich im Markovmodell die Odds Ratios auf einen noch kürzeren Zeitraum beziehen. Und zwar beziehen sich im Markovmodell diese Odds Ratios auf das erste Jahr, nachdem der Gesundheitsstatus „KHK liegt vor“/„liegt nicht vor“ festgelegt wurde. Im darauf folgenden Jahr/Zyklus findet im Markovmodell die Zuordnung einer Person auf die Zustände  $g$  und  $h$  erneut statt. Optimal wäre es demnach, wenn Sterbefälle und Personenjahre für ein Ein-Jahres-Follow-up angegeben wären. Da die Daten des Vier-Jahres-Follow-ups aus der Studie von Rosengren *et al.* (1998) am geeignetsten sind, werden diese zur Schätzung von  $\omega_k$  und  $\omega_t$  verwendet.

Personenjahre werden in der Studie nicht explizit angegeben. Sie können aber, wie in Abschnitt 5.5.1 bereits geschehen, aus der Anzahl der KHK-Todesfälle und KHK-Toten pro Personenjahr berechnet werden. Während die Anzahl der KHK-Todesfälle für die Gruppen „Männer ohne KHK“, „Männer mit Angina Pectoris“ und „Männer, die bereits einen Herzinfarkt hatten“ angegeben ist, muss die Anzahl der sonstigen Todesfälle für die drei Gruppen berechnet werden. Dazu wird von der Gesamtzahl der Todesfälle die Anzahl der KHK-Todesfälle abgezogen. Todesfälle durch Herzinsuffizienz sind nicht separat aufgeführt, wodurch es zu einer Überschätzung des Parameters  $\omega_t$  kommt. Diese wird toleriert, da der Anteil der Todesfälle durch Herzinsuffizienz verhältnismäßig gering ist (Statistisches Bundesamt 2005). Die beiden Personengruppen mit Angina Pectoris und mit Herzinfarkt werden zur Gruppe der Personen mit KHK zusammengefasst (siehe Tabelle 5.6). Frauen wurden nicht in die Studie eingeschlossen. Da keine vergleichbaren Angaben

	Personenjahre	Anzahl KHK-Todesfälle	Anzahl sonstige Todesfälle
ohne KHK	28571,43	60	176
Angina Pectoris	1351,35	15	9
Herzinfarkt	780,14	33	5
KHK	2131,49	48	14

Tabelle 5.6: Aus der Studie von Rosengren *et al.* (1998) berechnete Personenjahre und Anzahl ursachenspezifischer Todesfälle

	KHK-Sterberate	Sterberate sonstiger Todesursachen
ohne KHK	2,10	6,16
Angina Pectoris	11,10	6,66
Herzinfarkt	42,30	6,41
KHK	22,52	6,57

Tabelle 5.7: Sterberaten: Todesfälle pro 1.000 Personenjahren

für Frauen vorliegen, werden die Ergebnisse für Männer auf Frauen übertragen.

Wie in Abschnitt 5.5.1 können Sterberaten für Personen mit und ohne KHK berechnet werden. In Tabelle 5.7 sind die daraus resultierenden Sterberaten dargestellt. Es seien  $T_0^{\text{KHK}}$  die Anzahl der Personenjahre bei Personen mit KHK und  $T_0^{\text{ohne KHK}}$  die Anzahl der Personenjahre bei Personen ohne KHK. Weiter werde die Anzahl der KHK Todesfälle als  $K_k^{\text{ohne KHK}}$  bzw.  $K_k^{\text{KHK}}$  und die Anzahl der sonstigen Todesfälle als  $K_t^{\text{ohne KHK}}$  bzw.  $K_t^{\text{KHK}}$  bezeichnet. Daraus werden Verhältniswerte (relative Risiken) berechnet, die wiederum als Schätzer für die Odds Ratios verwendet werden. Es ergeben sich folgende Schätzwerte:

$$\hat{\omega}_k := \frac{K_k^{\text{ohne KHK}}}{K_k^{\text{KHK}}} \cdot \frac{T_0^{\text{KHK}}}{T_0^{\text{ohne KHK}}} \approx 0,093 \quad \text{und} \quad (5.6)$$

$$\hat{\omega}_t := \frac{K_t^{\text{ohne KHK}}}{K_t^{\text{KHK}}} \cdot \frac{T_0^{\text{KHK}}}{T_0^{\text{ohne KHK}}} \approx 0,938. \quad (5.7)$$

#### 5.5.4 Die Wahl von $\omega_i$

Bei dem Parameter  $\omega_i$  handelt es sich um ein Odds Ratio, das in Abschnitt 4.2 eingeführt wurde. Es bezieht sich auf die Wahrscheinlichkeit, dass eine Person an der Herzinsuffizienz stirbt, gegeben dass zuvor kein Tod einer anderen Ursache (KHK-Tod oder sonstiger Tod) eintritt. Bei den Personengruppen, die miteinander verglichen werden, handelt es sich um Personen mit bzw. ohne KHK, adjustiert nach Alter und Geschlecht. Geeignete Daten zur Mortalität durch die Herzinsuffizienz liegen nicht vor. Stattdessen wird ein Prävalenzdatensatz der Herzinsuffizienz verwendet. Das aus diesem Datensatz gewonnene Odds Ratio wird auf die Mortalität übertragen.

Als Prävalenzdatensatz wird der Krankenkassendatensatz aus Abschnitt 5.1.3 (alle sechs Krankenkassen, zusammengefügt Datensatz aus aggregierten und personenspezifischen Daten) verwendet. Zu jeder Kombination aus Alter, Geschlecht und KHK (ja/nein) wird der Anteil der Versichertentage von Personen mit Herzinsuffizienz an den Versichertentagen aller Personen bestimmt. Dieser Anteil wird als Responsevariable in einem GAM verwendet. Als unabhängige Variablen fließen das Geschlecht, der Status der KHK, ein glatter Term des Alters sowie ein glatter Term des Produkts aus dem Alter und der Indikatorvariable Geschlecht ein. Die logit-Funktion wird als Linkfunktion gewählt, und die Versichertentage fließen als GewichtungsvARIABLE in die Regression ein. Als Basis für die glatten Terme werden „thin plate regression splines“ (vgl. Abschnitt 2.1.6) gewählt. Die Glättungsparameter werden mittels GCV (vgl. Abschnitt 2.1.8) bestimmt. Die Varianzfunktion (vgl. Formel 2.4 in Abschnitt 2.1.1)  $V(\mu) := \mu \cdot (1 - \mu)$  wird von der Binomialfamilie übernommen. Die geschätzten Parameter des Modells werden in Tabelle 5.8, die glatten Terme in Abbildung 5.10 dargestellt. Die Modellanpassung ist sehr gut, das adjustierte

Parametrische Koeffizienten:				
	Schätzer	Standardfehler	t-Statistik	p-Wert
(Achsenabschnitt)	-3,363	0,152	-22,08	<0,001
weiblich	0,070	0,297	0,24	0,81
KHK	1,438	0,021	69,55	<0,001
Approximative Signifikanz der glatten Terme:				
	gesch. Freiheitsgrade	F-Statistik	p-Wert	
Alter	8,918	1101,97	<0,001	
Alter · weiblich	5,351	5,82	<0,001	
adjustiertes R-Quadrat	0,997	GCV-Score	28967	
erklärte Devianz	99,6 %	Skalenschätzer	27578	

Tabelle 5.8: Modellzusammenfassung des GAM zur Quantifizierung von  $\omega_i$

R-Quadrat beträgt 0,997. Das Odds Ratio zwischen Personen mit und ohne KHK, ob Herzinsuffizienz vorliegt, wird berechnet, indem der Parameter der KHK mit der Exponentialfunktion transformiert wird (Kleinbaum 1994, S. 76ff). Da  $\omega_i$  als Referenz die Personen mit KHK und nicht die Personen ohne KHK verwendet, wird zusätzlich der Kehrwert gebildet. Es ergibt sich:

$$\hat{\omega}_i := \frac{1}{\exp(1,438)} \approx 0,237 . \quad (5.8)$$

Die Modellvoraussetzungen des GAM werden mit Hilfe der Abbildung 5.11 überprüft. Die Verteilung der Devianzresiduen weicht zwar leicht von der Normalverteilung ab, die Abweichung ist aber akzeptabel. Bei Werten des linearen Prädiktors im Bereich von -4 bis 0 scheinen die Residuen stärker zu streuen. Dieser Bereich entspricht geschätzten Wahrscheinlichkeiten zwischen 0,02 und 0,5. Aus der „Response vs. Prädiktion“ ist zu erkennen, weshalb die Residuen in diesem Bereich stärker streuen: Zwei Kurven weichen systematisch von der Winkelhalbierenden ab. Unter den Personen mit KHK kommt es in einigen Alters-Geschlechtsklassen vor,

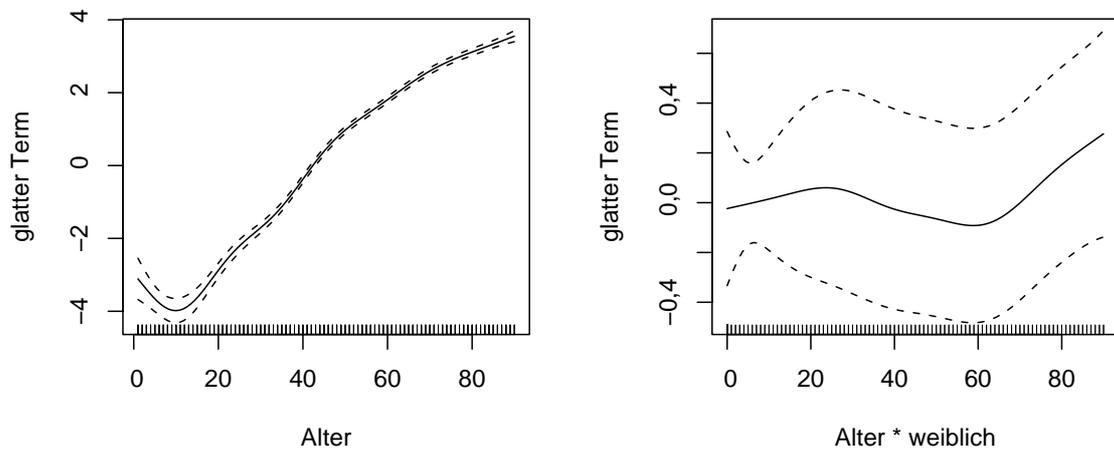


Abbildung 5.10: Einfluss des Alters auf den linearen Prädiktor des GAMs zur Quantifizierung des Parameters  $\omega_i$

dass alle Personen an mindestens einer der chronischen Krankheiten Bluthochdruck, Diabetes und Schlaganfall leiden. Da diese chronischen Krankheiten wiederum mit einem höheren Herzinsuffizienz-Risiko verbunden sind, kommt es zu der beobachtbaren Abweichung. Die obere abweichende Kurve steht für Frauen, die untere für Männer. Abweichungen bei Vorhersagewerten nahe 0 kommen hauptsächlich durch geringe Fallzahlen zustande. Da im Modell nicht vorgesehen ist, dass das Odds Ratio  $\omega_i$  bezüglich weiterer chronischer Erkrankungen adjustiert wird, ist es legitim  $\omega_i$  anhand des vorliegenden GAM zu schätzen.

## 5.6 KHK-bedingte Herzinsuffizienz

Die KHK gilt als wichtigste Ursache der Herzinsuffizienz. Hoppe *et al.* (2005) schreiben: „Die häufigste Ursache einer Herzinsuffizienz in westlichen Ländern ist die koronare Herzerkrankung (54-70%), die bei 35-52% dieser Patienten von einer arteriellen Hypertonie begleitet ist“. Somit wird bei einer Vermeidung der KHK durch

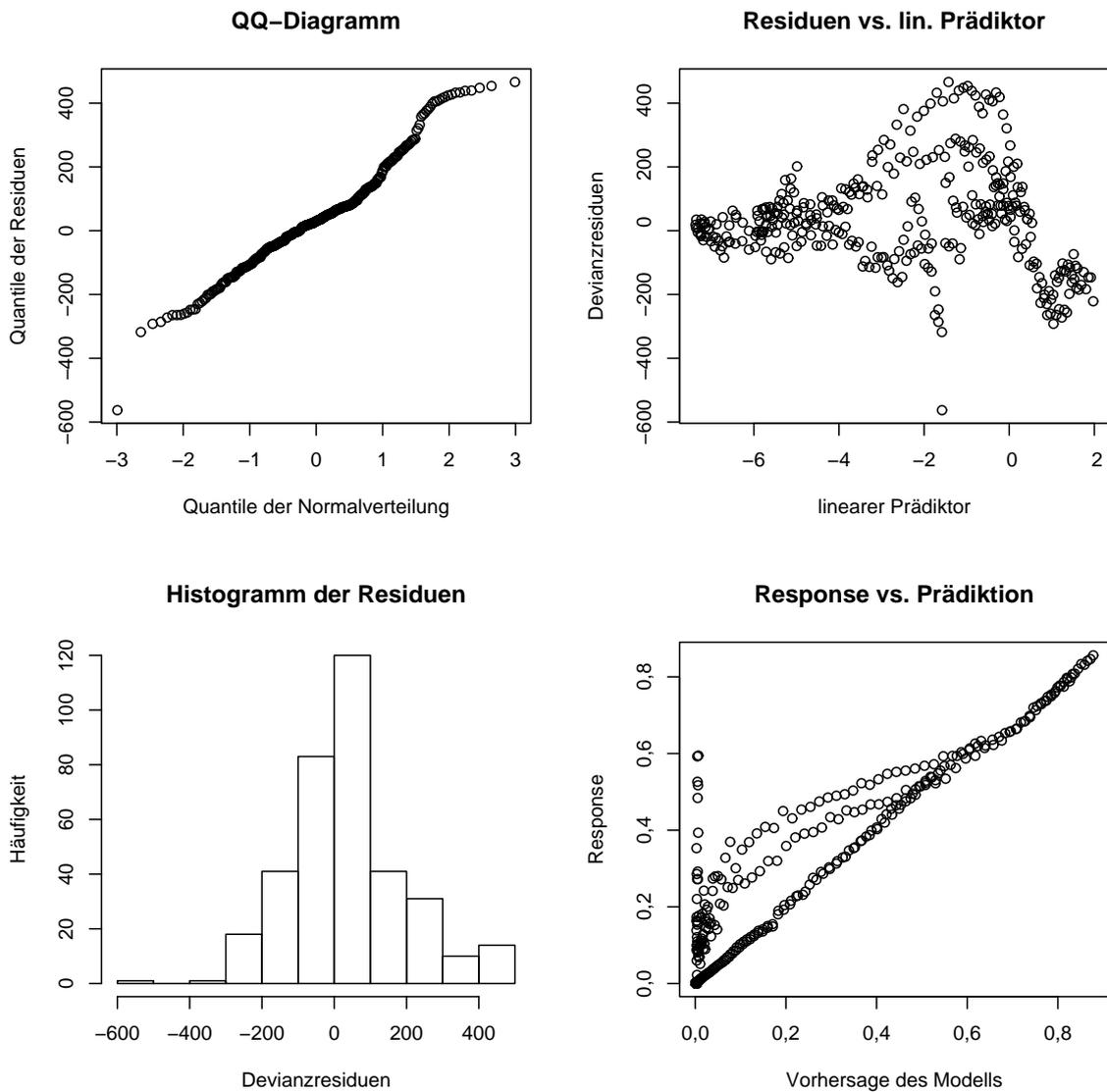


Abbildung 5.11: Graphiken zur Überprüfung der Modellvoraussetzungen des GAM, mit dem  $\omega_i$  geschätzt wird

Primärprävention auch ein Teil der Herzinsuffizienz-spezifischen Mortalität vermieden. Der Anteil, um den sich bei Wegfall der KHK die Mortalität durch Herzinsuffizienz verringern würde, wird mit  $\delta_{\text{HI}}$  bezeichnet und fließt in das Gleichungssystem aus Abschnitt 4.8 ein. Die Angaben von Hoppe *et al.* (2005) gehen auf McMurray *et al.* (2000) zurück und sind nicht dazu geeignet,  $\delta_{\text{HI}}$  zu schätzen. So stellen McMurray *et al.* (2000) beispielsweise fest, wie schwierig es ist, die primäre Ursache von Herzinsuffizienz festzulegen, wenn bei einem Patienten viele potentielle Ursachen vorliegen. Da die bisher veröffentlichten Angaben zur Approximierung von  $\delta_{\text{HI}}$  nicht gut geeignet sind, werden eigene Berechnungen durchgeführt. Da kein Datensatz, der sich auf die Mortalität der Herzinsuffizienz bezieht, zur Verfügung steht, wird wie in Abschnitt 5.5.4 bei der Quantifizierung von  $\omega_i$  ein Prävalenzdatensatz verwendet. Es wird angenommen, dass die aus diesem Datensatz gewonnenen Aussagen auf die Mortalität der Herzinsuffizienz übertragbar ist.

### 5.6.1 Modellierung der Prävalenz der Herzinsuffizienz

Es wird ein GAM angepasst, mit dem die Prävalenz der Herzinsuffizienz in Abhängigkeit von Alter, Geschlecht und Komorbiditäten modelliert wird. Als Datengrundlage wird der in Abschnitt 5.1.3 beschriebene Krankenkassendatensatz aller sechs Krankenkassen verwendet. Pro Kombination aus Alter, Geschlecht und den fünf Komorbiditäten (Hypertonie, KHK, Brustkrebs, Schlaganfall und Diabetes) wird der Anteil der Versichertentage von Personen mit Herzinsuffizienz an allen Versichertentagen bestimmt. Dieser Anteil fließt als Responsevariable in die Modellgleichung ein. Folgende Kovariablen werden verwendet: Ein glatter Term des Alters, die Indikatorvariablen der Komorbiditäten sowie die Indikatorvariable Geschlecht. Weiterhin fließen in das Modell ein: Die Wechselwirkungen jeweils zweier Komorbiditäten sowie glatte Terme des Produktes des Alters mit jeweils einer der Komorbiditäten bzw. dem Geschlecht. Als Basis der glatten Terme werden „thin plate regression splines“ verwendet, deren Glättungsparameter

mittels GCV bestimmt werden. Als Linkfunktion wird die logistische Funktion gewählt. Die Varianzfunktion wird von der Binomialverteilung übernommen, d. h.  $V(\mu) = \mu \cdot (1 - \mu)$ . Als GewichtungsvARIABLE fließen die Versichertentage in das Modell ein. Aus Konvergenzgründen<sup>‡</sup> werden nur Kombinationen der Kovariablen verwendet, in denen der Anteil der Personen mit Herzinsuffizienz ungleich 0 oder gleich 1 ist. Dadurch werden 0,1 % der Versichertentage nicht berücksichtigt. Dabei sind Versichertentage von Personen mit bzw. ohne KHK gleichermaßen betroffen (0,08 % bei Personen mit und 0,12 % bei Personen ohne KHK).

Durch das Modell wird 99,9 % der Devianz erklärt. Das adjustierte R-Quadrat beträgt 1 (siehe Tabelle 5.11), womit die Modellgüte extrem hoch ist. Die Parameterschätzer mit zugehörigen Statistiken sind in Tabelle 5.9 dargestellt, die Signifikanz der glatten Terme werden in Tabelle 5.10 dargestellt. In Abbildung 5.12 werden die glatten Terme des GAM graphisch veranschaulicht. Die Modellvoraussetzungen werden mit Hilfe von Abbildung 5.13 überprüft. Die Normalverteilungsannahme der Residuen ist approximativ erfüllt. Die Graphik „Residuen vs. linearer Prädiktor“ zeigt keinen stark abschüssigen oder ansteigenden Verlauf. Die Graphik „Response vs. Prädiktion“ zeigt den typischen Verlauf eines gut angepassten Modells. Die Modellvoraussetzungen können somit als gegeben angesehen werden.

---

<sup>‡</sup>In der iterativen Maximierung der Quasi-Likelihood kommt es bei Modellen mit logistischer Linkfunktion häufig zu Problemen, wenn Responsewerte außerhalb des offenen Intervalls (0; 1) auftreten. Auch durch die Festlegung von Startwerten kommt es zu keiner Konvergenz des Modells, bevor die entsprechenden Zeilen ausgeschlossen sind.

	Schätzer	Standardfehler	t-Statistik	p-Wert
(Achsenabschnitt)	-7,66	0,78	-9,85	<0,001
weiblich	0,03	0,10	0,28	0,78
Diabetes	-1,96	1,21	-1,61	0,11
KHK	1,94	0,20	9,55	<0,001
Hypertonie	8,89	0,18	48,48	<0,001
Asthma	0,60	0,07	8,12	<0,001
Brustkrebs	3,33	0,75	4,41	<0,001
Schlaganfall	5,15	1,79	2,87	0,004
Diabetes & KHK	0,06	0,01	4,05	<0,001
Diabetes & Hypertonie	-0,40	0,02	-25,55	<0,001
Diabetes & Asthma	-0,07	0,02	-3,82	<0,001
Diabetes & Brustkrebs	-0,03	0,08	-0,43	0,67
Diabetes & Schlaganfall	-0,13	0,03	-3,89	<0,001
KHK & Hypertonie	-0,67	0,01	-47,11	<0,001
KHK & Asthma	0,02	0,01	1,05	0,30
KHK & Brustkrebs	-0,03	0,07	-0,45	0,65
KHK & Schlaganfall	-0,10	0,03	-3,35	<0,001
Hypertonie & Asthma	-0,35	0,02	-22,72	<0,001
Hypertonie & Brustkrebs	-0,27	0,07	-4,05	<0,001
Hypertonie & Schlaganfall	0,13	0,03	4,01	<0,001
Asthma & Brustkrebs	-0,12	0,08	-1,56	0,12
Asthma & Schlaganfall	-0,12	0,04	-2,97	<0,001
Brustkrebs & Schlaganfall	-0,42	0,17	-2,49	0,01

Tabelle 5.9: Schätzer, Standardfehler und Teststatistiken der nicht glatten Terme des GAM, in dem die Prävalenz der Herzinsuffizienz modelliert wird

	gesch. Freiheitsgrade	F	p-Wert
Alter	9,0	8124,4	<0,001
Alter · Diabetes	9,0	21,4	<0,001
Alter · KHK	4,4	7,5	<0,001
Alter · Hypertonie	8,9	4157,2	<0,001
Alter · Asthma	3,6	2,6	0,005
Alter · Brustkrebs	4,2	3,1	0,001
Alter · Schlaganfall	7,0	11,2	<0,001
Alter · weiblich	6,6	165,9	<0,001

Tabelle 5.10: Approximative Signifikanzen der glatten Terme des GAM in dem die Prävalenz der Herzinsuffizienz modelliert wird

adjustiertes R-Quadrat	1	GCV-Score	1126
erklärte Devianz	99,9 %	Skalenschätzer	1105

Tabelle 5.11: Generelle Statistiken zum GAM, in dem die Prävalenz der Herzinsuffizienz modelliert wird

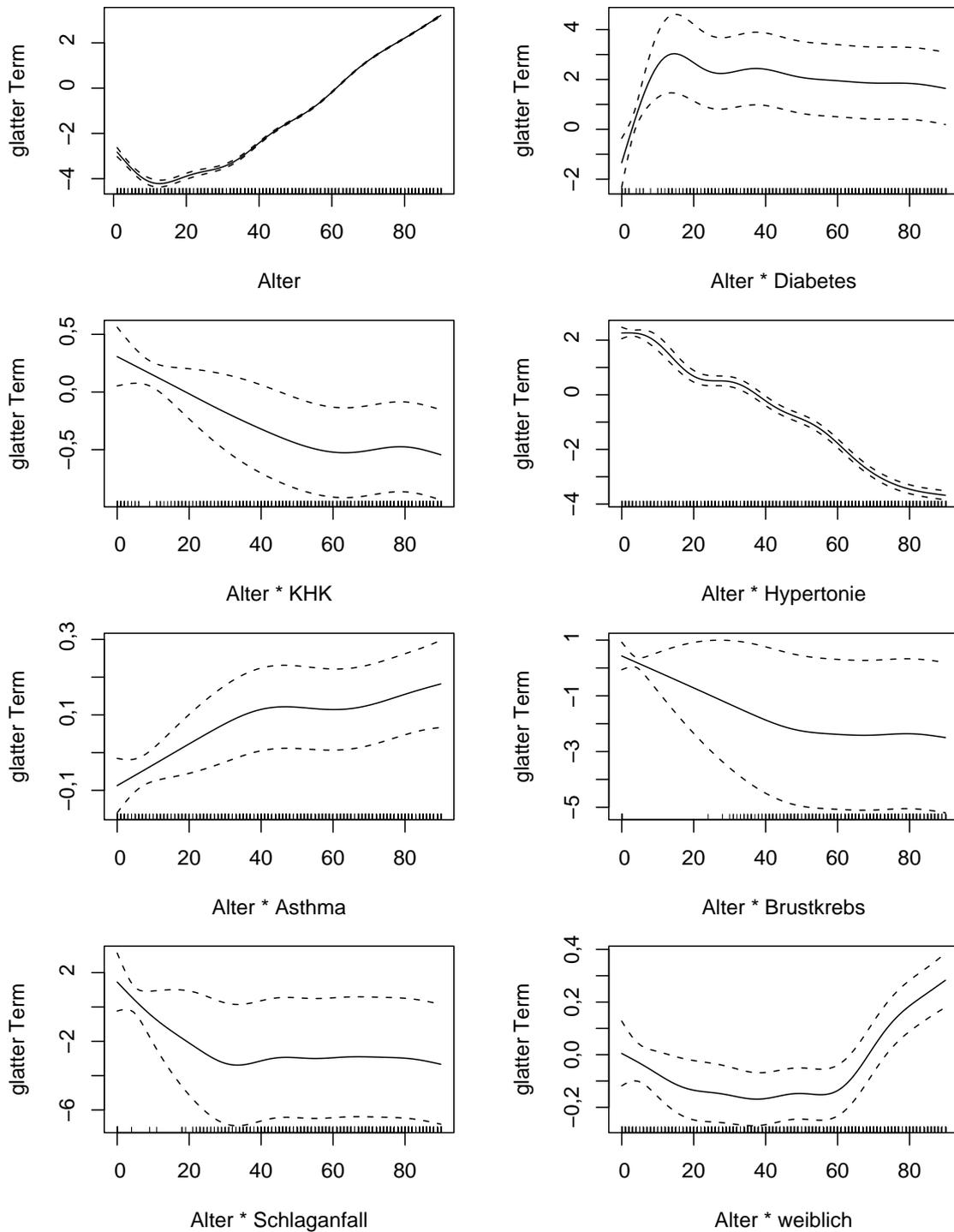


Abbildung 5.12: Glatte Terme und deren punktweise approximativen 95%-Konfidenzintervalle des GAM zur Modellierung der Prävalenz der Herzinsuffizienz

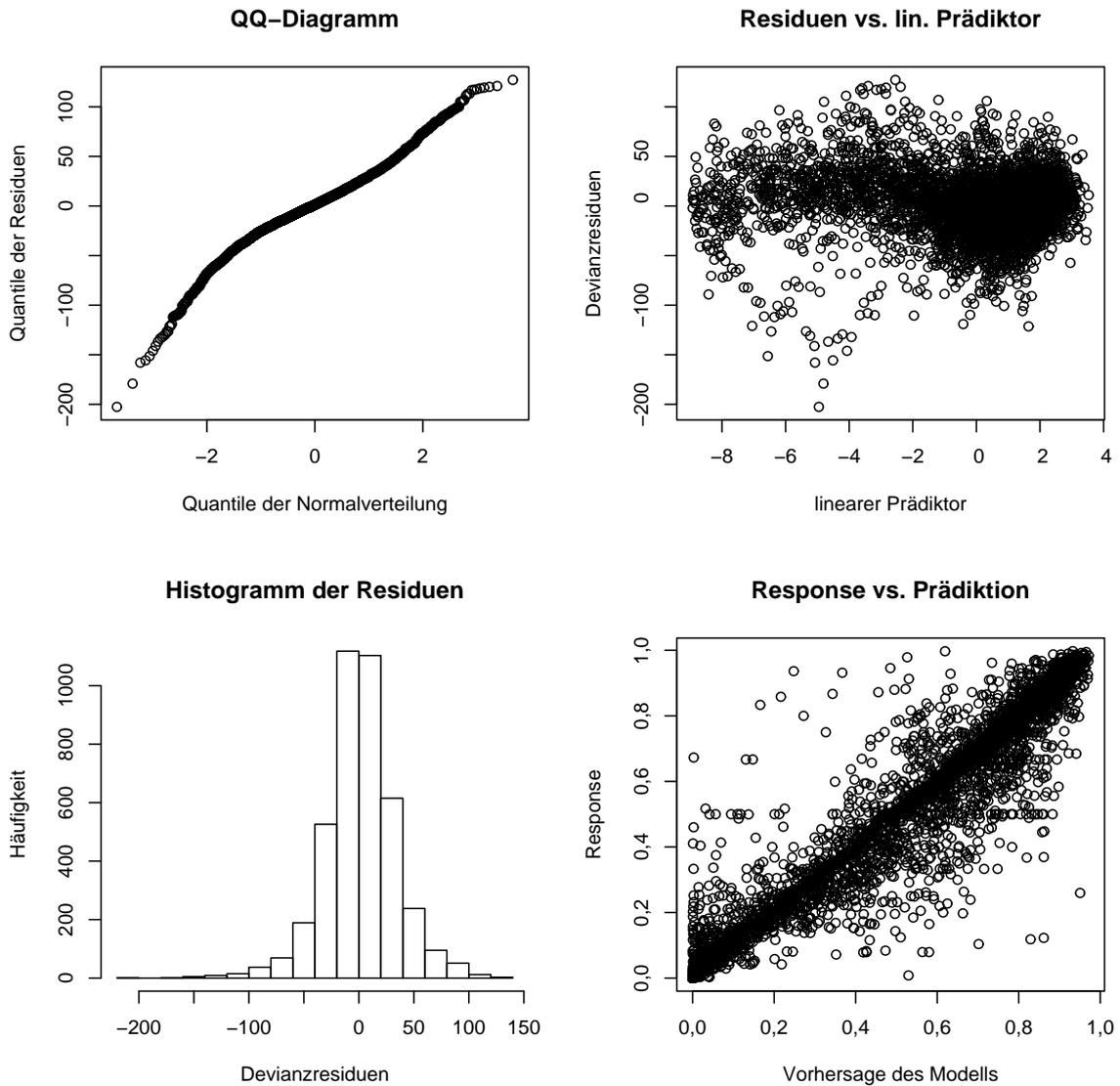


Abbildung 5.13: Graphiken zur Überprüfung der Modellvoraussetzungen des GAM, mit dem die Prävalenz der Herzinsuffizienz in Abhängigkeit von Alter, Geschlecht und Komorbidität inklusive einiger Wechselwirkungen, modelliert wird

### 5.6.2 Quantifizierung vermeidbarer Fälle von Herzinsuffizienz

Es sei  $\mathcal{P}^{\mathcal{K}}$  die Prävalenz der Herzinsuffizienz bei Personen mit KHK. Weiter sei  $\mathcal{P}^{\bar{\mathcal{K}}}$  die Prävalenz der Herzinsuffizienz bei Personen ohne KHK. Kann bei Personen mit KHK im Szenario ohne Prävention die KHK durch die Prävention vermieden werden, so sinkt bei ihnen die Prävalenz der Herzinsuffizienz ab. Bei ihnen wird jedoch nicht  $\mathcal{P}^{\bar{\mathcal{K}}}$  erreicht, da weitere Einflussfaktoren der Prävalenz der Herzinsuffizienz bestehen bleiben. Der Wert, auf den  $\mathcal{P}^{\mathcal{K}}$  sinkt, wenn von den Personen mit KHK das Alter, das Geschlecht und die Komorbidität<sup>§</sup> unverändert bleiben, wird mit  $\mathcal{P}^{\mathcal{A}_1}$  bezeichnet. Der Anteil  $\delta_{\text{HI}}$ , der angibt, wie stark die Prävalenz der Herzinsuffizienz bei Personen mit KHK sinken würde, wenn die koronare Herzkrankheit wegfallen würde, lässt sich dann folgendermaßen ausdrücken:

$$\delta_{\text{HI}} = 1 - \frac{\mathcal{P}^{\mathcal{A}_1}}{\mathcal{P}^{\mathcal{K}}} . \quad (5.9)$$

Um  $\delta_{\text{HI}}$  zu schätzen, werden mit Hilfe des in Abschnitt 5.6.1 angepassten GAM und mit Hilfe der Krankenkassendaten die zugrundeliegenden Prävalenzen  $\mathcal{P}^{\mathcal{K}}$  und  $\mathcal{P}^{\mathcal{A}_1}$  geschätzt. Die Quantifizierung von  $\delta_{\text{HI}}$  erfolgt separat für jede Alters- und Geschlechtsgruppe.

Es sei  $\mathcal{K}$  der Teildatensatz der Krankenkassendaten, der genau die Zeilen einer Alters-Geschlechtsklasse umfasst, die sich auf Personen mit KHK beziehen. Die Anzahl der Zeilen von  $\mathcal{K}$  wird mit  $N_{\mathcal{K}}$  bezeichnet. Es sei  $\mathcal{T}_i^{\mathcal{K}}$ ,  $i = 1, \dots, N_{\mathcal{K}}$ , die  $i$ -te Komponente der Versichertentage des Datensatzes  $\mathcal{K}$ . Weiter sei

$$\mathcal{T}^{\mathcal{K}} := (\mathcal{T}_1^{\mathcal{K}}, \dots, \mathcal{T}_{N_{\mathcal{K}}}^{\mathcal{K}})^{\top} \quad (5.10)$$

---

<sup>§</sup>Da es sich bei der Herzinsuffizienz um die Responsevariable handelt und der Status der KHK verändert wird, bezieht sich die Komorbidität in diesem Zusammenhang auf die vier chronischen Krankheiten Hypertonie, Diabetes, Schlaganfall und Brustkrebs.

ein Vektor der einzelnen Versichertentage der Personen mit KHK. Es wird ein Datensatz  $\mathcal{A}_1$  erstellt, dessen einziger Unterschied zu  $\mathcal{K}$  darin besteht, dass der Status KHK auf „nein“ gesetzt wird. Es sei  $\mathcal{P}_i^{\mathcal{K}}$  die erwartete Prävalenz der Herzinsuffizienz von Personen, deren Kovariablen der  $i$ -ten Zeile von  $\mathcal{K}$  entsprechen. In Analogie dazu beziehe sich  $\mathcal{P}_i^{\mathcal{A}_1}$  auf Personen der  $i$ -ten Zeile von  $\mathcal{A}_1$ . Wird das GAM aus Abschnitt 5.6.1 verwendet, um  $\mathcal{P}_i^{\mathcal{K}}$  bzw.  $\mathcal{P}_i^{\mathcal{A}_1}$  zu schätzen, so ergeben sich die Schätzwerte  $\hat{\mathcal{P}}_i^{\mathcal{K}}$  bzw.  $\hat{\mathcal{P}}_i^{\mathcal{A}_1}$ . Mit Hilfe dieser Größen lassen sich  $\mathcal{P}^{\mathcal{K}}$  und  $\mathcal{P}^{\mathcal{A}_1}$  folgendermaßen schätzen:

$$\hat{\mathcal{P}}^{\mathcal{K}} = \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{K}}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}}}, \quad (5.11)$$

$$\hat{\mathcal{P}}^{\mathcal{A}_1} = \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{A}_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}}}. \quad (5.12)$$

Werden die Schätzer der Gleichungen 5.11 und 5.12 in Gleichung 5.9 eingesetzt, so lässt sich ein Schätzer für  $\delta_{\text{HI}}$  konstruieren:

$$\begin{aligned} \hat{\delta}_{\text{HI}} &:= 1 - \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{A}_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}}} \cdot \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{K}}} \\ &= 1 - \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{A}_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{K}}}. \end{aligned} \quad (5.13)$$

Mittels des Fehlerfortpflanzungsgesetzes (vgl. Hartung *et al.* 2005) soll die Varianz des Schätzers  $\hat{\delta}_{\text{HI}}$  approximiert werden. Dazu muss  $\hat{\delta}_{\text{HI}}$  nach  $\mathcal{T}_j^{\mathcal{K}}$ ,  $\mathcal{P}_j^{\mathcal{K}}$  und  $\mathcal{P}_j^{\mathcal{A}_1}$ ,  $j = 1, \dots, N_{\mathcal{K}}$ , abgeleitet werden:

$$\frac{\partial \hat{\delta}_{\text{HI}}}{\partial \mathcal{T}_j^{\mathcal{K}}} = - \frac{\hat{\mathcal{P}}_j^{\mathcal{A}_1} \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{K}} - \mathcal{P}_j^{\mathcal{K}} \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{A}_1}}{\left( \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{K}} \right)^2}, \quad (5.14)$$

$$\frac{\partial \hat{\delta}_{\text{HI}}}{\partial \hat{\mathcal{P}}_j^{\mathcal{K}}} = \frac{\mathcal{T}_j^{\mathcal{K}} \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{A}_1}}{\left( \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{K}} \right)^2}, \quad (5.15)$$

$$\frac{\partial \hat{\delta}_{\text{HI}}}{\partial \mathcal{P}_j^{\mathcal{A}_1}} = - \frac{\mathcal{T}_j^{\mathcal{K}}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{K}}}. \quad (5.16)$$

Für die logistische Funktion:

$$\text{logit}(x) := \log \left( \frac{x}{1-x} \right), \quad \forall x \in (0; 1) \quad (5.17)$$

lautet die Umkehrfunktion:

$$\text{logit}^{-1}(x) := \frac{\exp(x)}{1 + \exp(x)}, \quad \forall x \in \mathbb{R}. \quad (5.18)$$

Für einige Berechnungen wird die Ableitung der Inversen der logistischen Funktion benötigt. Diese lässt sich wie folgt berechnen:

$$\frac{\partial \text{logit}^{-1}(x)}{\partial x} = \frac{\exp(x)}{(1 + \exp(x))^2}. \quad (5.19)$$

Die Definition der logistischen Funktion sowie ihrer Umkehrfunktion wird auf Vektoren erweitert. Bei Vektoren erfolgt eine komponentenweise Anwendung der logistischen Funktion bzw. ihrer Umkehrfunktion:

$$\text{logit}(x) := (\text{logit}(x_1), \dots, \text{logit}(x_n))^{\top} \quad \text{für } x \in (0; 1)^n \quad \text{und} \quad (5.20)$$

$$\text{logit}^{-1}(x) := (\text{logit}^{-1}(x_1), \dots, \text{logit}^{-1}(x_n))^{\top} \quad \text{für } x \in \mathbb{R}^n. \quad (5.21)$$

Es wird ein Vektor definiert, der alle geschätzten Prävalenzen enthält, die zur Quantifizierung von  $\delta_{\text{HI}}$  benötigt werden:

$$\hat{\mathcal{P}}^{\delta} := \left( \hat{\mathcal{P}}_1^{\mathcal{K}}, \dots, \hat{\mathcal{P}}_{N_{\mathcal{K}}}^{\mathcal{K}}, \hat{\mathcal{P}}_1^{A_1}, \dots, \hat{\mathcal{P}}_{N_{\mathcal{K}}}^{A_1} \right)^{\top}. \quad (5.22)$$

Der Vektor  $\hat{\mathcal{P}}^{\delta}$  lässt sich darstellen als

$$\hat{\mathcal{P}}^{\delta} = \text{logit}^{-1} \left( X_{\text{Präd}}^{\mathcal{P}^{\delta}} \cdot \hat{\beta}^{\mathcal{P}} \right), \quad (5.23)$$

wobei  $\hat{\beta}^{\mathcal{P}}$  den geschätzten Parametervektor des GAM aus Abschnitt 5.6.1 darstellt und  $X_{\text{Präd}}^{\mathcal{P}^{\delta}}$  die erforderliche Prädiktionsmatrix ist, mit der die in  $\delta_{\text{HI}}$  einfließenden Prävalenzen geschätzt werden. Bei

$$\hat{\eta}^{\mathcal{P}^{\delta}} := X_{\text{Präd}}^{\mathcal{P}^{\delta}} \cdot \hat{\beta}^{\text{HI}} \quad (5.24)$$

handelt es sich um den linearen Prädiktor von  $\hat{\mathcal{P}}^{\delta}$ . Die Kovarianzmatrix  $\hat{V}_{\eta^{\mathcal{P}^{\delta}}}$  des linearen Prädiktors lässt sich aus der Kovarianzmatrix  $\hat{V}_{\beta^{\mathcal{P}}}$  des Parameterschätzers  $\hat{\beta}^{\mathcal{P}}$  berechnen (vgl. Wood 2006):

$$\hat{V}_{\eta^{\mathcal{P}^{\delta}}} = X_{\text{Präd}}^{\mathcal{P}^{\delta}} \cdot \hat{V}_{\beta^{\mathcal{P}}} \cdot \left( X_{\text{Präd}}^{\mathcal{P}^{\delta}} \right)^{\top}. \quad (5.25)$$

Um das Fehlerfortpflanzungsgesetz mit den Ableitungen in Gleichungen 5.14 bis 5.16 anwenden zu können, wird die Kovarianzmatrix  $\hat{V}_{\mathcal{P}^\delta}$  des Vektors  $\hat{\mathcal{P}}^\delta$  benötigt. Für das Element von  $\hat{V}_{\mathcal{P}^\delta}$  in der  $i$ -ten Zeile und der  $j$ -ten Spalte  $\hat{V}_{\mathcal{P}^\delta}[i; j] = \text{Cov}(\hat{\mathcal{P}}_i^\delta; \hat{\mathcal{P}}_j^\delta)$  gilt:

$$\begin{aligned} \text{Cov}(\hat{\mathcal{P}}_i^\delta; \hat{\mathcal{P}}_j^\delta) &= \text{Cov}\left(\text{logit}^{-1}(\eta_i^{\mathcal{P}^\delta}); \text{logit}^{-1}(\eta_j^{\mathcal{P}^\delta})\right) \\ &\simeq \frac{\partial \text{logit}^{-1}(\eta_i^{\mathcal{P}^\delta})}{\partial \eta_i^{\mathcal{P}^\delta}} \cdot \frac{\partial \text{logit}^{-1}(\eta_j^{\mathcal{P}^\delta})}{\partial \eta_j^{\mathcal{P}^\delta}} \cdot \text{Cov}(\eta_i^{\mathcal{P}^\delta}; \eta_j^{\mathcal{P}^\delta}). \end{aligned} \quad (5.26)$$

Anstatt mit Hilfe von Formel 5.26 die Kovarianzmatrix  $\hat{V}_{\mathcal{P}^\delta}$  zu approximieren, ist es ebenfalls möglich  $\hat{\delta}_{\text{HI}}$  nicht nach  $\mathcal{P}_j^{\mathcal{K}}$  und  $\mathcal{P}_j^{\mathcal{A}_1}$ , sondern nach entsprechenden Komponenten von  $\eta^{\mathcal{P}^\delta}$  abzuleiten. Diese Vorgehensweise wird jedoch nicht gewählt, da sie zwar zu dem selben Ergebnis führt, in der Darstellung jedoch wesentlich komplexer ist. Die Komplexität entsteht durch die Terme  $\mathcal{P}_i^{\mathcal{K}}$  und  $\mathcal{P}_i^{\mathcal{A}_1}$  in Formel 5.13, die durch die Inverse der logistischen Funktion in Abhängigkeit von Komponenten von  $\eta^{\mathcal{P}^\delta}$  ersetzt werden müssen. Dadurch steigt die Komplexität der Formeln 5.14 bis 5.16 ebenfalls an. Dass beide Vorgehensweisen zu dem selben Ergebnis führen, kann durch die Anwendung der Kettenregel der Differentialrechnung leicht gezeigt werden: Die Funktion  $\hat{\delta}_{\text{HI}}(\hat{\mathcal{P}}_i^\delta)$  wird mit der Funktion  $\text{logit}^{-1}(\hat{\eta}_i^{\mathcal{P}^\delta})$  verkettet:

$$\hat{\delta}_{\text{HI}}(\hat{\mathcal{P}}_i^\delta) = \hat{\delta}_{\text{HI}}\left(\text{logit}^{-1}(\hat{\eta}_i^{\mathcal{P}^\delta})\right). \quad (5.27)$$

Die Anwendung der Kettenregel führt zu:

$$\begin{aligned} \frac{\partial \hat{\delta}_{\text{HI}}}{\partial \hat{\eta}_i^{\mathcal{P}^\delta}} &= \frac{\partial \hat{\delta}_{\text{HI}}}{\partial \hat{\mathcal{P}}_i^\delta(\hat{\eta}_i^{\mathcal{P}^\delta})} \cdot \frac{\partial \text{logit}^{-1}(\hat{\eta}_i^{\mathcal{P}^\delta})}{\partial \hat{\eta}_i^{\mathcal{P}^\delta}} \\ &= \frac{\partial \hat{\delta}_{\text{HI}}}{\partial \hat{\mathcal{P}}_i^\delta} \Bigg|_{\hat{\mathcal{P}}_i^\delta = \text{logit}^{-1}(\hat{\eta}_i^{\mathcal{P}^\delta})} \cdot \frac{\partial \text{logit}^{-1}(\hat{\eta}_i^{\mathcal{P}^\delta})}{\partial \hat{\eta}_i^{\mathcal{P}^\delta}}. \end{aligned} \quad (5.28)$$

Werden Formeln 5.26 und 5.28 miteinander verglichen und in Beziehung zum Fehlerfortpflanzungsgesetz gesetzt, so lässt sich folgendes erkennen:

Um die Varianz von  $\hat{\delta}_{\text{HI}}$  abzuschätzen, muss  $\text{Cov}(\hat{\mathcal{P}}_i^\delta; \hat{\mathcal{P}}_j^\delta)$  mit der Ableitung von  $\hat{\delta}_{\text{HI}}$  nach  $\hat{\mathcal{P}}_i^\delta$  sowie der Ableitung von  $\hat{\delta}_{\text{HI}}$  nach  $\hat{\mathcal{P}}_j^\delta$  multipliziert werden. Bei beiden Vorgehensweisen fließt der Term

$$\left. \frac{\partial \hat{\delta}_{\text{HI}}}{\partial \hat{\mathcal{P}}_i^\delta} \right|_{\hat{\mathcal{P}}_i^\delta} \cdot \left. \frac{\partial \hat{\delta}_{\text{HI}}}{\partial \hat{\mathcal{P}}_j^\delta} \right|_{\hat{\mathcal{P}}_j^\delta} \cdot \frac{\partial \text{logit}^{-1}(\eta_i^{\mathcal{P}^\delta})}{\partial \eta_i^{\mathcal{P}^\delta}} \cdot \frac{\partial \text{logit}^{-1}(\eta_j^{\mathcal{P}^\delta})}{\partial \eta_j^{\mathcal{P}^\delta}} \cdot \text{Cov}(\eta_i^{\mathcal{P}^\delta}; \eta_j^{\mathcal{P}^\delta}) \quad (5.29)$$

als Summand bei Anwendung des Fehlerfortpflanzungsgesetzes in die Varianzabschätzung von  $\hat{\delta}_{\text{HI}}$  ein.

Die Varianz der Versichertentage  $\mathcal{T}_i^{\mathcal{K}}$  wird mit Hilfe der Poissonverteilung approximiert. Die Poissonverteilung ergibt sich als Grenzwertbetrachtung der Binomialverteilung, bei der der Erwartungswert konstant gehalten wird und die Anzahl der zugrundeliegenden Bernoulliexperimente (Münzwürfe) gegen unendlich geht (Hafner 1989). Es wäre nicht angemessen anzunehmen, dass jeder einzelne Versichertentag zufällig zugeordnet wird. Schließlich werden durch jede Person auf einen Schlag bis zu 365 Versichertentage einer Morbiditäts-Kombination zugeordnet. Um die Varianz der einzelnen  $\mathcal{T}_i^{\mathcal{K}}$  nicht zu unterschätzen, wird angenommen, dass  $\mathcal{T}_i^{\mathcal{K}}/365$  approximativ Poisson-verteilt ist. Da Erwartungswert und Varianz einer Poisson-verteilten Zufallsvariablen identisch sind, gilt:

$$\begin{aligned} \text{Var} \left( \frac{\mathcal{T}_i^{\mathcal{K}}}{365} \right) &= E \left( \frac{\mathcal{T}_i^{\mathcal{K}}}{365} \right) \\ \Leftrightarrow \frac{1}{365^2} \cdot \text{Var}(\mathcal{T}_i^{\mathcal{K}}) &= \frac{E(\mathcal{T}_i^{\mathcal{K}})}{365} \\ \Leftrightarrow \text{Var}(\mathcal{T}_i^{\mathcal{K}}) &= 365 \cdot E(\mathcal{T}_i^{\mathcal{K}}). \end{aligned} \quad (5.30)$$

Da  $E(\mathcal{T}_i^{\mathcal{K}})$  unbekannt ist, wird stattdessen  $\mathcal{T}_i^{\mathcal{K}}$  verwendet, so dass sich der Schätzer

$$\widehat{\text{Var}}(\mathcal{T}_i^{\mathcal{K}}) = 365 \cdot \mathcal{T}_i^{\mathcal{K}} \quad (5.31)$$

ergibt.

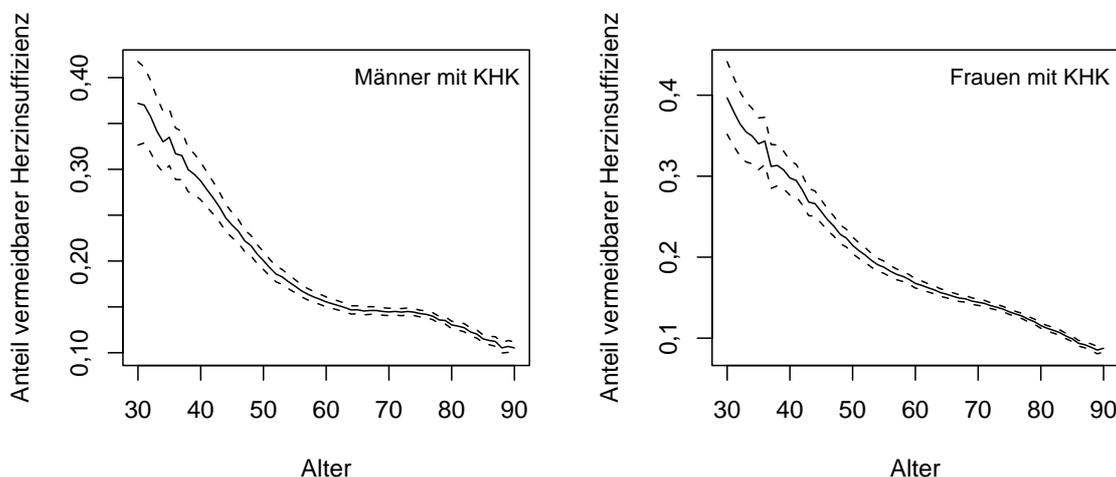


Abbildung 5.14: Anteil vermeidbarer Herzinsuffizienz bei Wegfall der KHK plus/minus das 1,96-fache der Standardabweichung

Es sei  $V_{\mathcal{T}^{\mathcal{K}}}$  die Kovarianzmatrix von  $\mathcal{T}^{\mathcal{K}}$ . Diese lässt sich folgendermaßen schätzen:

$$\hat{V}_{\mathcal{T}^{\mathcal{K}}} = \text{diag} \left( \widehat{\text{Var}} (\mathcal{T}_1^{\mathcal{K}}), \dots, \widehat{\text{Var}} (\mathcal{T}_{NK}^{\mathcal{K}}) \right) = \text{diag} (365 \cdot \mathcal{T}^{\mathcal{K}}) . \quad (5.32)$$

Weiterhin ist es sinnvoll anzunehmen, dass die Tage und die geschätzten Wahrscheinlichkeiten stochastisch unabhängig sind. Das heißt es gilt:

$$\text{Cov} \left( \mathcal{T}_i^{\mathcal{K}}, \hat{\mathcal{P}}_j^{\mathcal{K}} \right) = \text{Cov} \left( \mathcal{T}_i^{\mathcal{K}}, \hat{\mathcal{P}}_j^{\mathcal{A}_1} \right) = 0 . \quad (5.33)$$

Somit sind alle Varianzen und Ableitungen bestimmt, die für die Anwendung des Fehlerfortpflanzungsgesetz benötigt werden. Die Ergebnisse, die sich durch die direkte Anwendung des Fehlerfortpflanzungsgesetzes ergeben, werden in Abbildung 5.14 graphisch und in Anhang B tabellarisch dargestellt.

## 5.7 Effekt der Primärprävention

Es soll ein relatives Risiko  $\text{RR}^{\text{Präv}}$  festgelegt werden, um das sich bei Primärprävention mit Statinen das KHK-Erkrankungsrisiko verringert. Gemäß ei-

ner Meta-Analyse (Vreecer *et al.* 2003) beträgt in der Primärprävention das relative Risiko koronarer Ereignisse 0,67 (95 %-KI: 0,58 bis 0,77). Da nicht alle gescreenten Personen mit positivem Score tatsächlich an der Statintherapie teilnehmen, muss dieser Effekt noch reduziert werden. Es wird von einer Compliance  $C$  von 90 % ausgegangen. Damit schwächt sich das relative Risiko auf  $0,67 \cdot 0,9 + 1 \cdot 0,1 = 0,703$  ab.

# Kapitel 6

## Zuweisung der Kosten und Nutzen

### 6.1 Der Nutzen

Der Nutzen der vorliegenden Analyse wird in qualitätsadjustierten Lebensjahren (QALYs; Drummond *et al.* 2005) gemessen\*. Die Nutzwerte, die den einzelnen Zuständen zugewiesen werden, orientieren sich an entsprechenden Werten anderer Kosten-Nutzwert-Analysen. Dem Zustand „tot“ wird der Nutzwert 0 zugewiesen, alle Zustände ohne KHK erhalten vereinfachend den Nutzwert 1. Alle Zustände von Personen mit koronarer Herzkrankheit erhalten den Nutzwert 0,8. Dieser Wert ist größer als der von Hay und Sterling (2005) verwendete Nutzwert (0,77 für KHK) und kleiner als der für KHK verwendete Nutzwert der Analyse von Hayashino *et al.* (2004) (u. a. 0,88 für Personen nach einem Herzinfarkt).

### 6.2 Zweige der Sozialversicherungen

Die Sozialversicherungen sind in fünf Bereiche untergliedert: die gesetzliche Krankenversicherung (GKV), die gesetzliche Rentenversicherung (GRV), die Arbeitslo-

---

\*QALYs sind keine klinisch messbare Größen. Daher handelt es sich bei der anschließenden gesundheitsökonomischen Evaluation um eine Kosten-Nutzwert-Analyse und der Nutzen wird als Nutzwert bezeichnet.

senversicherung (AV), die soziale Pflegeversicherung (PflV) sowie die gesetzliche Unfallversicherung (UV). Eine Analyse aus der Perspektive der Sozialversicherungen steht vor dem Problem, dass die Versichertenkreise der einzelnen Sozialversicherungszweige nicht identisch sind, auch wenn es große Überschneidungen gibt. Welche Personen in der vorliegenden Arbeit als Versichertenkollektiv der Sozialversicherungen angesehen werden, hängt von der jeweils zugrunde liegenden Datenquelle ab. Zahlungen der Unfallversicherung werden als vernachlässigbar eingestuft und werden weder auf der Einnahmen- noch auf der Ausgabenseite berücksichtigt.

Es ist anzunehmen, dass sowohl die Einnahmen als auch die Ausgaben der einzelnen Sozialversicherungszweige neben dem Alter und dem Geschlecht auch vom Gesundheitszustand der betroffenen Person abhängen. Daher wird im Markovmodell berücksichtigt, dass Personen mit KHK höhere Gesundheitsausgaben haben als Personen ohne KHK. Es wird auch berücksichtigt, dass die Gesundheitsausgaben bei Vermeidung der KHK nicht auf das Niveau der Personen absinkt, die auch ohne Prävention nicht an der KHK erkrankt wären. Somit wird berücksichtigt, dass die Komorbidität bei Personen mit KHK höher ist als bei Personen ohne KHK. Es erscheint ebenfalls plausibel, dass Personen mit hoher Morbidität geringere Beitragszahlungen leisten als gesündere Personen, oder dass sie im Mittel höhere Pflegeausgaben verursachen. Da die zur Verfügung stehende Datengrundlage nicht dazu geeignet ist, diese Effekte zu quantifizieren, werden entsprechende Kosten nur alters- und geschlechtsspezifisch ermittelt. Die durch diese Vorgehensweise entstehenden Verzerrungen erscheinen objektiver zu sein, als ohne quantitative Grundlage eine Annahme zu treffen.

### 6.3 Die SOEP-Daten als Einnahmen- und Ausgabenquelle

Die Einnahmen der Sozialversicherungen sowie die Ausgaben der Renten- und der Arbeitslosenversicherung werden mit Hilfe der SOEP-Daten abgeschätzt. Da in den SOEP-Daten lediglich der Versichertenstatus der GKV dokumentiert ist, wird dieser für alle Zweige der Sozialversicherungen übernommen. Die SOEP-Daten enthalten keine morbiditätsspezifischen Angaben, daher werden Einnahmen und Ausgaben lediglich in Abhängigkeit von Alter und Geschlecht modelliert. Eine besondere Eigenschaft der SOEP-Daten ist es, dass für jede Person ein Hochrechnungsfaktor existiert, der angibt, wieviele Personen der gesamtdeutschen Bevölkerung durch die Person in der Stichprobe repräsentiert werden. Die Hochrechnungsfaktoren sollen in der Modellierung der Kostenkomponente berücksichtigt werden. Aus diesem Grund werden zunächst die durchschnittlichen Kosten einer Alters-Geschlechtsklasse ermittelt. Anschließend werden die aggregierten Daten mit Hilfe eines GAM geglättet. Bei der Glättung soll berücksichtigt werden, dass in jede Alters-Geschlechtsklasse eine unterschiedliche Anzahl von Personen mit verschiedenen Hochrechnungsfaktoren einfließen.

Es bezeichne  $K_{ij}$  die betrachtete Einnahmen- bzw. Ausgabenkomponente der  $i$ -ten Person in der  $j$ -ten Alters-Geschlechtsklasse und  $\mathcal{H}_{ij}$  den der Person zugewiesenen Hochrechnungsfaktor. Der Hochrechnungsfaktor wird im Folgenden als nicht stochastisch angenommen. Die Anzahl Personen in der  $j$ -ten Alters-Geschlechtsklasse werde mit  $N_j$  bezeichnet. Das gewichtete Mittel der Einnahmen- bzw. Ausgabenkomponente der  $j$ -ten Alters-Geschlechtsklasse beträgt dann:

$$K_j = \frac{\sum_{i=1}^{N_j} K_{ij} \cdot \mathcal{H}_{ij}}{\sum_{i=1}^{N_j} \mathcal{H}_{ij}}. \quad (6.1)$$

Es sei angenommen, dass die Kostenkomponenten  $K_{ij}$  alle dieselbe Varianz besitzen

und stochastisch unabhängig sind. Daraus folgt:

$$\begin{aligned}\text{Var}(K_j) &= \text{Var}\left(\frac{\sum_{i=1}^{N_j} K_{ij} \cdot \mathcal{H}_{ij}}{\sum_{i=1}^{N_j} \mathcal{H}_{ij}}\right) \\ &= \frac{\sum_{i=1}^{N_j} \mathcal{H}_{ij}^2}{\left(\sum_{i=1}^{N_j} \mathcal{H}_{ij}\right)^2} \cdot \text{Var}(K_{1j}).\end{aligned}\quad (6.2)$$

Um aus der Varianz von  $K_j$  geeignete Gewichte abzuleiten, wird eine herkömmliche Gewichtung betrachtet. Bei dieser besitzen alle Personen dasselbe Gewicht, d. h. sie haben den Hochrechnungsfaktor 1. In diesem Spezialfall ergibt sich durch Anwendung der Gleichung 6.2:

$$\text{Var}(K_j) = \frac{N_j}{(N_j)^2} \cdot \text{Var}(K_{1j}) = \frac{1}{N_j} \cdot \text{Var}(K_{1j}).\quad (6.3)$$

Da  $K_j$  dann nur noch  $1/N_j$  der Varianz der Kostenkomponente einer einzelnen Person aufweist, fließt  $K_j$  mit einem Gewicht von  $N_j$  in darauf aufbauende Modelle ein. In Analogie zum herkömmlichen Fall werden daher für den SOEP-Datensatz die Gewichte

$$\mathcal{W}_j = \frac{\left(\sum_{i=1}^{N_j} \mathcal{H}_{ij}\right)^2}{\sum_{i=1}^{N_j} \mathcal{H}_{ij}^2}\quad (6.4)$$

für die  $j$ -te Alters-Geschlechtsklasse verwendet.

## 6.4 Einnahmen der Sozialversicherungen

Zu den relevanten Kosten aus Sicht der Sozialversicherungen gehören neben den Ausgaben, die für Leistungen an die Versicherten gezahlt werden, auch Einnahmen, d. h. die Beitragszahlungen der Versicherten (negative Kosten). In gewonnenen Lebensjahren entstehen nicht nur zusätzliche Ausgaben, es gibt auch zusätzliche Beitragszahlungen. Die Einnahmen der Sozialversicherungen werden mit Hilfe der SOEP-Daten bestimmt.

Beiträge werden auf die folgenden im SOEP-Datensatz enthaltenen Einkünfte erhoben: Lohn, 13./14. Monatsgehalt, Weihnachtsgeld, Urlaubsgeld, Gewinnbeteiligung, Einkünfte aus selbständiger Tätigkeit, Nebenerwerb, Arbeitslosengeld (nur zu 80%), Arbeitslosenhilfe<sup>†</sup>, Arbeits- und Invaliditätsrente, Witwen- und Waisenrente, Betriebsrente. Die Einkunftsarten Lohn, 13./14. Monatsgehalt, Weihnachtsgeld, Urlaubsgeld und Gewinnbeteiligung werden zu einer Einkommensgruppe zusammengefasst, die im Folgenden als Lohn bezeichnet wird. Dabei werden die nicht monatlichen Zahlungen, auf die Monate aufgeteilt, in denen Lohn/Gehalt gezahlt wurde. Zugrunde gelegt werden die Beitragssätze gemäß des Verbands der Angestellten-Krankenkassen (VdAK) von 2005. Für die gesetzliche Krankenversicherung (GKV) wird der Wert 14,1% als Beitragssatz verwendet<sup>‡</sup> (siehe Tabelle 6.1). Der Zuschlag in Höhe von 0,25 Prozentpunkten bei Kinderlosen in der Sozialen Pflegeversicherung (SPV) wird vernachlässigt. Die unterschiedlichen Beitragsbemessungsgrenzen (BBG) für Ost- und Westdeutschland werden berücksichtigt. Es wird monatlich bis zur jeweiligen Beitragsbemessungsgrenze verbeitragt. Die Verbeitragung erfolgt separat für jede der Einkommensgruppen.

Nicht jeder Bestandteil des Einkommens wird für jede der Sozialversicherungen verbeitragt. So werden vom Arbeitslosengeld und der Arbeitslosenhilfe keine Beiträge an die Arbeitslosenversicherung (AV) entrichtet. Von Renten müssen keine Beiträge an die Rentenversicherung gezahlt werden.

In anderen Fällen wird nur der halbe Beitragssatz verwendet, um die Einnahmen der Sozialversicherungen zu bestimmen. Auf diese Weise werden Doppelzählungen vermieden. Zahlungen der Rentenversicherung (GRV) an die Krankenversicherungen (GKV) werden bei der Ausgabenbestimmung der GRV nicht erfasst.

---

<sup>†</sup>Mittlerweile sind Sozialhilfe und Arbeitslosenhilfe durch das Arbeitslosengeld II abgelöst worden. Die neue Gesetzeslage kann jedoch noch nicht mit den vorliegenden Daten modelliert werden.

<sup>‡</sup>Schätzwert. Tatsächlicher Wert liegt bei Durchführung der Berechnungen noch nicht vor.

	BBG (Euro pro Monat)	Beitragsatz in %
Rentenversicherung (GRV) West	5200	19,5
Arbeitslosenversicherung (AV) West	5200	6,5
Rentenversicherung (GRV) Ost	4400	19,5
Arbeitslosenversicherung (AV) Ost	4400	6,5
Krankenversicherung (GKV)	3525	14,1
Pflegeversicherung (SPV)	3525	1,7

Tabelle 6.1: Beitragssätze der einzelnen Zweige der Sozialversicherungen im Jahr 2005

Es handelt sich dabei um eine Transferzahlung zwischen einzelnen Zweigen der Sozialversicherungen. Im Datensatz wurden Rentenzahlungen, d.h. Ausgaben der Rentenversicherung erhoben (siehe Abschnitt 6.5). Von den ausgezahlten Renten müssen Beiträge an die anderen Zweige der Sozialversicherungen gezahlt werden. Die Hälfte der Beiträge von Rentnern zählt also als Transferzahlung zwischen einzelnen Versicherungszweigen, die andere Hälfte wird durch Beiträge auf die Renten ermittelt. Daher wird auf Renten nur der halbe Beitragssatz an die relevanten Sozialversicherungen, das sind die GKV und die Pflegeversicherung, entrichtet. Welcher Anteil einer Einkommensgruppe an den jeweiligen Zweig der Sozialversicherungen gezahlt wird, ist in Tabelle 6.2 dargestellt.

Die personenspezifischen Beitragszahlungen werden gemäß Abschnitt 6.3 alters- und geschlechtsspezifisch aggregiert. Da die alters- und geschlechtsspezifischen Beiträge noch stark streuen (siehe Abbildung 6.1) werden sie mit Hilfe eines GAM geglättet. Die jährlichen Beitragszahlungen pro Person werden als Response-Variable verwendet. Das Geschlecht fließt als parametrische Kovariable<sup>§</sup> in die Modellgleichung ein. Als glatte Terme fließen sowohl das Alter als auch das Produkt aus der Indika-

---

<sup>§</sup>Die Kovariablen werden in Anlehnung an Wood (2006) und das R-Paket `mgcv` in „parametrische Kovariablen“ und „glatte Terme“ unterteilt.

	GKV	SPV	AV	GRV
Lohn	1	1	1	1
Selbständigeneinkommen	1	1	0	1
Nebenerwerb	1	1	1	1
Arbeitslosengeld (80%)	0,5	0,5	0	0,5
Arbeitslosenhilfe	0,5	0,5	0	0,5
Arbeits- u. Invaliditätsrente	0,5	0,5	0	0
Witwen- und Waisenrente	0,5	0,5	0	0
Betriebsrente	1	1	0	0

Tabelle 6.2: Anteil des Beitragssatzes, der zwecks Bestimmung der Nettobeitragszahlungen an die Sozialversicherungen verwendet wird

torvariable Geschlecht mit dem Alter in die Modellgleichung ein. Als Basis für die glatten Terme werden „thin plate regression splines“ verwendet. Es wird die log-Linkfunktion gewählt und es wird unterstellt, dass die Responsevariable gammaverteilt ist. Es erfolgt eine Gewichtung, wobei die Gewichte gemäß Abschnitt 6.3 erstellt werden. Die Modellvoraussetzungen werden mit Hilfe der Graphiken in Abbildung 6.3 überprüft. Darin lassen sich keine Unregelmäßigkeiten feststellen. Eine Modellzusammenfassung wird in Tabelle 6.3 gegeben. Mit einem adjustierten R-Quadrat von 0,97 werden die Beitragszahlungen sehr gut durch das Modell beschrieben. Zum Vergleich werden die geglätteten und die ursprünglichen Werte in Abbildung 6.1 dargestellt.

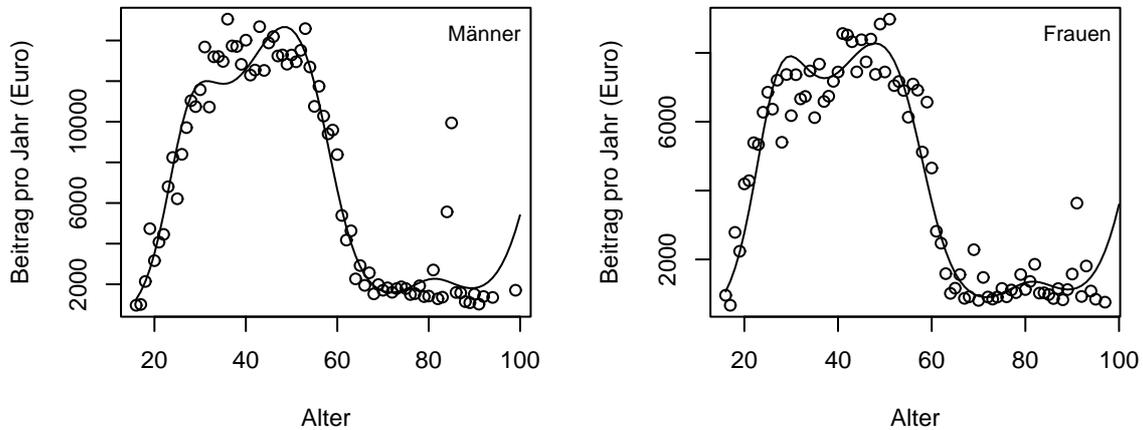


Abbildung 6.1: Jährliche Einnahmen der Sozialversicherungen alters- und geschlechtsspezifisch pro Person in Deutschland

## 6.5 Ausgaben der Renten- und der Arbeitslosenversicherung

Die Ausgaben der Renten- und der Arbeitslosenversicherung werden mit Hilfe der SOEP-Daten abgeschätzt. Für jede Person werden deren jährliche Einkünfte aus „Alters- und Invaliditätsrente“, „Witwen- und Waisenrente“ sowie „Arbeitslosengeld“, „Arbeitslosenhilfe“ und „Unterhaltsgeld“ addiert. Diese entsprechen den personenspezifischen Ausgaben der Renten- und der Arbeitslosenversicherung. Gemäß Abschnitt 6.3 werden für jede Alters-Geschlechtsklasse die durchschnittlichen Ausgaben und der zugehörige Gewichtungsfaktor ermittelt. An diese Daten wird ein GAM mit log-Linkfunktion angepasst. Als Responsevariable werden die jährlichen Zahlungen pro Person der Renten- und der Arbeitslosenversicherung gewählt. Als nicht glatte Kovariable fließt das Geschlecht in die Modellgleichung ein. Außerdem wird eine glatte Funktion des Alters sowie eine glatte Funktion des Produkts aus Alter und Geschlecht (Indikatorvariable) in die Modellgleichung aufgenommen. Als Basis für die glatten Terme werden „thin plate regression splines“ gewählt. Zunächst

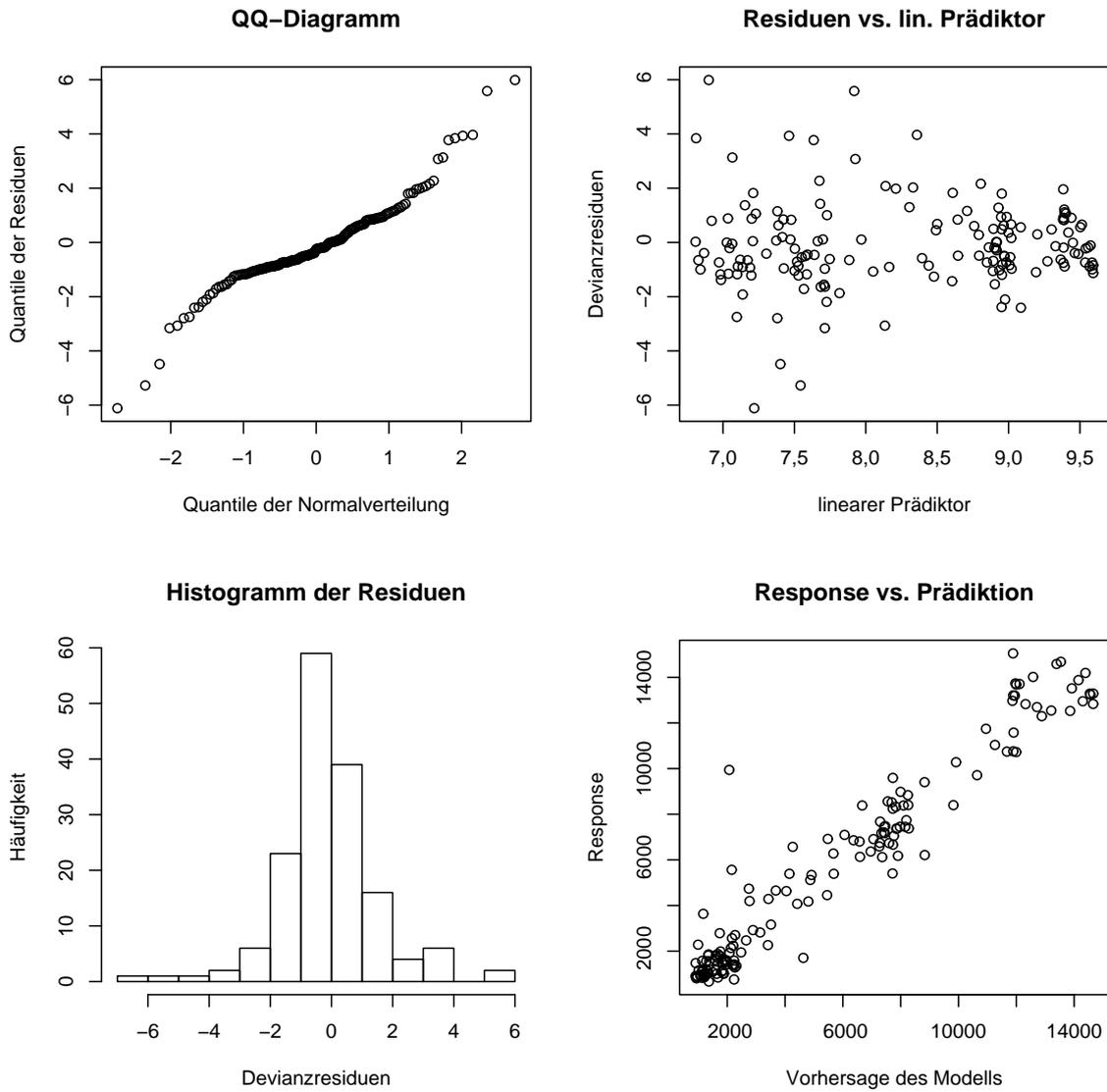


Abbildung 6.2: Graphiken zur Überprüfung der Modellvoraussetzungen des GAMs, in dem die Einnahmen der Sozialversicherungen als gammaverteilt angesehen werden

Parametrische Koeffizienten:				
	Schätzer	Standardfehler	t-Statistik	p-Wert
(Achsenabschnitt)	8,205	0,085	96,7	<0,001
weiblich	0,110	0,157	0,7	0,48
Approximative Signifikanz der glatten Terme:				
	gesch. Freiheitsgrade		F-Statistik	p-Wert
Alter	8,621		160,1	<0,001
Alter · weiblich	2,197		4,3	<0,001
adjustiertes R-Quadrat	0,97		GCV-Score	3,325
erklärte Devianz	92,6 %		Skalenschätzer	2,850

Tabelle 6.3: Modellzusammenfassung des GAM mit gammaverteilter Responsevariable „Beiträge an die Sozialversicherungen“ und log-Linkfunktion

wird von einer Gammaverteilung der Responsevariable ausgegangen. Wie in Abbildung 6.3 zu erkennen ist, sind die Modellvoraussetzungen unter diesen Annahmen verletzt. In der Graphik „Residuen vs. linearer Prädiktor“ ist eine trichterförmige Struktur der Residuen zu erkennen. Das legt die Vermutung nahe, dass die durch die Gammaverteilung implizierte Varianzfunktion  $V(\mu) = \mu^2$  unangemessen ist. Da bei kleineren Werten des linearen Prädiktors die Devianz-Residuen eine größere Streuung aufweisen, ist die Varianzfunktion  $V(\mu) = \mu$  möglicherweise besser geeignet. Das QQ-Diagramm und das Histogramm deuten auf eine leichte Abweichung der Devianzresiduen von der Normalverteilung hin. Aus diesen Gründen wird das GAM modifiziert: Anstatt eine Gammaverteilung anzunehmen werden die Varianzfunktion  $V(\mu) = \mu$  gewählt und die restringierte Quasi-Likelihood maximiert. In Abbildung 6.4 ist zu sehen, dass die Modellvoraussetzungen des veränderten Modells wesentlich besser erfüllt sind. Eine Modellzusammenfassung dieses GAMs wird in Tabelle 6.4 gegeben. Die Anpassung ist verhältnismäßig gut (adjustiertes R-Quadrat von 0,987). Die geglätteten Kosten werden in Abbildung 6.5 graphisch veranschaulicht.

Parametrische Koeffizienten:				
	Schätzer	Standardfehler	t-Statistik	p-Wert
(Achsenabschnitt)	8,407	0,356	23,6	<0,001
weiblich	-1,168	0,713	-1,6	0,10
Approximative Signifikanz der glatten Terme:				
	geschätzte Freiheitsgrade		F-Statistik	p-Wert
Alter	8,31		323,0	<0,001
Alter · weiblich	5,64		2,0	<0,001
adjustiertes R-Quadrat	0,987		GCV-Score	6423
erklärte Devianz	98,4 %		Skalenschätzer	5783

Tabelle 6.4: Modellzusammenfassung des GAM mit quasivverteilter Responsevariable „Ausgaben der Renten- und der Arbeitslosenversicherung“, log-Linkfunktion und Varianzfunktion  $V(\mu) = \mu$

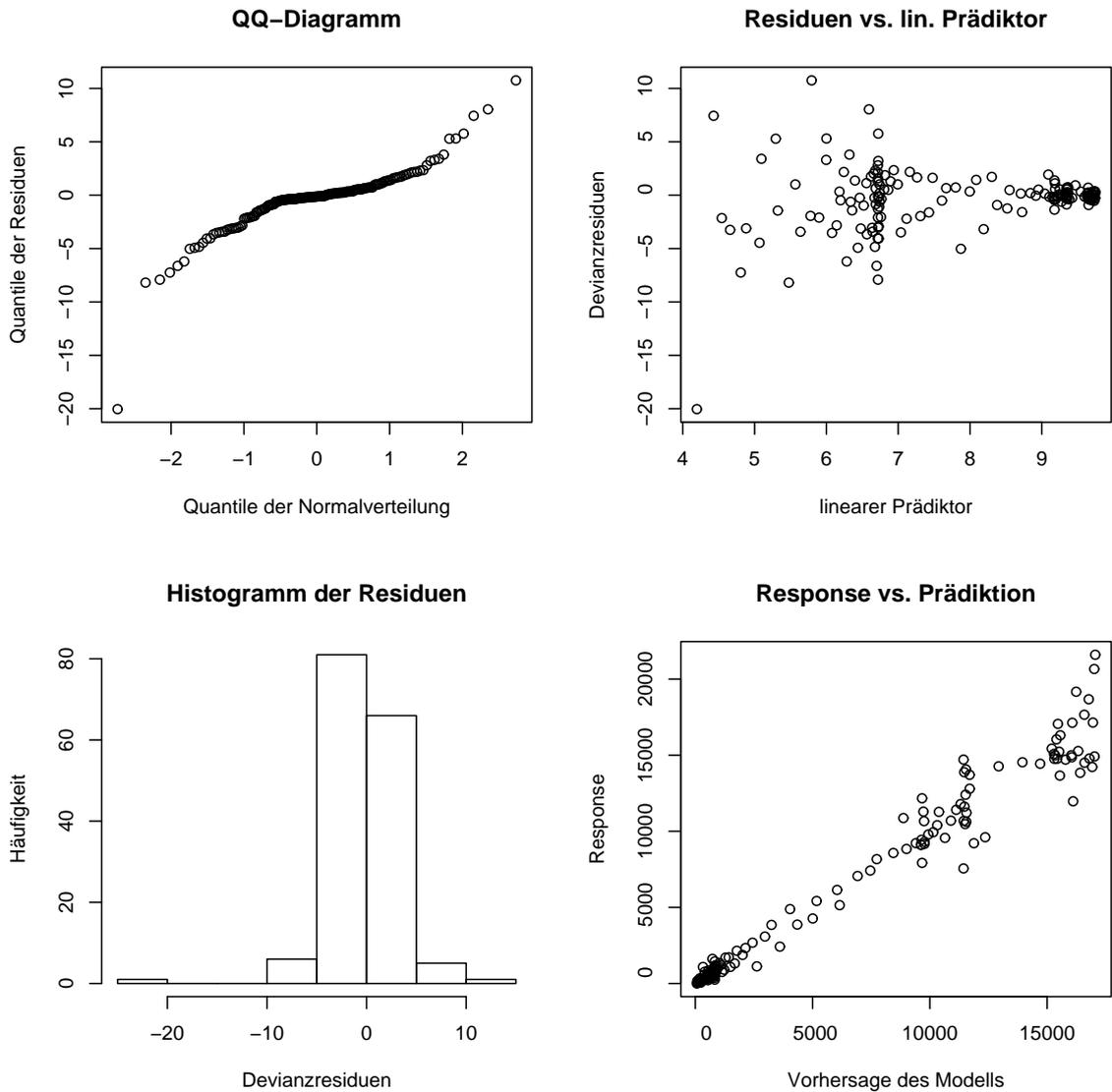


Abbildung 6.3: Graphiken zur Überprüfung der Modellvoraussetzungen; GAM mit Responsevariable „Ausgaben der Renten- und der Arbeitslosenversicherung“; Annahme einer Gammaverteilung

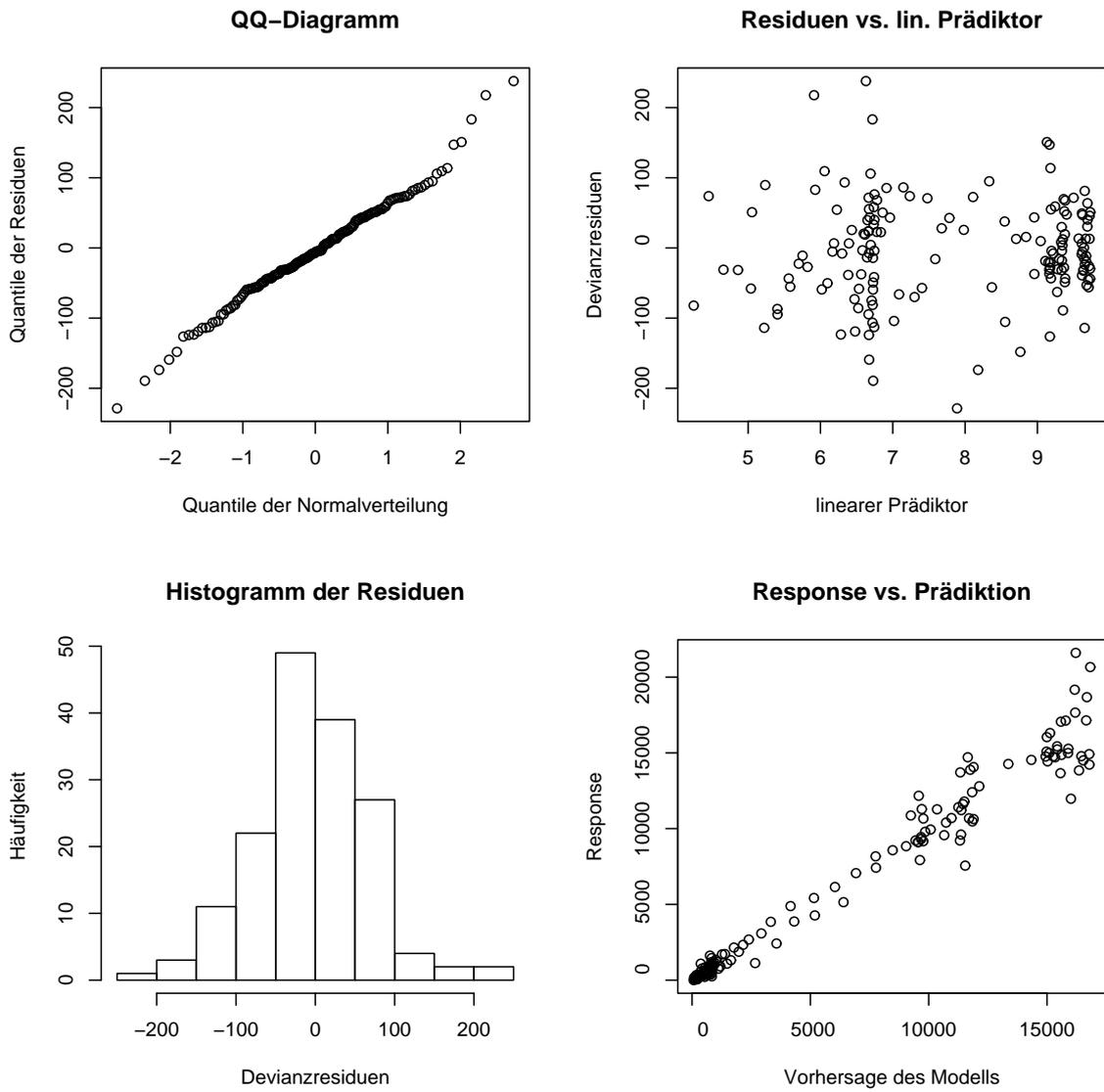


Abbildung 6.4: Graphiken zur Überprüfung der Modellvoraussetzungen; GAM mit Responsevariable „Ausgaben der Renten- und der Arbeitslosenversicherung“, Quasi-Familie

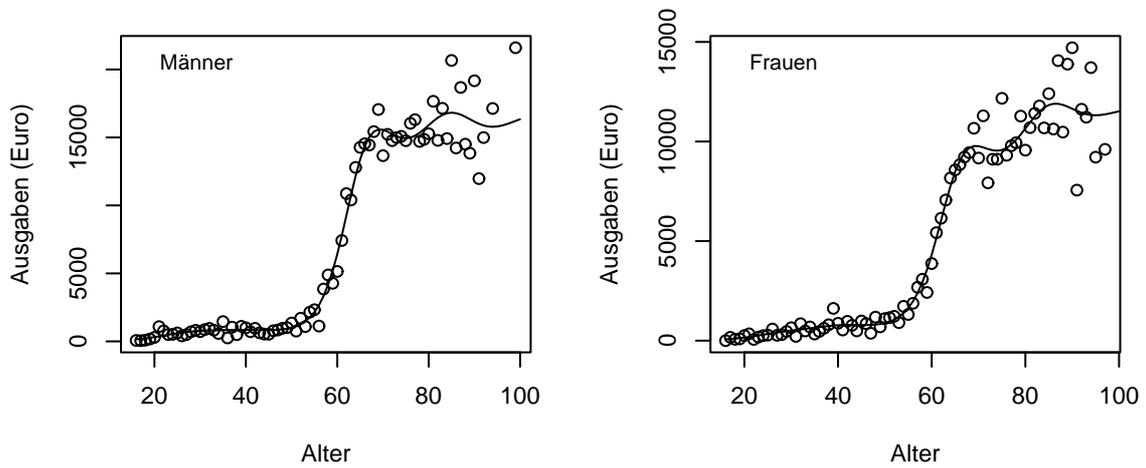


Abbildung 6.5: Ausgaben der Renten- und der Arbeitslosenversicherung pro Person pro Jahr

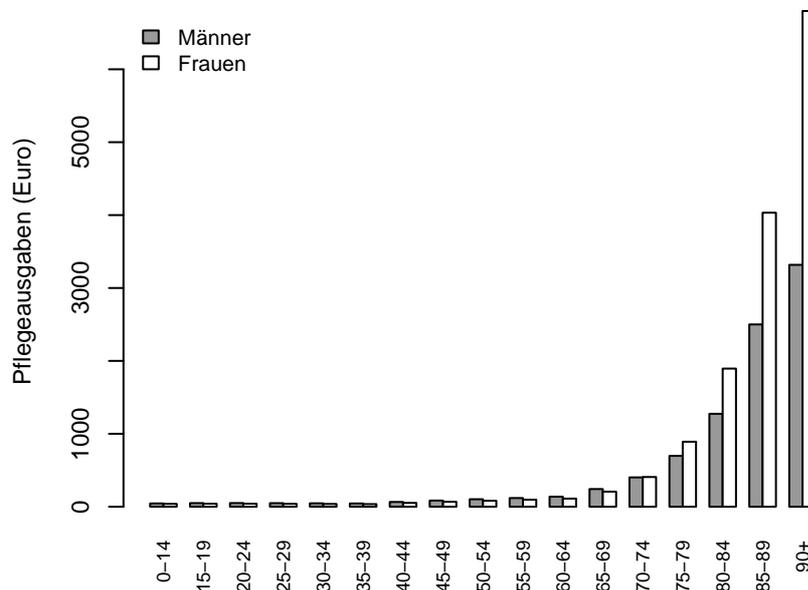


Abbildung 6.6: Durchschnittliche Pflegeausgaben pro Person pro Jahr in Deutschland

## 6.6 Ausgaben der Pflegeversicherung

Die Ausgaben der Pflegeversicherung werden alters- und geschlechtsspezifisch gemäß Lauterbach *et al.* (2005) bestimmt. Die berechneten Werte basieren auf den offiziellen Statistiken des Bundesministerium für Gesundheit und Soziales (BMGS) des Bezugsjahres 2003 und werden in Abbildung 6.6 dargestellt.

## 6.7 Ausgaben der gesetzlichen Krankenversicherung

Die Ausgaben der gesetzlichen Krankenversicherung werden mit Hilfe von zwei Quellen abgeschätzt: Den Risikostrukturausgleich-Daten des BVA (siehe Abschnitt 5.1.4) und den Krankenkassendaten (siehe Abschnitt 5.1.3). Die Krankenkassendaten sind älter als die Daten des BVA und umfassen nur drei Leistungsbereiche

(Arzneimittel, Krankenhaus und Krankengeld). Im Gegensatz zu den Daten des BVA enthalten sie jedoch morbiditätsspezifische Informationen. Aus ihnen kann abgeleitet werden, um wieviel die Gesundheitsausgaben bei Personen mit KHK höher sind als bei Personen ohne KHK. Außerdem kann aus ihnen abgeleitet werden, welcher Anteil der Gesundheitsausgaben bei Personen mit KHK wegfallen würde, wenn bei diesen Personen die KHK durch Prävention vermieden würde. Diese Kostenproportionen können verwendet werden, um die aktuelleren und umfassenderen Gesundheitsausgaben gemäß BVA auf die einzelnen Zustände des Markovmodells aufzuteilen.

Für die Ausgaben der Leistungsbereiche Zahnärzte und nichtärztliche Leistungsausgaben für Dialyse wird angenommen, dass sie nicht durch die KHK beeinflusst werden. Die übrigen Ausgaben, d. h. die der Leistungsbereiche Ärzte, Apotheken, Krankenhäuser, Mehrausgaben DMP (Disease Management Programme) und sonstige Leistungsausgaben werden unter Rückgriff auf die in diesem Abschnitt gewonnenen Erkenntnisse aufgeteilt.

Aus den Daten des BVA werden alters- und geschlechtsspezifisch sowohl die Versichertenzeiten als auch die Ausgaben für jeden der Leistungsbereiche ermittelt. Für jeden Leistungsbereich und jede Alters-/Geschlechtsgruppe werden die durchschnittlichen Ausgaben bestimmt, indem die Ausgaben durch die Versichertenzeiten geteilt werden. Der Verlauf der Gesamtausgaben aller Leistungsbereiche wird in Abbildung 6.7 graphisch dargestellt. Der in der Abbildung erkennbare Sprung bei den 35-Jährigen ist auf die Risikostruktur-Ausgleichsverordnung zurückzuführen (Bundesversicherungsamt 2005). Die Versichertenzeiten der Gruppen<sup>¶</sup> VG04, VG05 und VG06 sind „mit einem Alter unter 35 dem Alter 35 zuzuordnen“ (Bundesversicherungsamt 2005, § 2, Abs. 3). Diese drei Gruppen stehen für Personen, die eine Er-

---

<sup>¶</sup>Die genauen Bedeutungen der einzelnen Gruppenbezeichnungen werden in der Datensatzbeschreibung des Bundesversicherungsamtes erläutert.

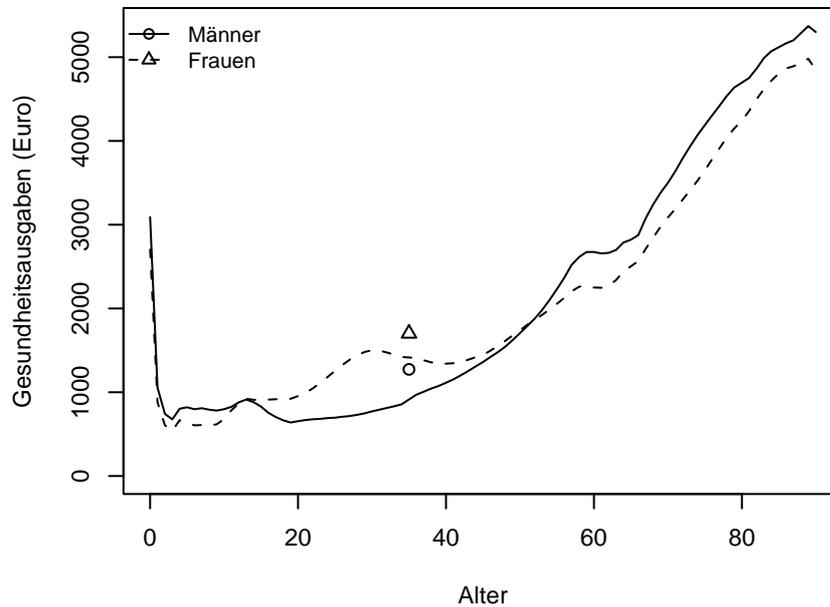


Abbildung 6.7: Jährliche Leistungsausgaben der GKV pro Versichertem; bei 35-Jährigen wird aufgrund der starken Abweichung ein geglätteter Wert verwendet

werbsminderungsrente beziehen. Personen dieser Gruppen verursachen höhere Gesundheitsausgaben als Personen, die keine Erwerbsminderungsrente beziehen. Da unbekannt ist, wie sich die entsprechenden Versichertenzeiten auf die unter 35-Jährigen verteilen, werden die Gesundheitsausgaben der unter 35-Jährigen beibehalten und die der 35-Jährigen auf den Mittelwert der 34- und 36-Jährigen gesetzt. Die Gesundheitsausgaben der Personen im Alter von 35 und jünger werden somit leicht unterschätzt.

### 6.7.1 GAM zur Modellierung der Gesundheitsausgaben

Die Gesundheitsausgaben pro Person pro Tag werden mit einem GAM modelliert. Es werden die Krankenkassendaten der vier Kassen verwendet, für die sowohl Angaben zur Morbidität als auch zu den Ausgaben vorliegen (vgl. Abschnitt 5.1.3). Bei einigen Zeilen (9 von 9513) werden Kosten in Höhe von 0 realisiert. Um eine

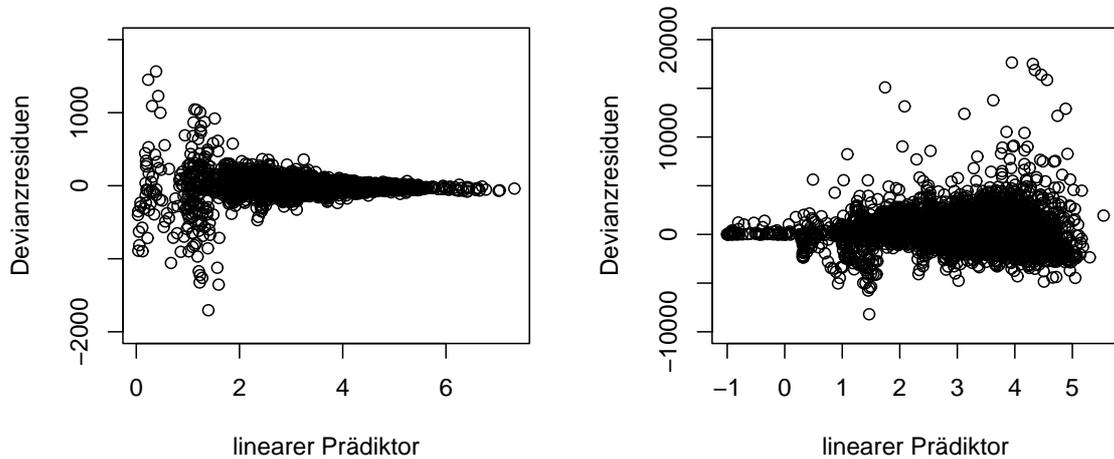


Abbildung 6.8: Unangemessene Varianzfunktionen; Links:  $V(\mu) = \mu^2$ , rechts:  $V(\mu) = 1$

Transformation der Responsevariablen, wie z. B. die Addition eines geringen Wertes, zu umgehen, werden die Kosten nicht als gammaverteilt angenommen, sondern es wird ein Modell der Quasi-Familie angepasst. Als Varianzfunktion wird  $V(\mu) = \mu$  gewählt, da es bei dieser Varianzfunktion nicht zu einem trichterförmigen Anstieg bzw. einer trichterförmigen Abnahme der Devianz-Residuen in Abhängigkeit des linearen Prädiktors kommt (vgl. Abbildung 6.8). Wie bei Modellierungen von Kostendaten üblich, wird eine log-Linkfunktion gewählt. Als Basis für die glatten Terme werden „thin plate regression splines“ verwendet. Die Glättungsparameter werden mittels GCV bestimmt. Folgende Kovariablen fließen in das Modell ein: Ein glatter Term des Alters, das Geschlecht, die chronischen Erkrankungen, Wechselwirkungen zwischen jeweils zwei chronischen Erkrankungen, ein glatter Term des Produkts aus Alter und Geschlecht, sowie glatte Terme des Produkts aus Alter mit jeweils einer chronischen Erkrankung. Die Versichertentage werden als Gewichte verwendet. Die Modellvoraussetzungen können mit Hilfe der Abbildung 6.9 überprüft werden. Die

Varianzfunktion kann als angemessen betrachtet werden (vgl. Graphik „Residuen vs. linearer Prädiktor“, oben rechts). Die Devianzresiduen weichen von einer Normalverteilung ab. Dennoch ist die Modellanpassung mit einem adjustierten R-Quadrat von 0,895 verhältnismäßig gut (siehe Tabelle 6.7), so dass das Modell als legitim angesehen wird. Die Schätzwerte der parametrischen Koeffizienten befinden sich in Tabelle 6.5. Der Verlauf der glatten Terme wird in Abbildung 6.10 veranschaulicht. Die approximative Signifikanz der glatten Terme kann der Tabelle 6.6 entnommen werden.

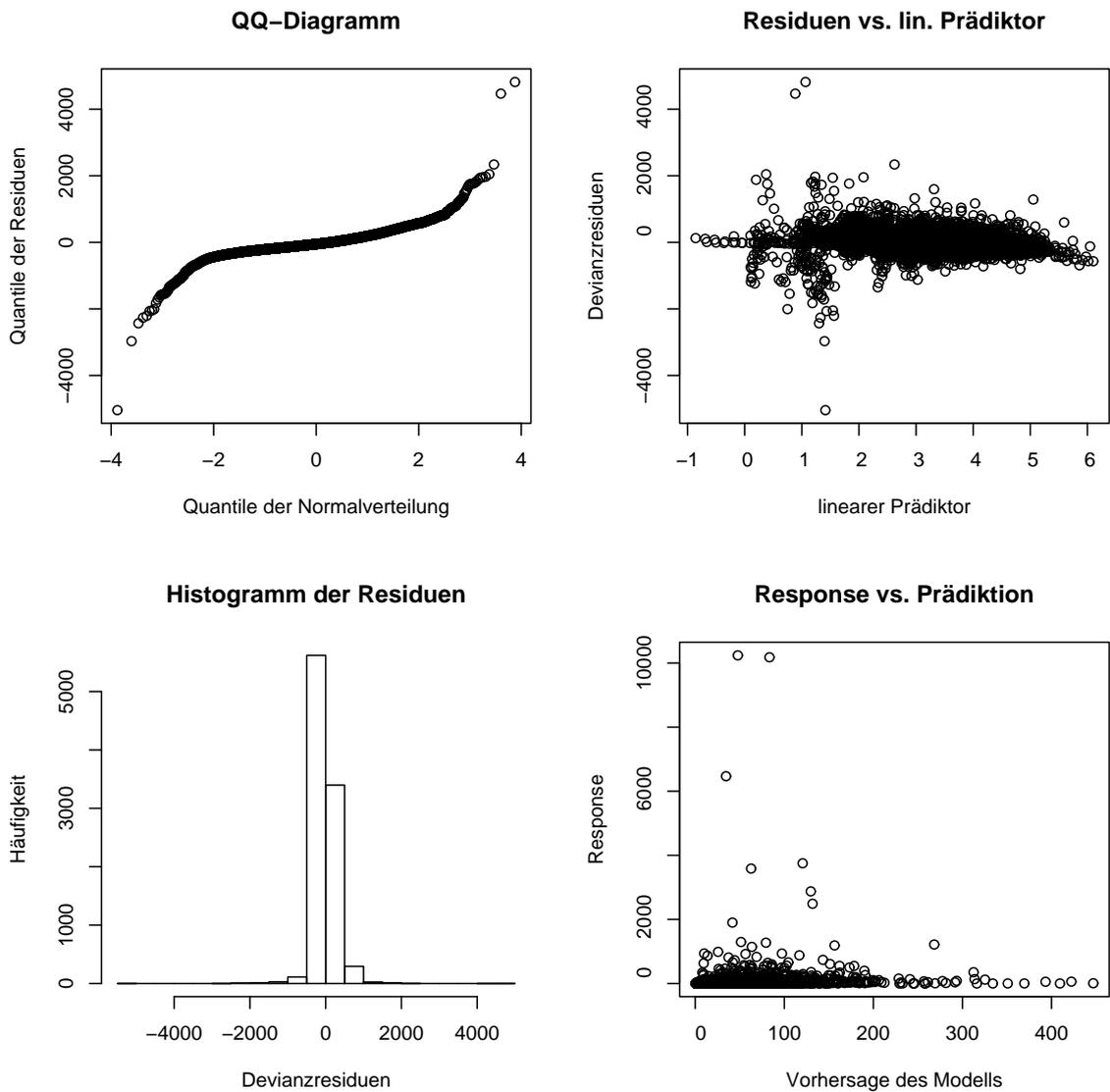


Abbildung 6.9: Graphiken zur Überprüfung der Modellvoraussetzungen des GAM, mit dem die Ausgaben der GKV modelliert werden

	Schätzer	Standardfehler	t-Statistik	p-Wert
(Achsenabschnitt)	1,19	0,65	1,83	0,07
weiblich	-0,20	0,04	-5,09	<0,001
Diabetes	1,81	0,23	8,02	<0,001
KHK	-1,25	0,71	-1,75	0,08
Hypertonie	1,01	0,19	5,30	<0,001
Asthma	1,22	0,04	29,70	<0,001
Herzinsuffizienz	2,74	0,16	16,66	<0,001
Brustkrebs	-1,12	2,94	-0,38	0,70
Schlaganfall	1,68	0,31	5,40	<0,001
Diabetes & KHK	-0,07	0,01	-5,66	<0,001
Diabetes & Hypertonie	-0,43	0,02	-23,19	<0,001
Diabetes & Asthma	-0,14	0,01	-9,57	<0,001
Diabetes & Herzinsuffizienz	-0,03	0,02	-2,07	0,04
Diabetes & Brustkrebs	-0,46	0,04	-10,80	<0,001
Diabetes & Schlaganfall	-0,24	0,02	-14,04	<0,001
KHK & Hypertonie	-0,10	0,02	-5,45	<0,001
KHK & Asthma	-0,13	0,01	-10,20	<0,001
KHK & Herzinsuffizienz	0,02	0,01	1,25	0,21
KHK & Brustkrebs	-0,41	0,04	-10,16	<0,001
KHK & Schlaganfall	-0,30	0,02	-18,72	<0,001
Hypertonie & Asthma	-0,34	0,02	-20,62	<0,001
Hypertonie & Herzinsuffizienz	-0,12	0,02	-5,47	<0,001
Hypertonie & Brustkrebs	-0,44	0,04	-12,08	<0,001
Hypertonie & Schlaganfall	-0,41	0,02	-16,50	<0,001
Asthma & Herzinsuffizienz	-0,03	0,02	-2,10	0,04
Asthma & Brustkrebs	-0,38	0,04	-10,10	<0,001
Asthma & Schlaganfall	-0,25	0,02	-12,01	<0,001
Herzinsuffizienz & Brustkrebs	-0,19	0,04	-5,04	<0,001
Herzinsuffizienz & Schlaganfall	-0,26	0,02	-11,57	<0,001
Brustkrebs & Schlaganfall	-0,82	0,07	-12,51	<0,001

Tabelle 6.5: Schätzer und Teststatistiken der parametrischen Kovariablen des GAM der GKV-Ausgaben

	gesch. Freiheitsgrade	F	p-Wert
Alter	9,0	1283,9	<0,001
Alter · Diabetes	8,9	197,2	<0,001
Alter · KHK	8,6	120,4	<0,001
Alter · Hypertonie	8,9	65,4	<0,001
Alter · Asthma	9,0	16,5	<0,001
Alter · Herzinsuffizienz	8,1	29,6	<0,001
Alter · Brustkrebs	8,9	3,3	0,001
Alter · Schlaganfall	8,8	36,6	<0,001
Alter · weiblich	8,9	282,5	<0,001

Tabelle 6.6: Approximative Signifikanz der glatten Terme des GAM mit Responsevariable „Gesundheitsausgaben pro Person pro Tag“

adjustiertes R-Quadrat	0,985	GCV-Score	80470
erklärte Devianz	97 %	Skalenschätzer	79548

Tabelle 6.7: Statistiken des gesamten GAM mit Responsevariable „Gesundheitsausgaben pro Person pro Tag“

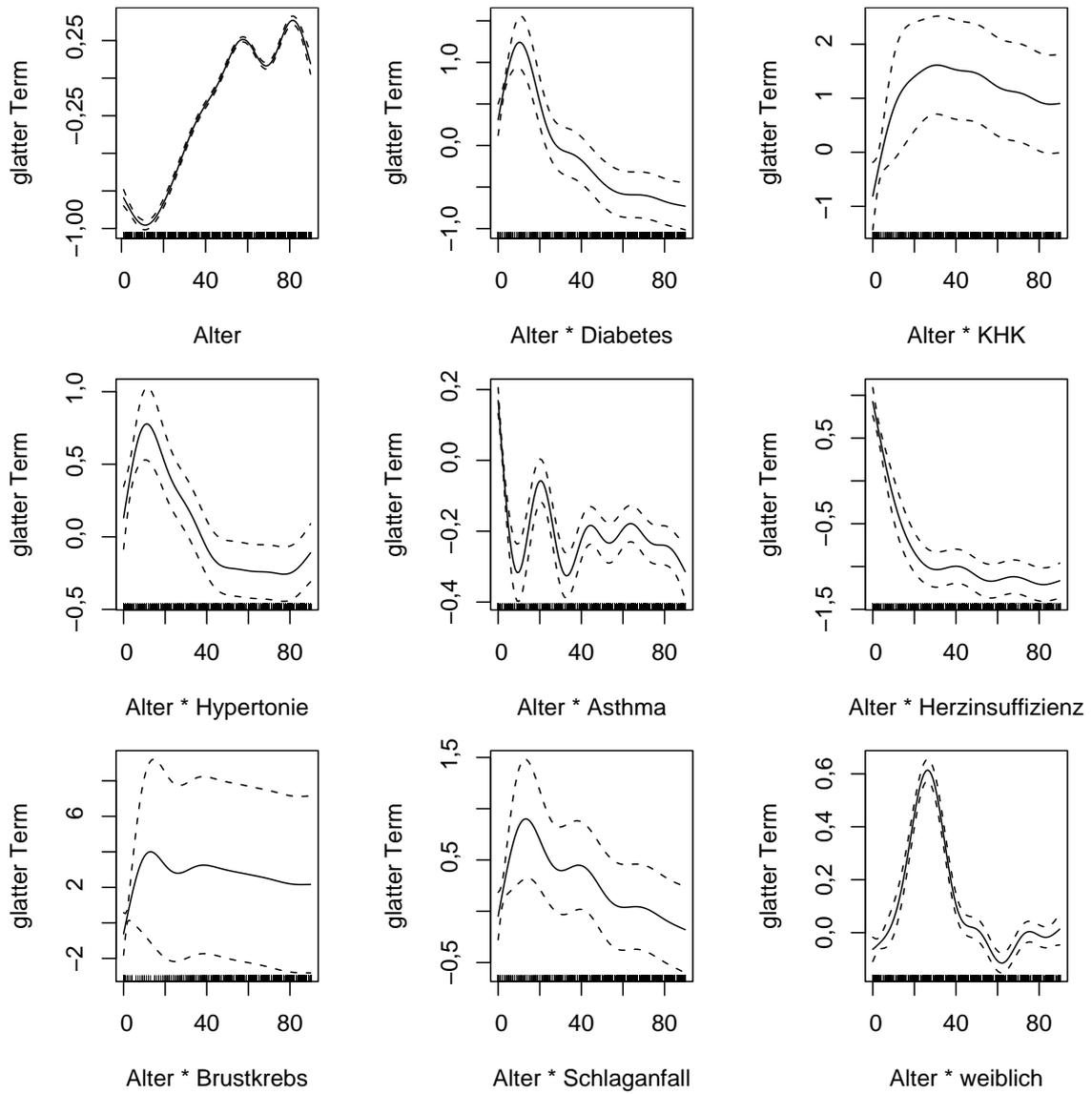


Abbildung 6.10: Glatte Terme des GAM, mit dem die Ausgaben der GKV modelliert werden

### 6.7.2 Kostenproportionen

Es sollen Proportionen von Gesundheitsausgaben zwischen gewissen Personenkollektiven bestimmt werden. Von Interesse ist erstens, um das Wievielfache die Gesundheitsausgaben pro Person mit KHK die Gesundheitsausgaben pro Person ohne KHK übersteigen, und zweitens, welcher Anteil der Gesundheitsausgaben bei Personen mit KHK (Szenario ohne Prävention) vermieden werden kann, wenn die koronare Herzkrankheit (durch Prävention) wegfällt. Die Proportionen werden alters- und geschlechtsspezifisch bestimmt. In ihre Berechnung fließen zwei GAMs ein: Das GAM aus Abschnitt 5.6.1, in dem die Prävalenz der Herzinsuffizienz in Abhängigkeit von Alter, Geschlecht und Komorbidität modelliert wird, sowie das GAM aus Abschnitt 6.7.1 mit der Responsevariablen Gesundheitsausgaben pro Person pro Tag.

Es seien  $g_K$  die zu erwartenden Gesundheitsausgaben pro Person pro Versichertentag der Personen mit KHK. Analog seien  $g_{\bar{K}}$  die entsprechenden Ausgaben von Personen ohne KHK. Wird das Kollektiv der Personen mit KHK betrachtet, aber davon ausgegangen, dass bei ihnen die KHK vermieden wurde, so ergeben sich bei ihnen adjustierte Ausgaben die mit  $g_A$  bezeichnet werden. In  $g_A$  sollen die Komorbiditäten der KHK im Prinzip bestehen bleiben, es wird jedoch davon ausgegangen, dass ein Teil der Herzinsuffizienz wegfällt, und zwar der Teil der Herzinsuffizienz, der als Folgeerkrankung auf die KHK zurückzuführen ist. Deshalb werden zwei Adjustierungen durchgeführt: In einer ersten Adjustierung wird das Kollektiv der Personen mit KHK derart modifiziert, dass der Status KHK auf „nein“ gesetzt wird, die übrigen Morbiditäten aber bestehen bleiben. Die entsprechenden Gesundheitsausgaben werden mit  $g_{A_1}$  bezeichnet. In der zweiten Adjustierung fällt zusätzlich zur KHK die Herzinsuffizienz weg. Die Gesundheitsausgaben werden mit  $g_{A_2}$  bezeichnet. Diese beiden adjustierten Gesundheitsausgaben werden mit  $\delta_{HI}$ , dem Anteil vermeidbarer

Herzinsuffizienz (durch Wegfall von KHK), gewichtet. Es gelte:

$$g_{\mathcal{A}} := \delta_{\text{HI}} \cdot g_{\mathcal{A}_2} + (1 - \delta_{\text{HI}}) \cdot g_{\mathcal{A}_1}. \quad (6.5)$$

Die gesuchten Proportionen lassen sich dann folgendermaßen ausdrücken:

$$Q_1 := \frac{g_{\mathcal{K}}}{g_{\bar{\mathcal{K}}}} \quad \text{und} \quad Q_2 := \frac{g_{\mathcal{K}} - \delta_{\text{HI}} \cdot g_{\mathcal{A}_2} - (1 - \delta_{\text{HI}}) \cdot g_{\mathcal{A}_1}}{g_{\mathcal{K}}}. \quad (6.6)$$

Die zugrunde liegenden Größen  $g_{\bar{\mathcal{K}}}$ ,  $g_{\mathcal{K}}$ ,  $g_{\mathcal{A}_1}$  und  $g_{\mathcal{A}_2}$  können aus den Krankenkassendaten mit Hilfe des GAM mit Responsevariablen „Gesundheitsausgaben pro Person pro Tag“ geschätzt werden.

Es sei

$$\begin{pmatrix} \mathcal{K} \\ \bar{\mathcal{K}} \end{pmatrix} \quad (6.7)$$

ein Teildatensatz der Krankenkassendaten, der genau die Zeilen enthält, die sich auf eine Kohorte von Personen gleichen Alters und gleichen Geschlechts beziehen. Weiter enthalte  $\mathcal{K}$  (vgl. Abschnitt 5.6.2) nur Zeilen, die sich auf Personen mit KHK, und  $\bar{\mathcal{K}}$  nur Zeilen, die sich auf Personen ohne KHK beziehen. Die Anzahl der Zeilen von  $\mathcal{K}$  sei mit  $N_{\mathcal{K}}$ , die Anzahl der Zeilen von  $\bar{\mathcal{K}}$  mit  $N_{\bar{\mathcal{K}}}$  bezeichnet. Weiter sei  $\mathcal{T}_j^{\mathcal{K}}$ ,  $j = 1, \dots, N_{\mathcal{K}}$ , die  $j$ -te Komponente der Versichertentage des Datensatzes  $\mathcal{K}$  und  $\mathcal{T}_\ell^{\bar{\mathcal{K}}}$ ,  $\ell = 1, \dots, N_{\bar{\mathcal{K}}}$ , die  $\ell$ -te Komponente der Versichertentage des Datensatzes  $\bar{\mathcal{K}}$ . Es seien  $g_j^{\mathcal{K}}$  die zu erwartenden Gesundheitsausgaben pro Versichertentag einer Person mit Kovariablen, die der  $j$ -ten Zeile von  $\mathcal{K}$  entsprechen. Analog seien  $g_\ell^{\bar{\mathcal{K}}}$  die entsprechenden Gesundheitsausgaben einer Person mit den Kovariablen der  $\ell$ -ten Zeile von  $\bar{\mathcal{K}}$ . Die Größen  $g_j^{\mathcal{K}}$  und  $g_\ell^{\bar{\mathcal{K}}}$  können mit Hilfe des GAM mit der Responsevariablen „Gesundheitsausgaben pro Person pro Tag“ geschätzt werden. Diese Schätzer werden mit  $\hat{g}_j^{\mathcal{K}}$  und  $\hat{g}_\ell^{\bar{\mathcal{K}}}$  bezeichnet.

Die Ausgaben  $g_{\mathcal{K}}$  und  $g_{\bar{\mathcal{K}}}$  lassen sich folgendermaßen schätzen:

$$\hat{g}_{\mathcal{K}} = \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}}}, \quad (6.8)$$

$$\hat{g}_{\bar{\mathcal{K}}} = \frac{\sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}} \cdot \hat{g}_i^{\bar{\mathcal{K}}}}{\sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}}}. \quad (6.9)$$

Die Ausgaben  $g_{\mathcal{A}_1}$  und  $g_{\mathcal{A}_2}$  können auf eine ähnliche Art geschätzt werden. Dazu muss der Datensatz  $\mathcal{K}$  der Personen mit KHK modifiziert werden. In einer ersten Modifikation wird der Status KHK auf „nein“ gesetzt. Der modifizierte Datensatz wird mit  $\mathcal{A}_1$  bezeichnet (vgl. Abschnitt 5.6.2). In einer zweiten Modifikation wird zusätzlich der Status Herzinsuffizienz auf „nein“ gesetzt. Dieser Datensatz erhält die Bezeichnung  $\mathcal{A}_2$ . Analog zu den Größen  $g_j^{\mathcal{K}}$  und  $\hat{g}_j^{\mathcal{K}}$  werden die Größen  $g_j^{\mathcal{A}_1}$ ,  $g_j^{\mathcal{A}_2}$ ,  $\hat{g}_j^{\mathcal{A}_1}$ , und  $\hat{g}_j^{\mathcal{A}_2}$  definiert.

Die Ausgaben  $g_{\mathcal{A}_1}$  und  $g_{\mathcal{A}_2}$  werden folgendermaßen geschätzt:

$$\hat{g}_{\mathcal{A}_1} = \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{A}_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}}} \quad \text{und} \quad \hat{g}_{\mathcal{A}_2} = \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{A}_2}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}}}. \quad (6.10)$$

Somit lassen sich Schätzer für die Quotienten  $\mathcal{Q}_1$  und  $\mathcal{Q}_2$  entwickeln:

$$\hat{\mathcal{Q}}_1 := \frac{\hat{g}_{\mathcal{K}}}{\hat{g}_{\bar{\mathcal{K}}}} \quad \text{und} \quad (6.11)$$

$$\hat{\mathcal{Q}}_2 := \frac{\hat{g}_{\mathcal{K}} - \hat{\delta}_{\text{HI}} \cdot \hat{g}_{\mathcal{A}_2} - (1 - \hat{\delta}_{\text{HI}}) \cdot \hat{g}_{\mathcal{A}_1}}{\hat{g}_{\mathcal{K}}}. \quad (6.12)$$

Der Schätzer  $\hat{\mathcal{Q}}_1$  lässt sich als Funktion der Zufallsvariablen  $\hat{g}_j^{\mathcal{K}}$ ,  $\mathcal{T}_j^{\mathcal{K}}$ ,  $\hat{g}_\ell^{\bar{\mathcal{K}}}$  und  $\mathcal{T}_\ell^{\bar{\mathcal{K}}}$  darstellen:

$$\hat{\mathcal{Q}}_1 = \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}}} \bigg/ \frac{\sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}} \cdot \hat{g}_i^{\bar{\mathcal{K}}}}{\sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}}} \quad (6.13)$$

Um die Varianz von  $\hat{\mathcal{Q}}_1$  zu approximieren, muss  $\hat{\mathcal{Q}}_1$  nach  $\mathcal{T}_j^{\mathcal{K}}$ ,  $\hat{g}_j^{\mathcal{K}}$ ,  $\mathcal{T}_\ell^{\bar{\mathcal{K}}}$  und  $\hat{g}_\ell^{\bar{\mathcal{K}}}$  abgeleitet werden:

$$\frac{\partial \hat{\mathcal{Q}}_1}{\partial \mathcal{T}_j^{\mathcal{K}}} = \frac{\left( \sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}} \right) \left( \hat{g}_j^{\mathcal{K}} \cdot \left( \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \right) - \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}} \right)}{\left( \sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}} \cdot \hat{g}_i^{\bar{\mathcal{K}}} \right) \left( \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \right)^2}, \quad (6.14)$$

$$\frac{\partial \hat{Q}_1}{\partial \hat{g}_j^{\mathcal{K}}} = \frac{\left( \sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}} \right) \cdot \mathcal{T}_j^{\mathcal{K}}}{\left( \sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}} \cdot \hat{g}_i^{\bar{\mathcal{K}}} \right) \left( \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \right)}, \quad (6.15)$$

$$\frac{\partial \hat{Q}_1}{\partial \mathcal{T}_\ell^{\bar{\mathcal{K}}}} = \frac{\left( \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}} \right) \left( \left( \sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}} \cdot \hat{g}_i^{\bar{\mathcal{K}}} \right) - g_\ell^{\bar{\mathcal{K}}} \cdot \left( \sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}} \right) \right)}{\left( \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \right) \cdot \left( \sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}} \cdot \hat{g}_i^{\bar{\mathcal{K}}} \right)^2}, \quad (6.16)$$

$$\frac{\partial \hat{Q}_1}{\partial g_\ell^{\bar{\mathcal{K}}}} = \frac{\left( \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}} \right) \cdot \mathcal{T}_\ell^{\bar{\mathcal{K}}} \cdot \left( \sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}} \right)}{\left( \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \right) \cdot \left( \sum_{i=1}^{N_{\bar{\mathcal{K}}}} \mathcal{T}_i^{\bar{\mathcal{K}}} \cdot \hat{g}_i^{\bar{\mathcal{K}}} \right)^2}. \quad (6.17)$$

Es wird ein Vektor definiert, der alle Versichertentage enthält, die benötigt werden, um  $Q_1$  zu schätzen:

$$\mathcal{T}^{\mathcal{Q}_1} := \left( \mathcal{T}_1^{\mathcal{K}}, \dots, \mathcal{T}_{N_{\mathcal{K}}}^{\mathcal{K}}, \mathcal{T}_1^{\bar{\mathcal{K}}}, \dots, \mathcal{T}_{N_{\bar{\mathcal{K}}}}^{\bar{\mathcal{K}}} \right)^\top. \quad (6.18)$$

Um das Fehlerfortpflanzungsgesetz anwenden zu können, muss die Kovarianzmatrix  $V_{\mathcal{T}^{\mathcal{Q}_1}}$  von  $\mathcal{T}^{\mathcal{Q}_1}$  geschätzt werden. Da die Elemente von  $\mathcal{T}^{\mathcal{Q}_1}$  approximativ stochastisch unabhängig sind, wird in Analogie zu Abschnitt 5.6.2 der folgende Schätzer gewählt:

$$\hat{V}_{\mathcal{T}^{\mathcal{Q}_1}} = \text{diag} \left( \widehat{\text{Var}}(\mathcal{T}_1^{\mathcal{Q}_1}), \dots, \widehat{\text{Var}}(\mathcal{T}_{N_{\mathcal{K}}+N_{\bar{\mathcal{K}}}}^{\mathcal{Q}_1}) \right) = \text{diag} (365 \cdot \mathcal{T}^{\mathcal{Q}_1}). \quad (6.19)$$

Weiter wird ein Vektor definiert, der alle geschätzten Gesundheitsausgaben pro Person pro Tag enthält, die in  $Q_1$  bzw.  $\hat{Q}_1$  einfließen:

$$g^{\mathcal{Q}_1} := \left( g_1^{\mathcal{K}}, \dots, g_{N_{\mathcal{K}}}^{\mathcal{K}}, g_1^{\bar{\mathcal{K}}}, \dots, g_{N_{\bar{\mathcal{K}}}}^{\bar{\mathcal{K}}} \right)^\top \quad \text{bzw.} \quad (6.20)$$

$$\hat{g}^{\mathcal{Q}_1} := \left( \hat{g}_1^{\mathcal{K}}, \dots, \hat{g}_{N_{\mathcal{K}}}^{\mathcal{K}}, \hat{g}_1^{\bar{\mathcal{K}}}, \dots, \hat{g}_{N_{\bar{\mathcal{K}}}}^{\bar{\mathcal{K}}} \right)^\top. \quad (6.21)$$

Es sei  $\hat{\beta}^g$  der Parameterschätzer des GAM zur Bestimmung der Gesundheitsausgaben pro Versichertentag (GAM aus Abschnitt 6.7.1). Seine zugehörige Kovarianzmatrix sei mit  $\hat{V}_{\beta^g}$  bezeichnet. Um  $g^{\mathcal{Q}_1}$  zu schätzen, wird eine Prädiktionsmatrix  $X_{\text{Präd}}^{g^{\mathcal{Q}_1}}$  verwendet (siehe Wood 2006), die Zeilen für jede Morbiditätskombination der entsprechenden Alters-Geschlechtsklasse enthält. Es werden der lineare Prädiktor  $\hat{\eta}^{g^{\mathcal{Q}_1}}$  und dessen Kovarianzmatrix  $\hat{V}_{\eta^{g^{\mathcal{Q}_1}}}$  bestimmt:

$$\hat{\eta}^{g^{\mathcal{Q}_1}} = X_{\text{Präd}}^{g^{\mathcal{Q}_1}} \cdot \hat{\beta}^g, \quad (6.22)$$

$$\hat{V}_{\eta^g \mathcal{Q}_1} = X_{\text{Präd}}^{g \mathcal{Q}_1} \cdot \hat{V}_{\beta^g} \cdot \left( X_{\text{Präd}}^{g \mathcal{Q}_1} \right)^\top. \quad (6.23)$$

Der Vektor  $\hat{g}^{\mathcal{Q}_1}$  wird bestimmt, indem die inverse Linkfunktion angewendet wird:

$$\hat{g}^{\mathcal{Q}_1} = \exp\left(\hat{\eta}^{g \mathcal{Q}_1}\right). \quad (6.24)$$

Die Ableitung der Inversen der Linkfunktion wird verwendet, um die Kovarianzmatrix  $\hat{V}_{g \mathcal{Q}_1}$  von  $\hat{g}^{\mathcal{Q}_1}$  zu schätzen:

$$\hat{V}_{g \mathcal{Q}_1}[i, j] = \exp(\eta_i^{g \mathcal{Q}_1}) \cdot \exp(\eta_j^{g \mathcal{Q}_1}) \cdot \hat{V}_{\eta^g \mathcal{Q}_1}[i, j] \quad (6.25)$$

für alle  $i, j \in \{1, \dots, N_{\mathcal{K}} + N_{\overline{\mathcal{K}}}\}$ . Es wird angenommen, dass die Anzahl der Versicherungstage und die Gesundheitsausgaben unkorreliert sind:

$$\text{Cov}\left(\hat{g}_i^{\mathcal{Q}_1}, \mathcal{T}_j^{\mathcal{Q}_1}\right) = 0 \quad \forall i, j \in \{1, \dots, N_{\mathcal{K}} + N_{\overline{\mathcal{K}}}\}. \quad (6.26)$$

Daher sind mit den Formeln 6.14 bis 6.17 alle Ableitungen und mit den Formeln 6.19, 6.25 und 6.26 alle Varianzen und Kovarianzen bestimmt, um das Fehlerfortpflanzungsgesetz direkt anzuwenden und die Varianz von  $\hat{\mathcal{Q}}_1$  abzuschätzen. Für jede Altersgruppe und beide Geschlechter werden die Ergebnisse in Abbildung 6.11 sowie in Tabellen B.3 und B.4 (Anhang) dargestellt.

Werden die unbekannt Parameter in Formel 6.12 zur Quantifizierung von  $\mathcal{Q}_2$  gemäß Formeln 6.8 und 6.10 ersetzt, so ergibt sich der Schätzer:

$$\hat{\mathcal{Q}}_2 = 1 - \left( 1 - \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \hat{\mathcal{P}}_i^{\mathcal{A}_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \hat{\mathcal{P}}_i^{\mathcal{K}}} \right) \cdot \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \hat{g}_i^{\mathcal{A}_2}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \hat{g}_i^{\mathcal{K}}} - \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \hat{\mathcal{P}}_i^{\mathcal{A}_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \hat{\mathcal{P}}_i^{\mathcal{K}}} \cdot \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \hat{g}_i^{\mathcal{A}_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \hat{g}_i^{\mathcal{K}}}. \quad (6.27)$$

Es sei  $\mathcal{Z} \in \{\hat{g}_j^{\mathcal{K}}, \hat{g}_j^{\mathcal{A}_1}, \hat{g}_j^{\mathcal{A}_2}, \hat{\mathcal{P}}_j^{\mathcal{K}}, \hat{\mathcal{P}}_j^{\mathcal{A}_1}, \mathcal{T}_j^{\mathcal{K}} : j = 1, \dots, N_{\mathcal{K}}\}$ . Dann lässt sich  $\hat{\mathcal{Q}}_2$  nach  $\mathcal{Z}$

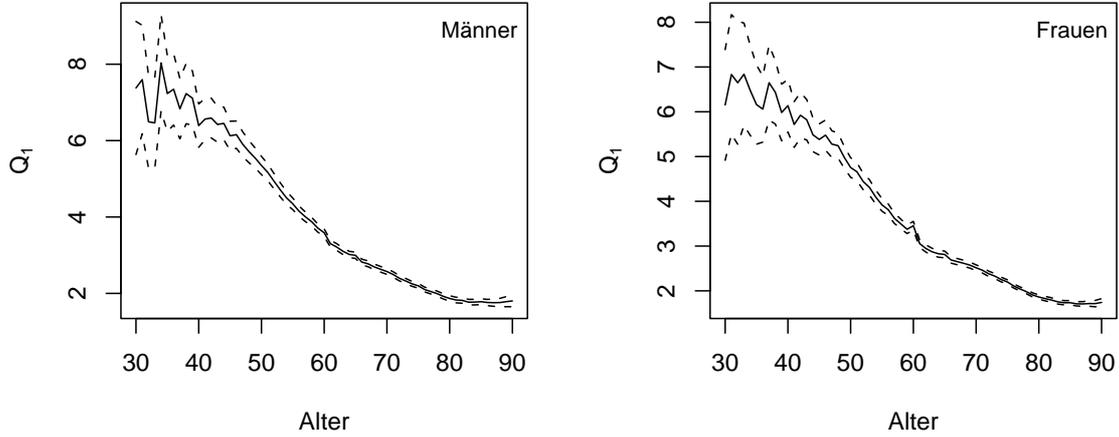


Abbildung 6.11:  $\hat{Q}_1$  plus/minus das 1,96-Fache der Standardabweichung

differenzieren:

$$\begin{aligned}
 \frac{\partial \hat{Q}_2}{\partial \mathcal{Z}} &= \frac{\partial \left( 1 - \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}}} \right)}{\partial \mathcal{Z}} \cdot \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{A_2}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \\
 &+ \left( 1 - \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}}} \right) \cdot \frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{A_2}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \right)}{\partial \mathcal{Z}} \\
 &- \frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}}} \right)}{\partial \mathcal{Z}} \cdot \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \\
 &- \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}}} \cdot \frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \right)}{\partial \mathcal{Z}}. \tag{6.28}
 \end{aligned}$$

Die Formel 6.28 enthält auf der rechten Seite der Gleichung vier Ableitungen. Diese werden für alle in Frage kommenden  $\mathcal{Z}$  bestimmt. Die Ableitungen des ersten Terms lauten:

$$\frac{\partial \left( 1 - \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}}} \right)}{\partial T_j^{\mathcal{K}}} = - \frac{\hat{p}_j^{A_1} \cdot \left( \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}} \right) - \hat{p}_j^{\mathcal{K}} \cdot \left( \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1} \right)}{\left( \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}} \right)^2},$$

(6.29)

$$\frac{\partial \left( 1 - \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}}} \right)}{\partial \hat{p}_j^{\mathcal{K}}} = \frac{T_j^{\mathcal{K}} \cdot \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\left( \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}} \right)^2}, \quad (6.30)$$

$$\frac{\partial \left( 1 - \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}}} \right)}{\partial \hat{p}_j^{A_1}} = - \frac{T_j^{\mathcal{K}}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}}}, \quad (6.31)$$

$$\frac{\partial \left( 1 - \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}}} \right)}{\partial \mathcal{Z}} = 0 \quad \text{für} \quad \mathcal{Z} = \hat{g}_j^{\mathcal{K}}, \hat{g}_j^{A_1}, \hat{g}_j^{A_2}. \quad (6.32)$$

Für den zweiten abzuleitenden Term gilt:

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{A_2}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \right)}{\partial T_j^{\mathcal{K}}} = \frac{\hat{g}_j^{A_2} \cdot \left( \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}} \right) - \hat{g}_j^{\mathcal{K}} \cdot \left( \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{A_2} \right)}{\left( \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}} \right)^2}, \quad (6.33)$$

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{A_2}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \right)}{\partial \hat{g}_j^{\mathcal{K}}} = - \frac{T_j^{\mathcal{K}} \cdot \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{A_2}}{\left( \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}} \right)^2}, \quad (6.34)$$

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{A_2}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \right)}{\partial \hat{g}_j^{A_2}} = \frac{T_j^{\mathcal{K}}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}}, \quad (6.35)$$

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{A_2}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \right)}{\partial \mathcal{Z}} = 0 \quad \text{für} \quad \mathcal{Z} = \hat{g}_j^{A_1}, \hat{p}_j^{\mathcal{K}}, \hat{p}_j^{A_1}. \quad (6.36)$$

Für den dritten abzuleitenden Term gilt:

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}}} \right)}{\partial T_j^{\mathcal{K}}} = \frac{\hat{p}_j^{A_1} \cdot \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}} - \hat{p}_j^{\mathcal{K}} \cdot \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\left( \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}} \right)^2}, \quad (6.37)$$

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}}} \right)}{\partial \hat{p}_j^{\mathcal{K}}} = - \frac{T_j^{\mathcal{K}} \cdot \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{A_1}}{\left( \sum_{i=1}^{N_{\mathcal{K}}} T_i^{\mathcal{K}} \cdot \hat{p}_i^{\mathcal{K}} \right)^2}, \quad (6.38)$$

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{K}}} \right)}{\partial \hat{\mathcal{P}}_j^{A_1}} = \frac{\mathcal{T}_j^{\mathcal{K}}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{K}}}, \quad (6.39)$$

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{\mathcal{P}}_i^{\mathcal{K}}} \right)}{\partial \mathcal{Z}} = 0 \quad \text{für } \mathcal{Z} = \hat{g}_j^{\mathcal{K}}, \hat{g}_j^{A_1}, \hat{g}_j^{A_2}. \quad (6.40)$$

Für den vierten abzuleitenden Term gilt:

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \right)}{\partial \mathcal{T}_j^{\mathcal{K}}} = \frac{\hat{g}_j^{A_1} \cdot \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}} - \hat{g}_j^{\mathcal{K}} \cdot \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{A_1}}{\left( \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}} \right)^2}, \quad (6.41)$$

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \right)}{\partial \hat{g}_j^{\mathcal{K}}} = - \frac{\mathcal{T}_j^{\mathcal{K}} \cdot \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{A_1}}{\left( \sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}} \right)^2}, \quad (6.42)$$

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \right)}{\partial \hat{g}_j^{A_1}} = \frac{\mathcal{T}_j^{\mathcal{K}}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}}, \quad (6.43)$$

$$\frac{\partial \left( \frac{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{A_1}}{\sum_{i=1}^{N_{\mathcal{K}}} \mathcal{T}_i^{\mathcal{K}} \cdot \hat{g}_i^{\mathcal{K}}} \right)}{\partial \mathcal{Z}} = 0 \quad \text{für } \mathcal{Z} = \hat{\mathcal{P}}_j^{\mathcal{K}}, \hat{\mathcal{P}}_j^{A_1}, \hat{g}_j^{A_2}. \quad (6.44)$$

Gemäß Abschnitt 5.6.2, Formel 5.10 und Formel 5.32 gilt:

$$\mathcal{T}^{\mathcal{K}} = (\mathcal{T}_1^{\mathcal{K}}, \dots, \mathcal{T}_{N_{\mathcal{K}}}^{\mathcal{K}})^{\top} \quad \text{und} \quad (6.45)$$

$$\hat{V}_{\mathcal{T}^{\mathcal{K}}} = \text{diag} \left( \widehat{\text{Var}}(\mathcal{T}_1^{\mathcal{K}}), \dots, \widehat{\text{Var}}(\mathcal{T}_{N_{\mathcal{K}}}^{\mathcal{K}}) \right) = \text{diag} (365 \cdot \mathcal{T}^{\mathcal{K}}). \quad (6.46)$$

Weiter wird ein Vektor definiert, der alle geschätzten Gesundheitsausgaben pro Person pro Tag enthält, die in  $\mathcal{Q}_2$  bzw.  $\hat{\mathcal{Q}}_2$  einfließen:

$$g^{\mathcal{Q}_2} := (g_1^{\mathcal{K}}, \dots, g_{N_{\mathcal{K}}}^{\mathcal{K}}, g_1^{A_1}, \dots, g_{N_{\mathcal{K}}}^{A_1}, g_1^{A_2}, \dots, g_{N_{\mathcal{K}}}^{A_2})^{\top} \quad \text{bzw.} \quad (6.47)$$

$$\hat{g}^{\mathcal{Q}_2} := (\hat{g}_1^{\mathcal{K}}, \dots, \hat{g}_{N_{\mathcal{K}}}^{\mathcal{K}}, \hat{g}_1^{A_1}, \dots, \hat{g}_{N_{\mathcal{K}}}^{A_1}, \hat{g}_1^{A_2}, \dots, \hat{g}_{N_{\mathcal{K}}}^{A_2})^{\top}. \quad (6.48)$$

Um  $\hat{g}^{\mathcal{Q}_2}$  zu schätzen, wird eine Prädiktionsmatrix  $X_{\text{Präd}}^{g^{\mathcal{Q}_2}}$  verwendet. Es wird der lineare Prädiktor  $\hat{\eta}^{g^{\mathcal{Q}_2}}$  und dessen Kovarianzmatrix  $\hat{V}_{\eta^{g^{\mathcal{Q}_2}}}$  bestimmt:

$$\hat{\eta}^{g^{\mathcal{Q}_2}} = X_{\text{Präd}}^{g^{\mathcal{Q}_2}} \cdot \hat{\beta}^g \quad (6.49)$$

$$\hat{V}_{\eta^g \mathcal{Q}_2} = X_{\text{Präd}}^{g \mathcal{Q}_2} \cdot \hat{V}_{\beta^g} \cdot \left( X_{\text{Präd}}^{g \mathcal{Q}_2} \right)^\top \quad (6.50)$$

Anwendung der Inversen der Linkfunktion:

$$\hat{g}^{\mathcal{Q}_2} = \exp\left(\hat{\eta}^{g \mathcal{Q}_2}\right) \quad (6.51)$$

Kovarianzmatrix  $\hat{V}_{g \mathcal{Q}_2}$  von  $\hat{g}^{\mathcal{Q}_2}$ :

$$\hat{V}_{g \mathcal{Q}_2}[i, j] = \exp(\eta_i^{g \mathcal{Q}_2}) \cdot \exp(\eta_j^{g \mathcal{Q}_2}) \cdot \hat{V}_{\eta^g \mathcal{Q}_2}[i, j] \quad (6.52)$$

Es wird angenommen, dass die Versichertentage sowohl mit den Schätzern für die Kosten pro Tag als auch für die Wahrscheinlichkeiten einer Herzinsuffizienz unkorreliert sind, d. h. eine Kovarianz von 0 besitzen. Zusätzlich wird angenommen, dass die Kovarianzmatrix zwischen den geschätzten Wahrscheinlichkeiten einer Herzinsuffizienz und den geschätzten Kosten pro Tag 0 beträgt:

$$\text{Cov}(\hat{\mathcal{P}}^\delta, \mathcal{T}^\mathcal{K}) = 0, \quad (6.53)$$

$$\text{Cov}(\hat{g}^{\mathcal{Q}_2}, \mathcal{T}^\mathcal{K}) = 0, \quad (6.54)$$

$$\text{Cov}(\hat{\mathcal{P}}^\delta, \hat{g}^{\mathcal{Q}_2}) = 0. \quad (6.55)$$

Somit sind alle Ableitungen, Varianzen und Kovarianzen bestimmt, um das Fehlerfortpflanzungsgesetz direkt anwenden zu können, um die Varianz von  $\hat{\mathcal{Q}}_2$  abzuschätzen. Separat nach Alter und Geschlecht wird das Ergebnis der Varianzabschätzung in Abbildung 6.12 und in Tabellen B.5 und B.6 (Anhang) dargestellt.

### 6.7.3 Anwendung der Quotienten

Es seien  $\tilde{g}$  die aufzuteilenden Durchschnittskosten einer Alters-Geschlechtsklasse. Zunächst sollen diese auf die Subpopulationen der Personen mit KHK und der Personen ohne KHK aufgeteilt werden. Die Durchschnittskosten  $\tilde{g}_\mathcal{K}$  der Personen mit KHK und die Durchschnittskosten  $\tilde{g}_{\bar{\mathcal{K}}}$  der Personen ohne KHK müssen insgesamt  $\tilde{g}$  ergeben:

$$\tilde{g} = \pi \cdot \tilde{g}_\mathcal{K} + (1 - \pi) \cdot \tilde{g}_{\bar{\mathcal{K}}}. \quad (6.56)$$

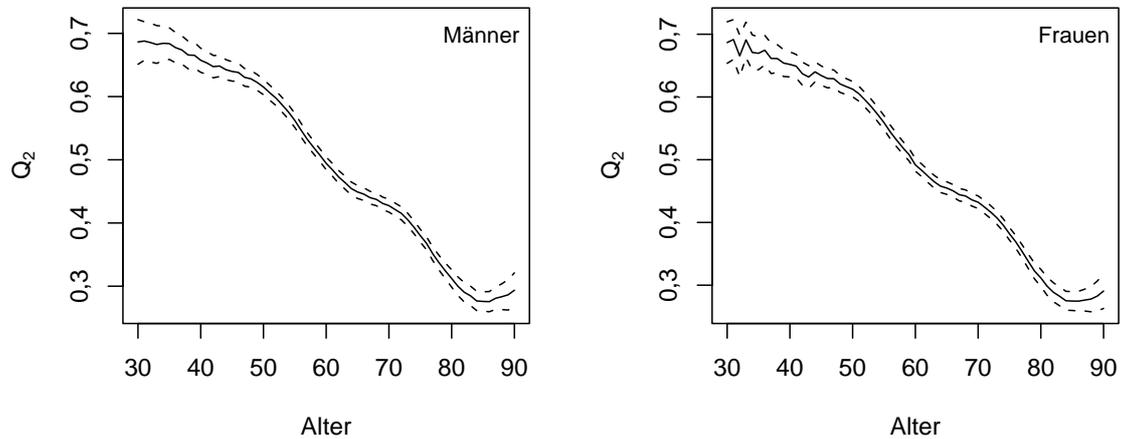


Abbildung 6.12:  $\hat{Q}_2$  plus/minus das 1,96-Fache der Standardabweichung

Außerdem muss der in Abschnitt 6.7.2 geschätzte Quotient  $Q_1$  eingehalten werden:

$$\begin{aligned} \hat{Q}_1 &= \frac{\tilde{g}_K}{\tilde{g}_K} \\ \Leftrightarrow \tilde{g}_K &= \hat{Q}_1 \cdot \tilde{g}_K. \end{aligned} \quad (6.57)$$

Wird Gleichung 6.57 in Gleichung 6.56 eingesetzt und nach  $\tilde{g}_K$  gelöst, so folgt:

$$\tilde{g}_K = \frac{\tilde{g}}{\pi \cdot \hat{Q}_1 + 1 - \pi}. \quad (6.58)$$

Die attributablen KHK-Kosten  $\tilde{g}_\Delta$  fallen für eine Person mit KHK weg, falls bei ihr die KHK durch Prävention vermieden wird. Um diese Kosten zu ermitteln wird der Schätzwert für  $Q_2$  verwendet:

$$\begin{aligned} \hat{Q}_2 &= \frac{\tilde{g}_\Delta}{\tilde{g}_K} \\ \Leftrightarrow \tilde{g}_\Delta &= \hat{Q}_2 \cdot \tilde{g}_K. \end{aligned} \quad (6.59)$$

## 6.8 Screeningkosten

Die Screeningkosten hängen vom jeweiligen Screeningverfahren ab. Ein Screening mittels der Framingham-Risikogleichung würde in etwa 60 bis 75 Euro kosten. Die Blutuntersuchungen würden dabei den geringsten Kostenanteil ausmachen. Der Hauptteil der Kosten würde durch den ärztlichen Zeitaufwand entstehen, der neben Blutdruckmessungen etc. auch ein Beratungsgespräch beinhaltet. In der gesundheitsökonomischen Analyse wird beispielhaft von Kosten von 70 Euro pro Screening ausgegangen. Dieser Wert wird im Rahmen der Sensitivitätsanalyse variiert.

## 6.9 Kosten der Statintherapie

Mit Wirkung zum 01.01.2005 liegen auch für patentgeschützte Statine Festbeträge vor (Klever-Deichert *et al.* 2006). Den Präventionskosten in der vorliegenden Modellierung liegen die durchschnittlichen Kosten (Festbeträge) aller am Markt verfügbaren Statine in der jeweils mittleren Dosierung zugrunde. Diese belaufen sich auf ca. 300 Euro pro Patient pro Jahr (Klever-Deichert *et al.* 2006). Werden zusätzlich zu den Medikationskosten die Kosten der behandelnden Ärzte berücksichtigt, so ist mit jährlichen Präventionskosten von ca. 400 Euro je Patient zu rechnen. Da bei der Wirksamkeit der Prävention von einer Compliance<sup>||</sup> von 90% ausgegangen wird, muss diese Compliance auch bei den Präventionskosten gelten. Deshalb werden im Markovmodell für die Zustände  $g^+$  und  $g^{V+}$  jährliche Präventionskosten von 360 Euro angesetzt.

---

<sup>||</sup>Bereitschaft eines Patienten zur aktiven Mitwirkung an therapeutischen Maßnahmen.

# Kapitel 7

## Sensitivität und Spezifität des Modells

### 7.1 Modell zur Bestimmung der Sensitivität und der Spezifität

Die Güte eines diagnostischen Tests in klinischen Studien wird in der Regel durch dessen Sensitivität und Spezifität angegeben. Unter der Sensitivität wird die Wahrscheinlichkeit verstanden, dass die betrachtete Krankheit erkannt wird (Testergebnis positiv), gegeben, dass sie tatsächlich vorliegt. Die Spezifität gibt die Wahrscheinlichkeit an, dass das Testergebnis negativ ist, gegeben, dass die Krankheit tatsächlich nicht vorliegt (Schumacher und Schulgen 2002).

Die Screeningverfahren, die in dieser Arbeit von Interesse sind, unterscheiden sich leicht von herkömmlichen diagnostischen Tests, da die Krankheit zum Screeningzeitpunkt noch nicht vorliegt. Das Ereignis „die Krankheit liegt tatsächlich vor“ wird durch das Ereignis „die Krankheit wird innerhalb eines vorgegebenen Zeitraumes eintreten“ ersetzt. Als Zeitraum werden 10 Jahre gewählt, da Framingham- und PROCAM-Risiken (Assmann *et al.* 2002) in der Regel als 10-Jahres-Risiken angegeben werden.

In klinischen Studien werden Sensitivität und Spezifität aus einer Vier-Felder-Tafel geschätzt. Die Spalten geben an, ob die Krankheit bei einer Person vorliegt, und die Zeilen geben an, ob der klinische Test positiv oder negativ ausgefallen ist. Für die hier zu untersuchenden Screeningverfahren kann eine ähnliche Vier-Felder-Tafel erstellt werden. Aus den Übergangswahrscheinlichkeiten des Markovmodells können die Anteile jeder Zelle der Vier-Felder-Tafel berechnet werden. Dabei geben die Randsummen der Zeilen an, welchem Anteil der Population zum Screeningzeitpunkt ein positiver bzw. negativer Score zugeordnet wurde. Wenn mit einer abgewandelten Form des Markovmodells die entsprechenden Zyklen berechnet werden, lässt sich bestimmen, wie hoch der Anteil der Population in jeder der vier Zellen ist. Aus diesen Anteilen lassen sich Sensitivität und Spezifität berechnen.

Das Markovmodell, mit dem die Sensitivität und die Spezifität bestimmt werden sollen, ist nicht eindeutig. Es muss festgelegt werden, wie mit Personen verfahren wird, die an einer anderen Ursache als der KHK sterben, bevor der vorgegebene Zeitraum abgelaufen ist. Eine Möglichkeit wäre es, sie als Personen ohne KHK zu betrachten, da bei ihnen die KHK nicht aufgetreten ist. Diese Möglichkeit wird nicht gewählt, da sie den gängigsten KHK-Risikoscreeningverfahren wie der Framingham-Risikogleichung widerspricht: In der Analyse von Überlebenszeiten ist es üblich, Personen die vorzeitig aus einer Studie ausscheiden, zu zensieren. Ist als Endpunkt einer Studie z. B. das Eintreten der KHK festgelegt worden, so wird eine Person zensiert, falls sie vor Eintritt der KHK die Teilnahme an der Studie abbricht oder an einer anderen Ursache als der KHK verstirbt. Diese Art der Zensierung wurde auch bei der Framingham-Studie durchgeführt. Da die Framinghamrisiken eine zentrale Rolle in dieser Arbeit einnehmen, wird hier eine entsprechende Vorgehensweise gewählt. Das 10-Jahres-Risiko gemäß der Framingham-Risikogleichung besagt, wie hoch die Wahrscheinlichkeit ist, dass eine Person innerhalb der nächsten zehn Jahre an der KHK erkrankt, *gegeben*, dass diese Person nicht zuvor an einer anderen Ursache als der KHK verstirbt.

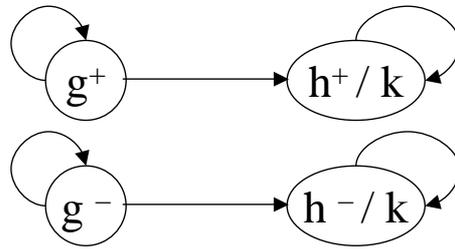


Abbildung 7.1: Verändertes Markovmodell zwecks Ermittlung der Sensitivität und Spezifität

Aus diesem Grund wird folgendes Modell gewählt (siehe Abbildung 7.1): Zum Zeitpunkt  $t_0$  befindet sich die zu screenende Population in den Zuständen  $g^+$  und  $g^-$ . Erkrankten Personen innerhalb eines Zyklus nicht, so verbleiben sie in  $g^+$  bzw.  $g^-$ . Der Eintritt der KHK kann sowohl „fatal“ als auch „nicht fatal“ erfolgen\*. Deshalb werden zwei neue Zustände  $h^+/k$  und  $h^-/k$  definiert: Bei Eintritt der KHK wechseln Personen von  $g^+$  nach  $h^+/k$  bzw. Personen von  $g^-$  nach  $h^-/k$ . Die benötigten Übergangswahrscheinlichkeiten lassen sich aus denen des zugrundeliegenden Modells (siehe Kapitel 4) berechnen:

$$p_{g^+ h^+/k} := \frac{p_{g^+ h^+} + p_{g^+ k}}{1 - p_{g^+ i} - p_{g^+ t}}, \quad (7.1)$$

$$p_{g^- h^-/k} := \frac{p_{g^- h^-} + p_{g^- k}}{1 - p_{g^- i} - p_{g^- t}}. \quad (7.2)$$

Die Übergangswahrscheinlichkeit von  $g^+$  zu sich selbst wird in diesem Modell mit  $p_{g^+ g^+}^S$  bezeichnet. Analog heißt die Übergangswahrscheinlichkeit von  $g^-$  zu sich selbst  $p_{g^- g^-}^S$ . Der Zusatz „S“ wird in den Namen eingefügt, um Verwechslungen analog bezeichneter Größen des Modells aus Abschnitt 4.4 zu vermeiden. Es gilt:

$$p_{g^+ g^+}^S = 1 - p_{g^+ h^+/k}, \quad (7.3)$$

$$p_{g^- g^-}^S = 1 - p_{g^- h^-/k}. \quad (7.4)$$

---

\*Ein „fatales“ KHK-Ereignis bedeutet, dass die Person an der KHK stirbt. Somit wechselt sie in den Subzustand  $k$  (KHK-Tod) des Zustands  $T$  (Tod).

	KHK tritt ein	KHK tritt nicht ein	Gesamt
Score positiv	$a_{g^0,s}^{h^+/k}(t_{10})$	$a_{g^0,s}^{g^+}(t_{10})$	$a_g^{g^+}(t_0)$
Score negativ	$a_{g^0,s}^{h^-/k}(t_{10})$	$a_{g^0,s}^{g^-}(t_{10})$	$a_g^{g^-}(t_0)$
Gesamt	$a_{g^0,s}^{h^+/k}(t_{10}) + a_{g^0,s}^{h^-/k}(t_{10})$	$a_{g^0,s}^{g^+}(t_{10}) + a_{g^0,s}^{g^-}(t_{10})$	1

Tabelle 7.1: Vier-Felder-Tafel der Bevölkerungsanteile zur Berechnung der (10-Jahres-)Sensitivität und der (10-Jahres-)Spezifität des Screeningverfahrens bei gegebenem Markovmodell

Um die Sensitivität und die Spezifität berechnen zu können, werden die Einträge der Vier-Felder-Tafel in Tabelle 7.1 gesucht. Die Einträge  $a_g^{g^+}(t_0)$  und  $a_g^{g^-}(t_0)$  sind bekannt, die übrigen Einträge werden folgendermaßen berechnet:

$$\begin{aligned} a_{g^0,s}^{g^+}(t_{10}) &= a_g^{g^+}(t_0) \cdot p_{g^+g^+}^S(t_0) \cdot \dots \cdot p_{g^+g^+}^S(t_9) \\ &= a_g^{g^+}(t_0) \cdot \prod_{i=1}^{10} p_{g^+g^+}^S(t_{i-1}), \end{aligned} \quad (7.5)$$

$$a_{g^0,s}^{g^-}(t_{10}) = a_g^{g^-}(t_0) \cdot \prod_{i=1}^{10} p_{g^-g^-}^S(t_{i-1}), \quad (7.6)$$

$$a_{g^0,s}^{h^+/k}(t_{10}) := a_g^{g^+}(t_0) - a_{g^0,s}^{g^+}(t_{10}), \quad (7.7)$$

$$a_{g^0,s}^{h^-/k}(t_{10}) := a_g^{g^-}(t_0) - a_{g^0,s}^{g^-}(t_{10}). \quad (7.8)$$

Die Sensitivität und die Spezifität werden unter Benutzung dieser Werte folgendermaßen berechnet:

$$\text{Sensitivität} = \frac{a_{g^0,s}^{h^+/k}(t_{10})}{a_{g^0,s}^{h^+/k}(t_{10}) + a_{g^0,s}^{h^-/k}(t_{10})}, \quad (7.9)$$

$$\text{Spezifität} = \frac{a_{g^0,s}^{g^-}(t_{10})}{a_{g^0,s}^{g^+}(t_{10}) + a_{g^0,s}^{g^-}(t_{10})}. \quad (7.10)$$

Es sei angemerkt, dass sich die Ermittlung der Sensitivität und der Spezifität auf den Fall beziehen, dass keine Prävention durchgeführt wird. Im Szenario mit Prävention erkranken weniger Personen mit hohem Score.

## 7.2 Steuerung der Sensitivität und der Spezifität im Markovmodell

Anhand des Markovmodells sollen unterschiedliche Szenarien miteinander verglichen werden. Zu diesen gehören das Szenario ohne Prävention sowie eine Vielzahl unterschiedlicher Präventionsszenarien. Die Präventionsszenarien können sich durch die Wahl der Screeningmaßnahme sowie durch die Wahl des Trennpunktes unterscheiden. Durch den Trennpunkt wird festgelegt, welche Personen einem hohen oder niedrigen Risiko zugeordnet werden. Jedes Präventionsszenario wird durch dessen Kombination aus Sensitivität und Spezifität charakterisiert. Das Markovmodell muss mit unterschiedlichen Kombinationen der Parameter  $a_g^{g^+}(t_0)$  und  $\gamma_0$  evaluiert werden, um unterschiedliche Sensitivität-Spezifität-Kombinationen zu generieren. Das bedeutet, es wird erstens der Anteil  $a_g^{g^+}(t_0)$  der Personen variiert, die als Personen mit höherem Risiko eingestuft werden. Zweitens wird ein Odds Ratio ( $\gamma_0$ ) variiert. Dieses gibt an, wieviel größer das Risiko an der KHK zu erkranken für eine Person mit hohem Score im Vergleich zu einer Person mit niedrigem Score ist. Mit dem Anteil  $a_g^{g^+}(t_0)$  kann die Sensitivität auf Kosten der Spezifität erhöht werden und umgekehrt. Ist  $a_g^{g^+}(t_0)$  gleich 1, so werden alle Personen als Personen mit hohem Risiko eingestuft, und es ergeben sich eine Sensitivität von 1 und eine Spezifität von 0. Wird hingegen  $a_g^{g^+}(t_0)$  gleich 0 gewählt, so ergeben sich eine Sensitivität von 0 und eine Spezifität von 1. Mit dem Parameter  $\gamma_0$  kann die Qualität des Screeningverfahrens reguliert werden. Steigt  $\gamma_0$ , so steigen sowohl Sensitivität als auch Spezifität. Geht  $\gamma_0$  gegen unendlich, so gehen Sensitivität und Spezifität gegen 1. Ein Screeningverfahren, das rein zufällig Personen in eine Hochrisikogruppe und Personen mit geringem Risiko aufteilt, entspricht einem  $\gamma_0$  von 1. Um eine exogen vorgegebene „Receiver-Operating Characteristic“ (ROC)-Kurve (siehe Abschnitt 7.3) mit dem Markov-Modell zu erreichen, ist sowohl die Variation von  $a_g^{g^+}(t_0)$  als auch von  $\gamma_0$  erforderlich.

## 7.3 Berechnung der Sensitivität und der Spezifität anhand der Framinghamrisiken

In der Regel wird ein diagnostischer Test auf der Basis einer deterministischen Größe, z. B. der Konzentration eines Proteins im Blut, erstellt. Die durch ein Screeningverfahren ermöglichten Sensitivität-Spezifität-Kombinationen können durch ROC-Kurven graphisch dargestellt werden. Dabei werden auf der horizontalen Achse der Wert  $(1 - \text{Spezifität})$  und auf der vertikalen Achse die Sensitivität abgetragen (siehe z. B. Abbildung 7.2). Je weiter oberhalb die Kurve liegt, desto besser ist der klinische Test. Der Punkt  $\{0; 1\}$  (die obere linke Ecke) repräsentiert den Fall, dass jede Zuordnung richtig ist. Die Einheitsgerade entspricht dem Fall, dass zufällig entschieden wird, ob der Test positiv oder negativ ausfällt.

Mit Hilfe der Framingham-Risikogleichung können ROC-Kurven für 10-Jahres-Risiken erstellt werden. Dazu werden für alle Personen ohne KHK des Datensatzes des Bundesgesundheitsurvey 1998 die 10-Jahres-Risiken gemäß der Risikogleichung von Anderson *et al.* (1991) berechnet. Bei der Berechnung werden einzelne Kohorten betrachtet und für diese werden die ROC-Kurven erstellt. Dabei wird folgendermaßen vorgegangen:

Es sei  $\check{n}$  die Anzahl der Personen in der Kohorte. Weiter bezeichne  $p_i, i = 1, \dots, \check{n}$ , die Erkrankungswahrscheinlichkeit der  $i$ -ten Person. Der Trennpunkt wird mit  $p^{\text{cut}}$  bezeichnet. Weiter sei  $\check{n}^+$  die Anzahl der Personen mit positivem Score und  $\check{n}^-$  die Anzahl der Personen mit negativem Score. Diese Werte werden folgendermaßen ermittelt:

$$\check{n}^+ = \sum_{i=1}^{\check{n}} \mathcal{I}_{(p^{\text{cut}}, 1]}(p_i) \quad \text{und} \quad (7.11)$$

$$\check{n}^- = \sum_{i=1}^{\check{n}} \mathcal{I}_{[0; p^{\text{cut}})}(p_i), \quad (7.12)$$

	KHK tritt ein	KHK tritt nicht ein	Gesamt
Score positiv	$p^+ \cdot a^+$	$(1 - p^+) \cdot a^+$	$a^+$
Score negativ	$p^- \cdot a^-$	$(1 - p^-) \cdot a^-$	$a^-$
Gesamt	$p^+ \cdot a^+ + p^- \cdot a^-$	$(1 - p^+) \cdot a^+ + (1 - p^-) \cdot a^-$	1

Tabelle 7.2: Vier-Felder-Tafel der Anteile an der Kohorte zur Berechnung der Sensitivität und der Spezifität der Framingham-Risikogleichung

wobei  $\mathcal{I}$  die Indikatorfunktion bezeichnet.

Die durchschnittliche Erkrankungswahrscheinlichkeit der Personen mit positivem Score wird mit  $p^+$ , bei Personen mit negativem Score mit  $p^-$  bezeichnet. Es gilt:

$$p^+ = \frac{1}{\check{n}^+} \cdot \sum_{i=1}^{\check{n}} p_i \cdot \mathcal{I}_{(p^{\text{cut}}, 1]}(p_i), \quad (7.13)$$

$$p^- = \frac{1}{\check{n}^-} \cdot \sum_{i=1}^{\check{n}} p_i \cdot \mathcal{I}_{[0; p^{\text{cut}}]}(p_i). \quad (7.14)$$

Weiter sei  $a^+ = \check{n}^+/\check{n}$  der Anteil der Personen der Kohorte mit positivem Score und  $a^- = \check{n}^-/\check{n}$  der Anteil der Personen mit negativem Score. Es ergibt sich die Vier-Felder-Tafel in Tabelle 7.2. Die Sensitivität und Spezifität werden folgendermaßen berechnet:

$$\text{Sensitivität} = \frac{p^+ \cdot a^+}{p^+ \cdot a^+ + p^- \cdot a^-}, \quad (7.15)$$

$$\text{Spezifität} = \frac{(1 - p^-) \cdot a^-}{(1 - p^+) \cdot a^+ + (1 - p^-) \cdot a^-}. \quad (7.16)$$

Es seien  $p_1^{\text{Fram}}, \dots, p_k^{\text{Fram}}$  die für das jeweils betrachtete Personenkollektiv vorkommenden aufsteigend sortierten Framingham-10-Jahres-Risiken ohne Doppelungen (Es kann vorkommen, dass mehrere Personen dasselbe 10-Jahres-Risiko besitzen).

Es werden die folgenden Trennpunkte verwendet:

$$\frac{0 + p_1^{\text{Fram}}}{2}, \frac{p_1^{\text{Fram}} + p_2^{\text{Fram}}}{2}, \dots, \frac{p_{k-1}^{\text{Fram}} + p_k^{\text{Fram}}}{2} \quad \text{und} \quad \frac{p_k^{\text{Fram}} + 1}{2}. \quad (7.17)$$

Als kleinster und größter Trennpunkt werden nicht die Werte 0 und 1 verwendet, da in Kapiteln 8 und 9 eine Transformation mittels der logistischen Funktion erfolgt. Für jeden Trennpunkt werden Sensitivität und Spezifität berechnet. Um die verbleibenden Punkte der ROC-Kurve zu bestimmen, werden diese Punkte linear interpoliert. Des Weiteren wird mittels linearer Interpolation jeder (im späteren Verlauf der Berechnungen) explizit vorgegebenen Sensitivität ein Trennpunkt zugewiesen.

## 7.4 Analyse der Sensitivität und Spezifität aus Framinghamrisiken

Die in dieser Arbeit betrachtete Sensitivität und Spezifität von KHK-Screeningmaßnahmen bezieht sich auf Personen einer Kohorte desselben Alters und desselben Geschlechts. Bei Übertragung der Sensitivität und der Spezifität eines Screeningverfahrens auf das Markovmodell muss dies berücksichtigt werden. Die Alters- und Geschlechtsverteilung einer Kohorte beeinflusst in hohem Maße den Verlauf einer ROC-Kurve, wie im Folgenden am Beispiel der Framinghamrisiken demonstriert wird.

In Abbildung 7.2 wird beispielhaft für 45-jährige Personen veranschaulicht, dass bei gleichaltrigen Männern und Frauen die Qualität des Screeningverfahrens erkennbar voneinander abweicht. Die ROC-Kurve 45-jähriger Frauen liegt deutlich oberhalb der Kurve 45-jähriger Männer. Bei Frauen scheinen Framinghamrisiken aussagekräftiger zu sein als bei gleichaltrigen Männern<sup>†</sup>.

Das Alter hat ebenfalls einen deutlichen Effekt auf die Güte des Framingham-

---

<sup>†</sup>Abbildungen anderer Jahrgänge sehen ähnlich aus. Dass die Fläche unter der Kurve („area under the curve“, AUC) bei Frauen signifikant größer ist, als bei gleichaltrigen Männern, lässt sich mit Hilfe eines Regressionsmodells zeigen, wird hier aber nicht weiter verfolgt.

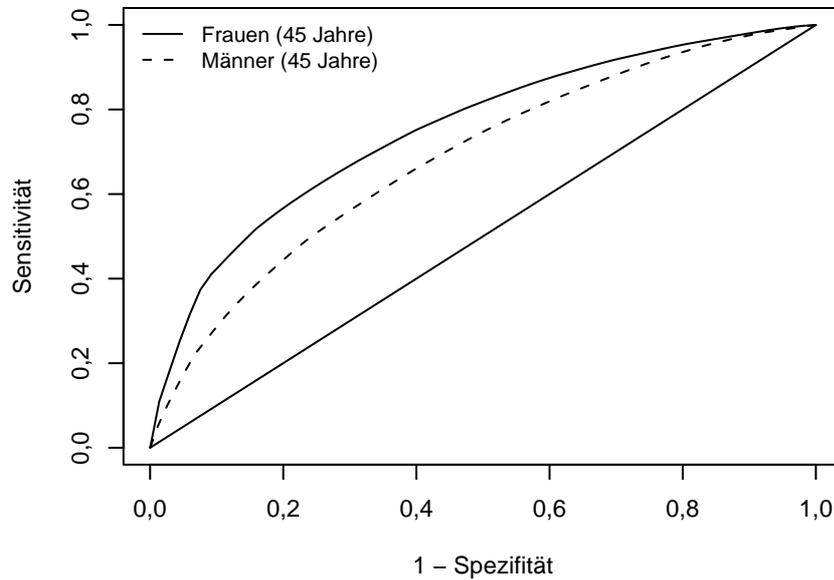


Abbildung 7.2: Aus den Framinghamrisiken der BGS-Daten ermittelte ROC-Kurven: Männer und Frauen im Vergleich

risikoscores, wie in Abbildung 7.3 für Männer gezeigt wird. Es werden fünf Geburtsjahrgänge mit jeweils gleichem Abstand zueinander ausgewählt. Die Altersspanne von 30 bis 74 Jahren wird gewählt, da es sich bei diesem Bereich um die Altersspanne handelt, auf die die Framingham-Risikogleichung angewendet werden darf (Anderson *et al.* 1990). Es ist erkennbar, dass die ROC-Kurve bei zunehmendem Alter sinkt. Daher ist es wichtig, neben dem Geschlecht auch nach dem Alter zu differenzieren.

Selbst wenn die Unterschiede der alters- und geschlechtsspezifischen ROC-Kurven als vernachlässigbar angesehen werden könnten, wäre es wichtig, die ROC-Kurven einzelner Alters-Geschlechtsklassen zu verwenden. Der Grund dafür wird durch einen Blick auf Abbildung 7.4 deutlich: Das Alter hat einen hohen Erklärungswert bezüglich der Wahrscheinlichkeit, an der koronaren Herzkrankheit zu erkranken. So liegt die ROC-Kurve, die alle Männer im Alter zwischen 30 und 74 Jahren umfasst,

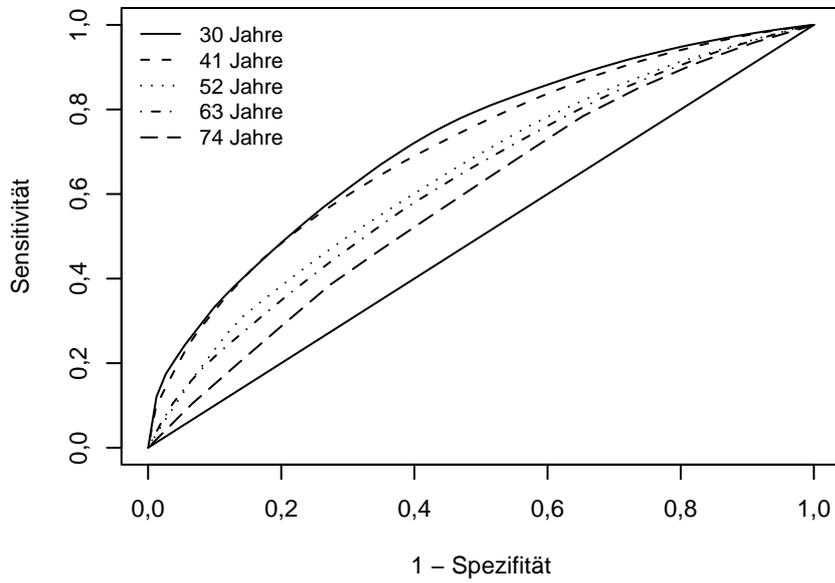


Abbildung 7.3: Aus den Framinghamrisiken der BGS-Daten ermittelte ROC-Kurven:  
Die Güte des Screenings mittels der Formel von Anderson sinkt mit dem Alter

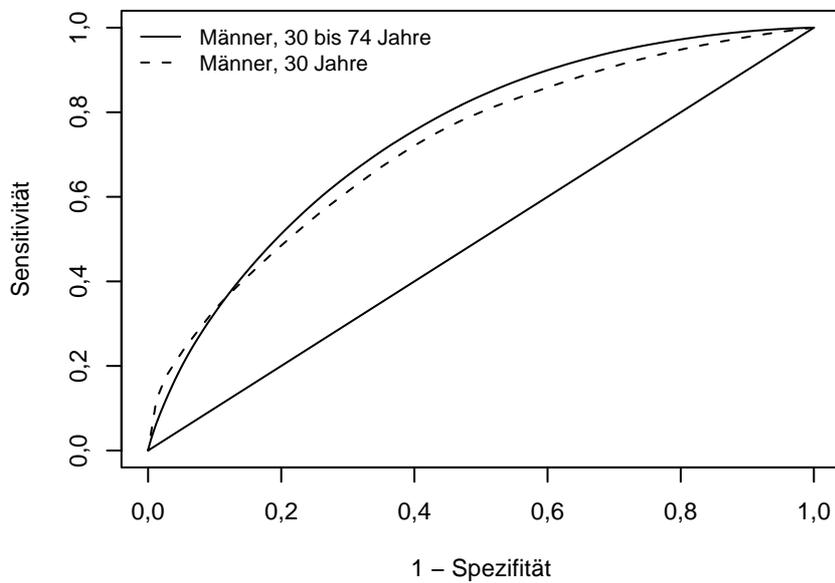


Abbildung 7.4: Aus den Framinghamrisiken der BGS-Daten ermittelte ROC-Kurven:  
Das Alter hat einen hohen Erklärungswert

zum größten Teil überhalb der besten Kurve, die erreicht werden kann, wenn nur ein einzelner Jahrgang zur Ermittlung der Sensitivität und Spezifität verwendet wird.

# Kapitel 8

## Durchführung der Kosten-Nutzwert-Analyse

### 8.1 Evaluation des Markovmodells

Das Markovmodell wird kohortenspezifisch evaluiert. Dabei besteht jede der Kohorten aus Personen gleichen Alters und Geschlechts. Sowohl für Männer als auch für Frauen werden für Kohorten der Screeningalter 30, 45 und 60 Jahre beispielhaft Analysen durchgeführt. Bei der Evaluierung des Markovmodells werden die Zyklen bis zum Alter von 95 Jahren berechnet. Die Kosten und der Nutzen höherer Jahrgänge werden als vernachlässigbar eingestuft. Für die über 90-Jährigen liegen nicht alle Inputgrößen, d. h. Größen zur Berechnung der Übergangswahrscheinlichkeiten und der Kosten, vor. Deshalb werden für diese Altersgruppen vereinfachend die Übergangswahrscheinlichkeiten und Auszahlungen der Gruppe der 90-Jährigen übernommen. Da die Übergänge des Markovmodells in der Mitte eines jeweiligen Zyklus erfolgen, wird eine Halbzykluskorrektur durchgeführt. Gemäß den Empfehlungen von Drummond und Jefferson (1996) werden sowohl Kosten als auch Nutzen mit einer Rate von 5 % diskontiert. Die Berechnungen werden für eine Zahlungsbereitschaft von 50.000 Euro pro gewonnenem

QALY durchgeführt. Diese Zahlungsbereitschaft orientiert sich u. a. an den Empfehlungen von Goldman *et al.* (1996). Dort werden Maßnahmen unter 20.000 US\$ pro gewonnenem QALY oder pro gewonnenem Lebensjahr als „sehr kosteneffektiv“, Maßnahmen zwischen 20.000 US\$ und 40.000 US\$ als „verhältnismäßig kosteneffektiv“ und Maßnahmen zwischen 40.000 US\$ und 60.000 US\$ (pro gewonnenem QALY oder pro gewonnenem Lebensjahr) als „grenzwertig“ eingestuft.

Jede Parameterkombination  $\{a_g^{g^+}(t_0), \gamma_0\}$ ,  $a_g^{g^+}(t_0) \in (0, 1)$ ,  $\gamma_0 \in \mathbb{R}^+$  entspricht genau einer Kombination aus Sensitivität und Spezifität für die entsprechende Kohorte\*. Für ausgewählte Kombinationen von  $\{a_g^{g^+}(t_0), \gamma_0\}$  wird das Markovmodell evaluiert. In der Anfangsverteilung wird nur den Zuständen  $g^+$  und  $g^-$  positives Gewicht zugewiesen ( $a_g^{g^+}(t_0)$  für  $g^+$  und  $a_g^{g^-}(t_0)$  für  $g^-$ ). Die Ergebnisse der Evaluation werden verwendet, um für vorgegebene Kombinationen aus Sensitivität und Spezifität die zugehörigen diskontierten Kosten und Nutzen zu bestimmen. Die Parameterkombinationen  $\{a_g^{g^+}(t_0), \gamma_0\}$  werden so gewählt, dass die zugehörigen Kombinationen aus Sensitivität und Spezifität einen großen Bereich potentieller ROC-Kurven abdecken. Für die folgenden Kombinationen von  $a_g^{g^+}(t_0)$  und  $\gamma_0$  wurde das Markovmodell evaluiert:

1. Für Kombinationen auf einem Parametergitter. Dabei nimmt der Parameter  $a_g^{g^+}(t_0)$  fünf unterschiedliche Ausprägungen mit gleichen Abständen zwischen 0,001 und 0,999 an. Der Parameter  $\gamma_0$  nimmt fünf Werte von 1 bis 200 an. Bei  $\gamma_0$  wurden die Abstände so gewählt, dass diese bezüglich der logarithmierten  $\gamma_0$  im gleichen Abstand zueinander liegen.
2. Es gibt nur ein  $a_g^{g^+}(t_0)$  bei dem sowohl die Sensitivität als auch die Spezifität

---

\*Der Parameter  $a_g^{g^+}(t_0)$  kann im Extremfall auch die Werte 0 und 1 annehmen. Dann sind die Gleichungssysteme zur Unterteilung des Modells nach erhöhtem bzw. geringem Risiko jedoch nicht mehr lösbar. In diesem Fall sind die Übergangswahrscheinlichkeiten der Personengruppe („+“ oder „-“), in der sich alle Personen befinden identisch mit den Übergangswahrscheinlichkeiten des Modells ohne Unterteilung nach Risikoscore.

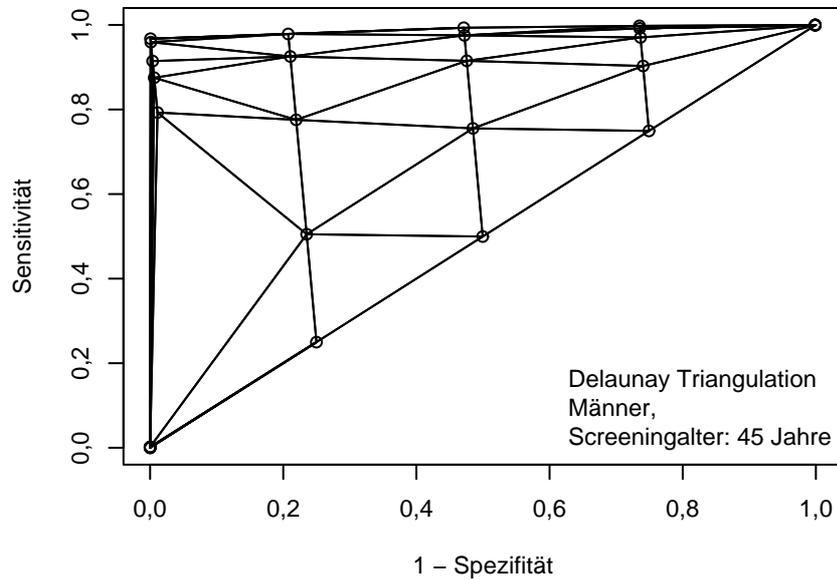


Abbildung 8.1: Delaunay Triangulation

bei steigenden Werten von  $\gamma_0$  gegen 1 streben. Dies ist der Anteil der Personen, die tatsächlich innerhalb der nächsten zehn Jahre an der KHK erkranken, gegeben, dass sie nicht zuvor an einer sonstigen Krankheit versterben. Dieser Anteil wird für jede Kohorte ermittelt. Zu diesem  $a_g^{g^+}(t_0)$  werden die Werte 200, 500, 1000, 5000 und 10000 für  $\gamma_0$  eingesetzt.

3. Zusätzlich werden die diskontierten Kosten und Nutzen für den Fall bestimmt, dass keine Prävention durchgeführt wird. In diesem Fall ist  $a_g^{g^+}(t_0)$  gleich 0, der Wert für  $\gamma_0$  ist irrelevant.

Zu den gegebenen Kombinationen von  $a_g^{g^+}(t_0)$  und  $\gamma_0$  werden die Sensitivität und die Spezifität gemäß der Definitionen in Kapitel 7 bestimmt. In Anlehnung an die ROC-Kurven wird anstelle der Spezifität der Wert  $(1 - \text{Spezifität})$  betrachtet. Es wird die Delaunay-Triangulation der Kombinationen aus Sensitivität und  $(1 - \text{Spezifität})$  bestimmt. Diese wird beispielhaft für 45-jährige Männer in Abbildung 8.1 dargestellt. Für die Eckpunkte eines Dreiecks der Delaunay Triangulation sind die

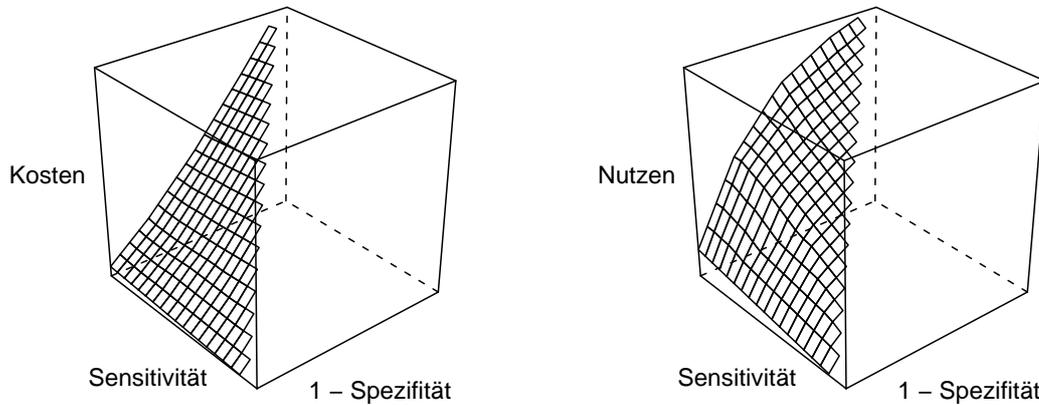


Abbildung 8.2: Verlauf der diskontierten Kosten und Nutzen in Abhängigkeit von der Sensitivität und 1 minus der Spezifität

diskontierten Kosten und der diskontierte Nutzen bekannt. Für die Sensitivität-Spezifität-Kombinationen, die nicht auf einem der Eckpunkte liegen, wird sowohl für die Kosten als auch für den Nutzen eine Ebene durch die drei Punkte im dreidimensionalen Raum gelegt. Durch diese beiden Ebenen werden die gesuchten Kosten und Nutzen approximiert. Beispielhaft wird der dreidimensionale Verlauf von Kosten und Nutzen für 45-jährige Männer in Abbildung 8.2 dargestellt. Zugehörige Iso-Kosten-Kurven (Kurven mit konstanten Kosten) befinden sich in Abbildung 8.3 und zugehörige Iso-Nutzen-Kurven in Abbildung 8.4.

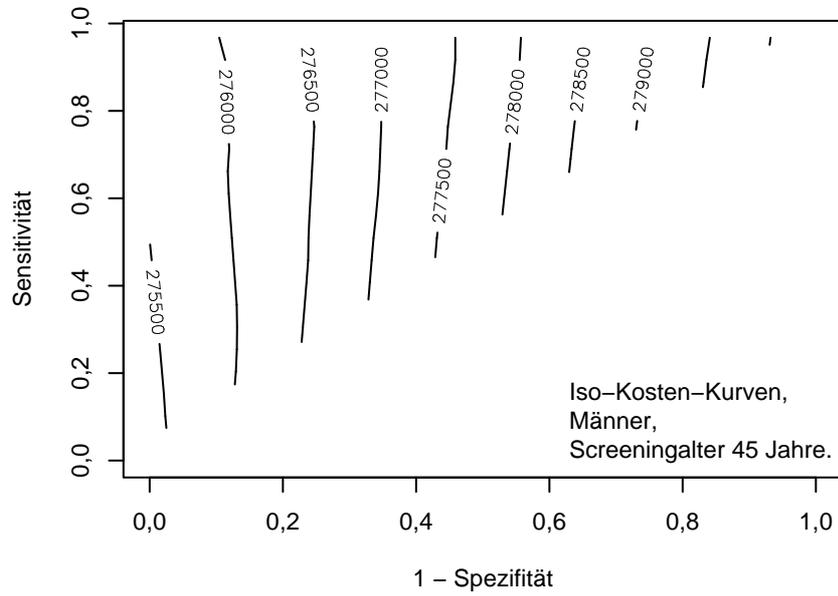


Abbildung 8.3: Kurven mit konstanten Kosten

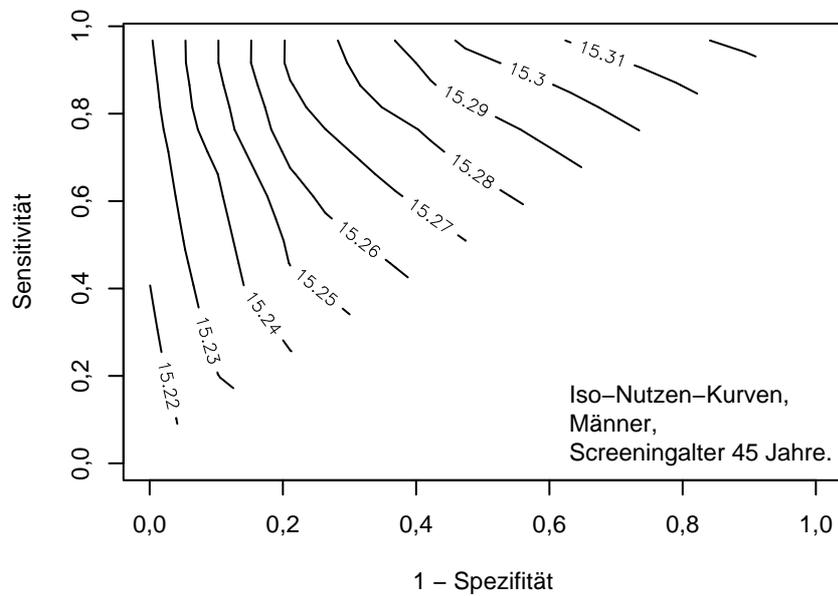


Abbildung 8.4: Kurven mit konstantem Nutzen

## 8.2 Kann sich Primärprävention lohnen?

Für jede Kombination aus Sensitivität und Spezifität kann eine inkrementelle Kosten-Nutzwert-Analyse im Vergleich zum Szenario ohne Prävention durchgeführt werden. Da sich diese Analysen auf kein bestimmtes Screeningverfahren beziehen, werden keine Screeningkosten angesetzt. Es wird geprüft, inwiefern sich Primärprävention „lohnen“ kann, wenn der Spezialfall eintritt, dass das Screening kostenlos ist. Dazu werden Kurven bestimmt, auf denen dieselbe Kosten-Nutzen-Relation im Vergleich zum Szenario ohne Prävention realisiert wird (siehe Abbildungen 8.5 bis 8.10).

Sensitivität-Spezifität-Kombinationen mit einer inkrementellen Kosten-Nutzen-Relation unterhalb der Zahlungsbereitschaft sind dem Szenario ohne Primärprävention vorzuziehen. Welche Sensitivität-Spezifität-Kombinationen bei einem Screening 30-jähriger Frauen „lohnenswert“ sein können, ist in Abbildung 8.5 dargestellt. Ausgehend von einer maximalen Zahlungsbereitschaft von 50.000 Euro pro QALY muss eine Sensitivität-Spezifität-Kombination auf einer verhältnismäßig hohen ROC-Kurve liegen und eine Sensitivität von mindestens 0,6 aufweisen, oder die Spezifität muss größer als 0,95 sein (entspricht „ $1 - \text{Spezifität} = 0,05$ “ in der Abbildung). Screeningmaßnahmen mit einer Spezifität von 0,7 oder niedriger scheinen sich bei 30-jährigen Frauen generell nicht zu lohnen.

Bei 30-jährigen Männern sieht die Situation ähnlich aus. Die Bereiche in denen sich ein Risikoscreening unter ökonomischen Gesichtspunkten lohnen, sind jedoch größer als bei gleichaltrigen Frauen (siehe Abbildung 8.6). Ein Screeningverfahren muss eine Spezifität von mindestens 0,4 aufweisen um sich auszuzahlen. Bei einer Spezifität von 0,95 muss ein Screeningverfahren nur noch eine Sensitivität von mindestens 0,4 aufweisen.

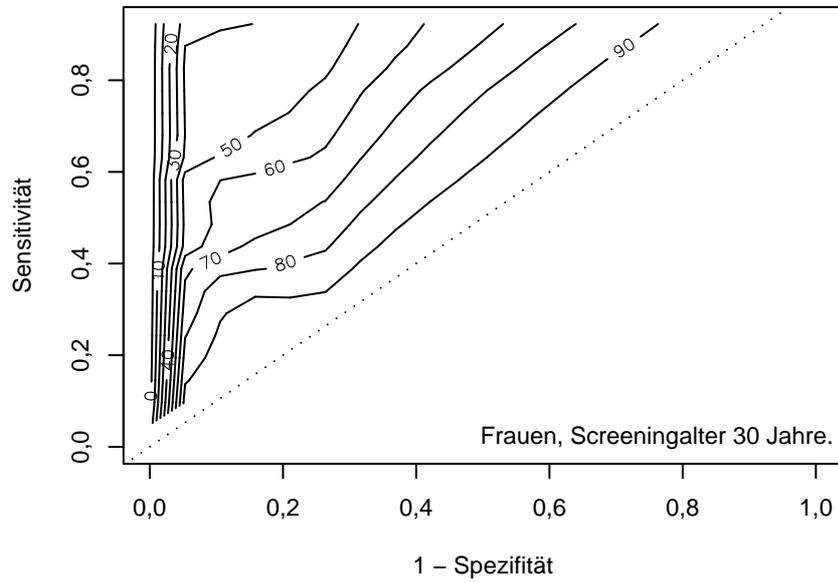


Abbildung 8.5: Kurven mit konstanter Kosten-Nutzen-Relation (in tausend Euro pro QALY) im Vergleich zum Szenario ohne Prävention; keine Screeningkosten

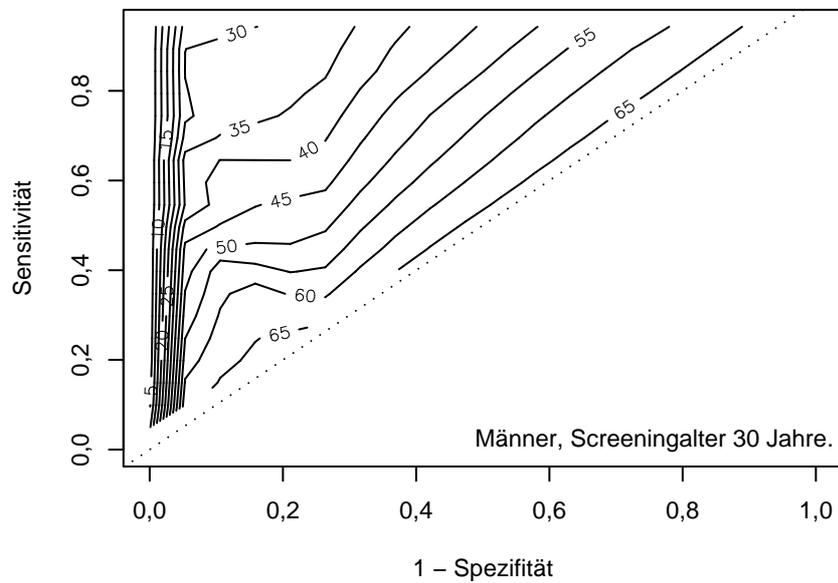


Abbildung 8.6: Kurven mit konstanter Kosten-Nutzen-Relation (in tausend Euro pro QALY) im Vergleich zum Szenario ohne Prävention; keine Screeningkosten

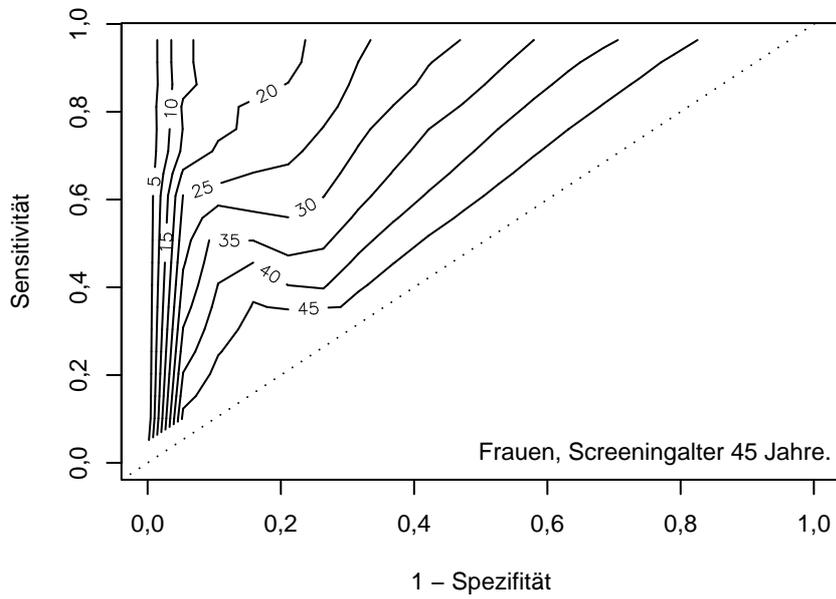


Abbildung 8.7: Kurven mit konstanter Kosten-Nutzen-Relation (in tausend Euro pro QALY) im Vergleich zum Szenario ohne Prävention; keine Screeningkosten

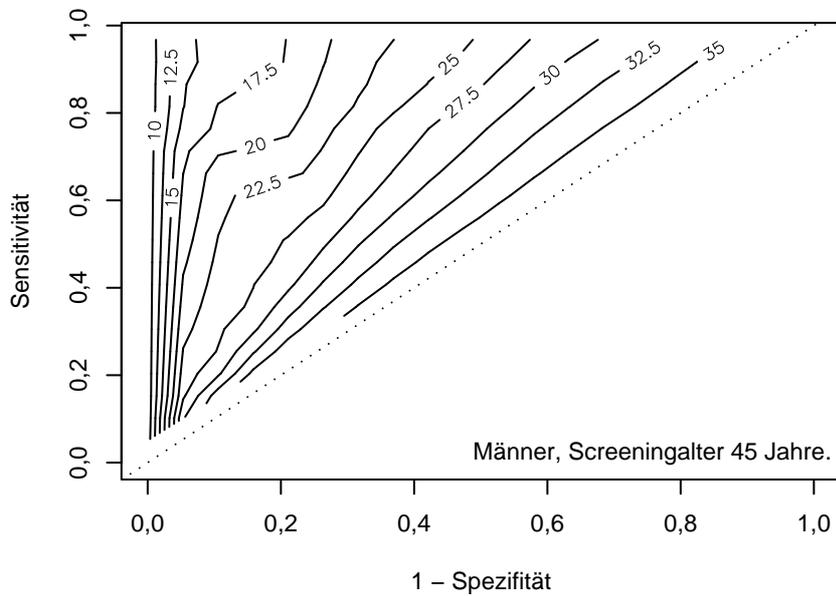


Abbildung 8.8: Kurven mit konstanter Kosten-Nutzen-Relation (in tausend Euro pro QALY) im Vergleich zum Szenario ohne Prävention; keine Screeningkosten

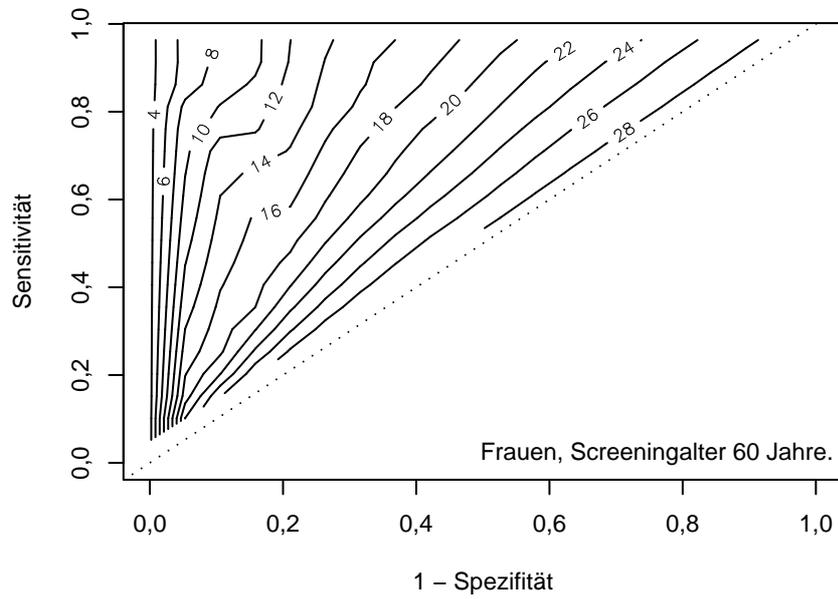


Abbildung 8.9: Kurven mit konstanter Kosten-Nutzen-Relation (in tausend Euro pro QALY) im Vergleich zum Szenario ohne Prävention; keine Screeningkosten

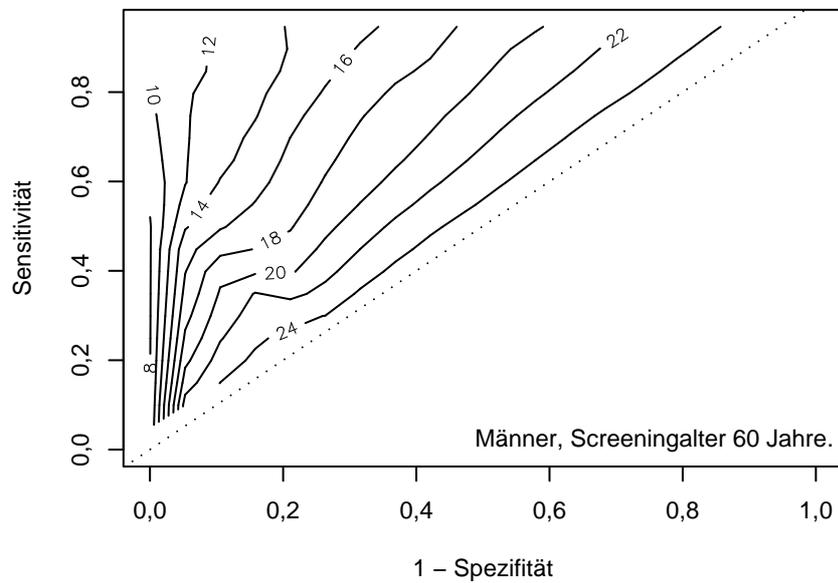


Abbildung 8.10: Kurven mit konstanter Kosten-Nutzen-Relation (in tausend Euro pro QALY) im Vergleich zum Szenario ohne Prävention; keine Screeningkosten

Bei einem Screening 45-jähriger Frauen sind alle abgebildeten Präventionsszenarien dem Szenario ohne Prävention vorzuziehen (siehe Abbildung 8.7). Einige Bereiche liegen aber nur sehr knapp unter der gewählten Zahlungsbereitschaft von 50.000 Euro pro QALY. Dieser Abstand vergrößert sich bei 45-jährigen Männern etwas: Die Kosten-Nutzen-Relationen liegen dort alle unterhalb von 40.000 Euro pro QALY (siehe Abbildung 8.8). Bei 60-jährigen Männern und Frauen sind ebenfalls alle abgebildeten Präventionsszenarien dem Szenario ohne Prävention deutlich vorzuziehen (vgl. Abbildungen 8.9 und 8.10). Die Szenarien mit der geringsten Kosten-Nutzen-Relation im Vergleich zum Szenario ohne Prävention liegen bei 28.000 Euro pro QALY bei 60-jährigen Frauen und bei 24.000 Euro pro QALY bei 60-jährigen Männern.

Es ist anzumerken, dass die Kosten-Nutzen-Relationen durch das Einbeziehen von Screeningkosten steigen, d. h. sich verschlechtern, würden. Außerdem können Szenarien, die im Vergleich zum Szenario ohne Prävention kosteneffektiv sind, einem anderen Präventionsszenario unterlegen sein.

### 8.3 Die Wahl eines „optimalen“ Szenarios

Wenn ein Screeningverfahren entwickelt wird, kommen unterschiedliche Trennpunkte in Frage, durch die die gescreente Population in Personen mit erhöhtem und nicht erhöhtem Risiko aufgeteilt wird. Es stellt sich die Frage, wie solch ein Trennpunkt gewählt werden soll, damit ein Screeningverfahren unter gegebenen Bedingungen „optimal“ ist. In einer herkömmlichen Kosten-Nutzen-Analyse werden nur zwei Szenarien miteinander verglichen. Wird die Wahl des Trennpunktes in das Entscheidungsproblem mit einbezogen, so müssten theoretisch unendlich viele Szenarien miteinander verglichen werden.

Die Punkte auf einer kohortenspezifischen ROC-Kurve geben an, welche Sensitivität-Spezifität-Kombinationen durch ein Risikoscreening für die entsprechende Kohorte erreicht werden können. Wie in Abschnitt 8.1 beschrieben, lassen sich für jede dieser Kombinationen die diskontierten Kosten und der diskontierte Nutzen durch Interpolation berechnen. Sind die ROC-Kurven unterschiedlicher Screeningmaßnahmen bekannt, so können einzelne Szenarien dieser Screeningmaßnahmen durch das in dieser Arbeit entwickelten Markovmodell miteinander verglichen werden. Da in der vorliegenden Analyse lediglich die ROC-Kurve der Framingham-Risiken bekannt ist, beschränkt sich die Suche des „optimalen“ Szenarios auf Präventionsszenarien, die dieser ROC-Kurve entsprechen, sowie dem Szenario ohne Primärprävention. Die anfallenden Screeningkosten werden berücksichtigt. Näherungsweise wird eine endliche Zahl von Trennpunkten betrachtet. Dazu werden 200 Werte der Sensitivität von 0 bis 1 in gleichen Abständen vorgegeben und die zugehörigen Spezifitätswerte bestimmt. Die realisierbaren Kosten-Nutzen-Kombinationen werden in Abbildung 8.11 dargestellt. Eine maximale Zahlungsbereitschaft lässt sich als Steigung innerhalb dieser Abbildung darstellen. Das „optimale“ Szenario resultiert, indem eine Gerade mit der Steigung der Zahlungsbereitschaft von links oben nach rechts unten verschoben wird. Der Punkt, der als erster von der Gerade berührt wird, stellt das optimale Szenario dar. Werden mehrere Punkte gleichzeitig berührt, so ist von diesen Szenarien das mit dem höchsten Nutzen zu bevorzugen. Eine andere Möglichkeit, das „optimale“ Szenario zu ermitteln, ist es, paarweise Kosten-Nutzwert-Analysen durchzuführen und unterlegene Szenarien zu entfernen (vgl. Abschnitt 2.4). Diese Methode wurde in der Analyse verwendet, um das jeweilige „optimale“ Szenario zu identifizieren. Wird keine Zahlungsbereitschaft festgelegt, so ist es prinzipiell möglich, dominierte Szenarien zu entfernen, worauf in dieser Analyse jedoch verzichtet wird. Bei 45- und 60-jährigen Männern und Frauen ist in Abbildung 8.11 das Szenario ohne Primärprävention zu erkennen (links unten): Bei gleichbleibendem Nutzen gibt es einen Sprung zu einem Szenario mit geringeren Kosten. Der Kostenunterschied kommt dadurch zustande, dass keine

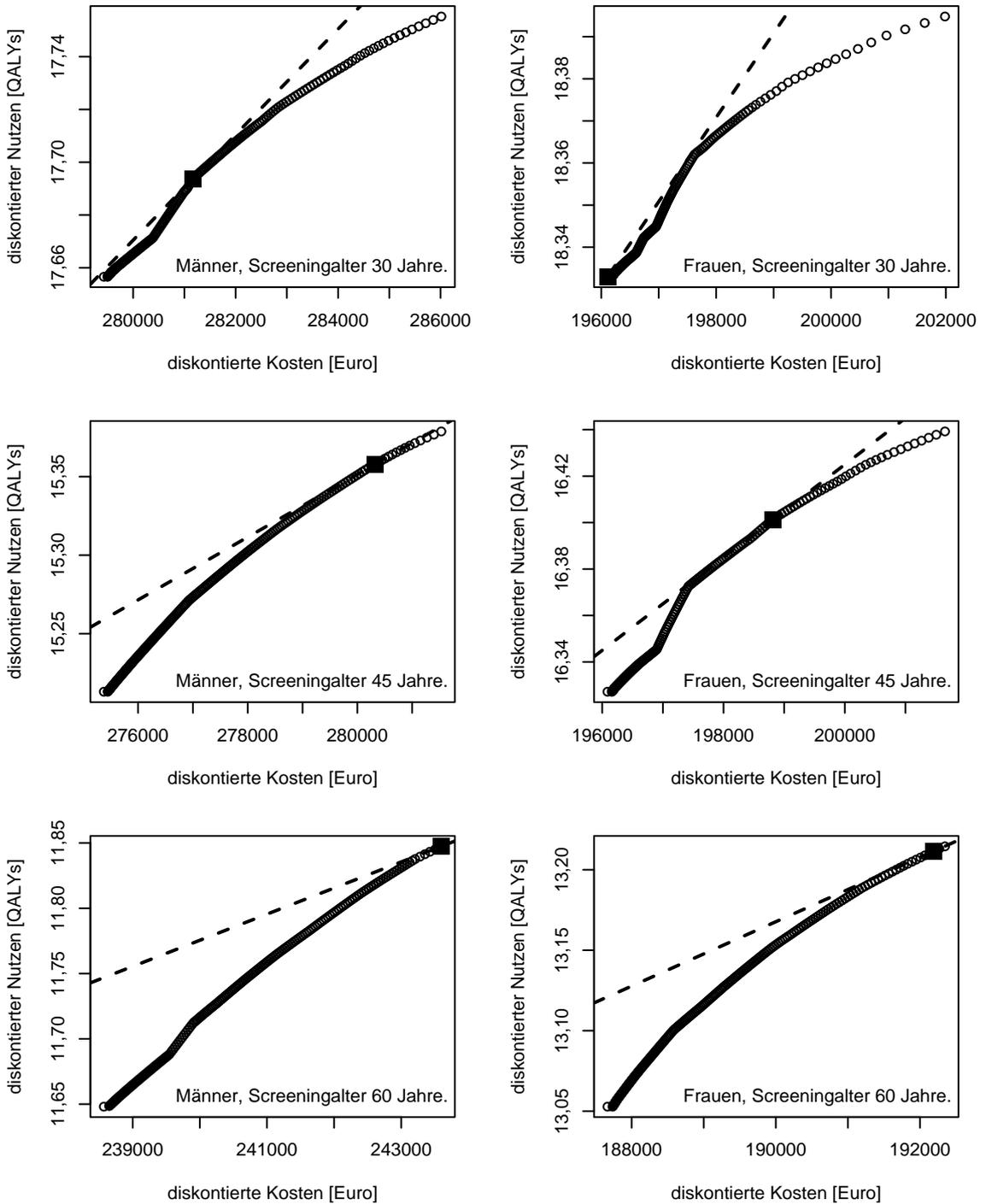


Abbildung 8.11: Realisierbare Kosten-Nutzen-Kombinationen, mit und ohne Prävention, Screeningkosten: 70 Euro, Zahlungsbereitschaft 50.000 Euro/QALY; „Optimales“ Szenario

Screeningkosten entstehen, wenn keine Screening durchgeführt wird.

Weiterhin lässt sich erkennen, dass die zusätzlichen Kosten bei 30-jährigen Frauen in keinem günstigen Verhältnis zum gewonnenem Nutzen stehen. Für die übrigen Kohorten existieren in der Basiswertanalyse Framingham-Screeningszenarien, die den Szenarien ohne Screening vorzuziehen sind. Rein theoretisch wäre es denkbar, dass durch Primärprävention Folgekosten erspart würden. Dieser Zusammenhang ist in Abbildung 8.11 nicht zu erkennen. Es ist jedoch ersichtlich, dass Primärprävention bei steigendem Alter lohnenswerter wird, da pro entstehender Kosteneinheit ein höherer Nutzen erzielt werden kann. Die Trennpunkte werden in Kapitel 9 in Tabelle 9.1 zusammen mit den durch die Sensitivitätsanalyse ermittelten Prognoseintervallen für jede Kohorte dargestellt.

# Kapitel 9

## Sensitivitätsanalyse

### 9.1 Festlegung der Monte-Carlo-Simulation

Als Sensitivitätsanalyse wird eine Monte-Carlo-Simulation durchgeführt. Bei dieser werden die im Modell verwendeten Parameterschätzer durch simulierte Parameterschätzer ersetzt. Wegen des hohen Rechenaufwandes, der pro Parameterkonstellation entsteht, werden nur 10 unterschiedliche Parameterkonstellationen pro Kohorte simuliert. Im Folgenden werden die Verteilungsannahmen der einfließenden Parameterschätzer getroffen.

#### 9.1.1 $\lambda$

Der Parameter  $\lambda$ , der angibt, wie stark die Aussagekraft des Screenings im Laufe der Zeit nachlässt, sollte zwischen 0 und 1 liegen. Bei einem Wert von 1 verschlechtert sich die Aussagekraft nicht, konvergiert  $\lambda$  gegen 0, so verschwindet als Grenzwert die Aussagekraft bereits im ersten Zyklus nach dem Screening komplett. Es bietet sich an, die Verteilung des Parameterschätzers  $\hat{\lambda}$  mittels der Betaverteilung (Abramowitz und Stegun 1972) zu approximieren, da die zugehörige Dichtefunktion außerhalb des Intervalls  $[0; 1]$  den Wert 0 annimmt. Die Parameter der Betaverteilung werden so gewählt, dass der Erwartungswert dem Schätzer für  $\lambda$  des Basisszenarios entspricht

und die in Abschnitt 5.4 geschätzte Standardabweichung von 5.4 erzielt wird.

### 9.1.2 $\psi_t^{\text{sup}}$

Die neuen Schätzer des Parameters  $\psi_t^{\text{sup}}$ , d. h. die simulierten Schätzer des durch potentielle Screeningverfahren maximal erzielbaren Odds Ratios bezüglich der sonstigen Mortalität zwischen Personen mit positivem und negativem Screeningergebnis, werden mittels der Gammaverteilung simuliert. Die Parameter der Gammaverteilung werden so gewählt, dass sich als Erwartungswert der Schätzer des Basisszenarios sowie als Standardabweichung der in Abschnitt 5.5.1 ermittelte Schätzer ergibt. Die Gammaverteilung wird gewählt, da diese nur positive Werte annehmen kann und der Wertebereich von Odds Ratios die positiven reellen Zahlen sind.

### 9.1.3 Transformation betaverteilter Zufallsvariablen

Im Zuge der Monte-Carlo-Simulation werden Werte simuliert, die innerhalb eines vorgegebenen Intervalls  $[v_1, v_2]$ ,  $v_1, v_2 \in \mathbb{R}$ , liegen sollen. Auch der Erwartungswert wird vorgegeben. Es sei  $X$  eine betaverteilte Zufallsvariable (Abramowitz und Stegun 1972) mit den Parametern  $\Upsilon_1$  und  $\Upsilon_2$ . Dann hat  $X$  den Erwartungswert  $\Upsilon_1/(\Upsilon_1 + \Upsilon_2)$  und die Zufallsvariable

$$Y := X \cdot (v_2 - v_1) + v_1 \quad (9.1)$$

hat den Erwartungswert

$$E(Y) = \frac{\Upsilon_1 \cdot (v_2 - v_1)}{\Upsilon_1 + \Upsilon_2} + v_1. \quad (9.2)$$

Außerdem ist  $Y$  beschränkt auf dem vorgegebenen Intervall  $[v_1, v_2]$ , was sich für  $X = 0$  und für  $X = 1$  und den Regeln für monotone Transformationen leicht zeigen lässt.

Wird der Erwartungswert von  $Y$  durch  $\mu$  explizit vorgegeben, so lässt sich der Parameter  $\Upsilon_1$  in Abhängigkeit des Parameters  $\Upsilon_2$  darstellen:

$$\begin{aligned} \mu &= \frac{\Upsilon_1 \cdot (v_2 - v_1)}{\Upsilon_1 + \Upsilon_2} + v_1 \\ \Leftrightarrow \Upsilon_1 &= \frac{\Upsilon_2 \cdot (\mu - v_1)}{v_2 - \mu}. \end{aligned} \quad (9.3)$$

#### 9.1.4 $\psi_i^{\text{sup}}$

Im Basisszenario wurde davon ausgegangen, dass  $\psi_i^{\text{sup}}$ , das durch Screeningverfahren maximal erzielbare Odds Ratio bezüglich der Herzinsuffizienzspezifischen Mortalität zwischen Personen mit positivem und negativem Screeningergebnis, doppelt so groß wie  $\psi_t^{\text{sup}}$  ist. In der Sensitivitätsanalyse wird der jeweils simulierte Parameterschätzer  $\hat{\psi}_t^{\text{sup}}$  für  $\psi_t^{\text{sup}}$  mit einer Zufallsvariablen multipliziert und als Parameterschätzer  $\hat{\psi}_i^{\text{sup}}$  für  $\psi_i^{\text{sup}}$  verwendet. Diese Zufallsvariable soll keine Werte kleiner als 1 annehmen, da das Odds Ratio bezüglich sonstiger Mortalität nicht kleiner als das Odds Ratio bezüglich Herzinsuffizienzspezifischer Mortalität ausfallen soll. Nach oben sollen stärkere Abweichungen zugelassen werden: Es wird die Obergrenze 5 gewählt. Als Erwartungswert wird die 2, der Wert des Basisszenarios gewählt. Die festzulegende Zufallsvariable wird durch die Transformation einer betaverteilten Zufallsvariable definiert. Für den Parameter  $\Upsilon_2$  wird der Wert 6 vorgegeben, da bei diesem Parameter am rechten Rand des positiven Wertebereichs der Übergang verhältnismäßig glatt ist, die Streuung aber noch groß genug ist. Die Dichtefunktion des Multiplikators wird zur Veranschaulichung in Abbildung 9.1 a) graphisch dargestellt.

#### 9.1.5 $\omega_i$

Das Odds Ratio  $\omega_i$  zwischen Personen mit und ohne KHK bezüglich der Herzinsuffizienzspezifischen Mortalität wurde mittels eines Regressionsmodells mit logistischer Linkfunktion geschätzt. Dabei wurde die Standardabweichung des Schätzers für den Parameter  $\log(1/\omega_i)$  bestimmt. Da der Wertebereich für  $\log(1/\omega_i)$  die gesamten re-

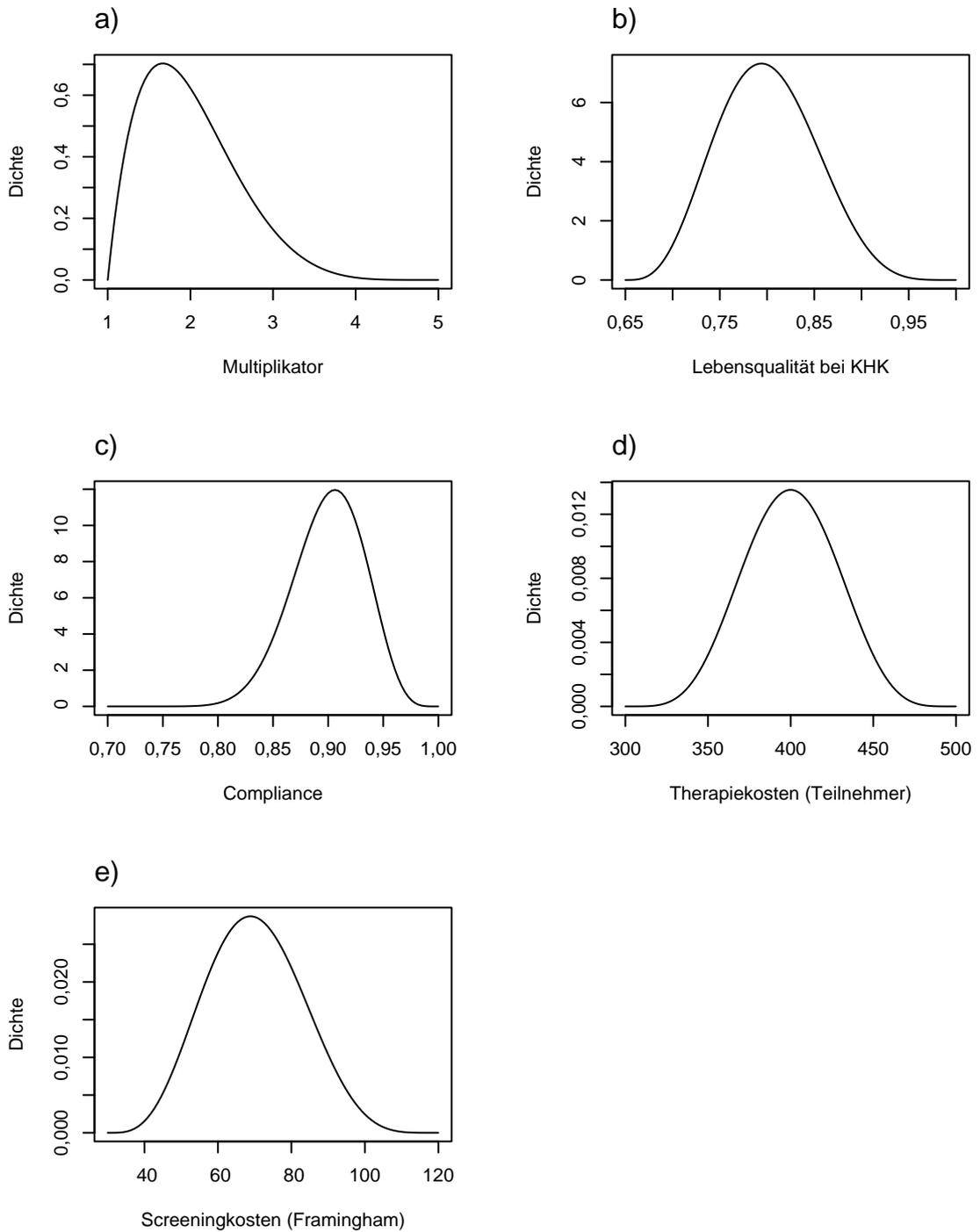


Abbildung 9.1: Dichtefunktionen der in der Sensitivitätsanalyse verwendeten transformierten Betaverteilungen

ellen Zahlen umfasst, ist es sinnvoll zu unterstellen, dass  $\log(1/\hat{\omega}_i)$  normalverteilt ist. Bei der Simulation neuer  $\log(1/\hat{\omega}_i)$  und daraus gewonnener  $\hat{\omega}_i$  werden die in Abschnitt 5.5.4 (Tabelle 5.8) ermittelten Schätzer als Erwartungswert und Standardabweichung von  $\log(1/\hat{\omega}_i)$  eingesetzt.

### 9.1.6 $\omega_k$ und $\omega_t$

Die Odds Ratio  $\omega_k$  und  $\omega_t$  bezüglich der KHK-spezifischen bzw. der sonstigen Mortalität zwischen Personen mit und ohne KHK wurden auf der Grundlage von Todesfällen sowie gewisser Beobachtungszeiten geschätzt (siehe Abschnitt 5.5.3). Neue Schätzer für  $\omega_k$  und  $\omega_t$  werden generiert, indem die zugrunde liegenden Todesfälle ( $K_k^{\text{KHK}}$ ,  $K_k^{\text{ohne KHK}}$ ,  $K_t^{\text{KHK}}$  und  $K_t^{\text{ohne KHK}}$ ) simuliert werden. Da es für Häufigkeiten seltener Ereignisse üblich ist, eine Poissonverteilung zu unterstellen, wird die Poissonverteilung zur Simulation der Todesfälle verwendet. Als Erwartungswerte der entsprechenden Poissonverteilungen werden die im Basisszenario verwendeten Schätzer eingesetzt.

### 9.1.7 Die Prävalenz der KHK

In der Sensitivitätsanalyse wird der Parameter  $\Omega$  aus Abschnitt 5.2.3 variiert, um unterschiedliche KHK-Prävalenzen zu erzeugen. Durch den Parameter  $\Omega$  wird festgelegt, zu welchen Anteilen der Prävalenzschätzer auf Basis der Kassendaten und der Prävalenzschätzer auf Basis des Bundesgesundheits surveys in die Prävalenz des Markovmodells einfließen. Die „wahre“ Größe von  $\Omega$  ist unbekannt. Der Parameterschätzer  $\hat{\Omega}$  wird simuliert, indem für ihn eine Rechteckverteilung auf dem Intervall  $[0; 1]$  unterstellt wird.

### 9.1.8 $\delta_{\text{HI}}$ , $Q_1$ und $Q_2$

Für die folgenden drei Parameter wird in der Monte-Carlo-Simulation ein einheitliches Vorgehen gewählt: Für  $\delta_{\text{HI}}$ , den Anteil vermeidbarer Herzinsuffizienz bei Personen mit KHK, unter der Annahme, dass die KHK vermieden werden kann, für  $Q_1$ , dem Faktor, um den die Gesundheitsausgaben bei Personen mit KHK im Vergleich zu Personen ohne KHK erhöht sind sowie für  $Q_2$ , dem Anteil der Gesundheitsausgaben, der bei Personen mit KHK vermieden werden kann, wenn die KHK wegfällt. Für jeden dieser drei Parameter wird pro Simulationsdurchlauf eine standardnormalverteilte Zufallsvariable generiert. Diese wird mit der Standardabweichung des jeweiligen Parameterschätzers multipliziert und zum Parameterschätzer des Basis-szenarios addiert. Für die unterschiedlichen Altersklassen wird dieselbe Zufallsvariable verwendet. Die simulierten Parameterschätzer ähneln den approximativen Konfidenzbändern der Parameter (siehe Abbildungen 5.14, 6.11 und 6.12), mit dem Unterschied, dass nicht das 2,5 %- und das 97,5 %-Quantil mit der Standardabweichung multipliziert wird, sondern die simulierten standardnormalverteilten Zufallsvariablen als Multiplikator verwendet werden.

### 9.1.9 Die Lebensqualität bei KHK

Als Schätzer für die Lebensqualität  $u^{\text{KHK}}$  von Personen mit KHK werden in der Monte-Carlo-Simulation linear transformierte betaverteilte Zufallsvariablen generiert, die auf dem Intervall  $[0,65; 1]$  positive Wahrscheinlichkeit besitzen. Für den Parameter  $\Upsilon_2$  wird der Wert 6 vorgegeben. Die Dichte des Parameterschätzers  $\hat{u}^{\text{KHK}}$  wird zur Veranschaulichung in Abbildung 9.1 b) graphisch dargestellt.

### 9.1.10 Die Compliance

Nicht alle Personen, die am Risikoscreening teilnehmen und als Personen mit erhöhtem Risiko identifiziert werden, nehmen tatsächlich an der Primärprävention teil. Der Anteil derjenigen, die am Screening teilnehmen, wird als Compliance  $C$

bezeichnet. Im Basisszenario wird für  $A$  der Schätzwert  $\hat{C} = 0,9$  verwendet (vgl. Abschnitt 5.7 und Abschnitt 6.9). In der Monte-Carlo-Simulation wird für  $\hat{C}$  der Erwartungswert 0,9 vorgegeben. Als potentiell Minimum und Maximum von  $\hat{C}$  werden 0,7 und 1 festgelegt. Weiterhin wird angenommen, dass es sich bei  $\hat{C}$  um eine lineare Transformation einer betaverteilten Zufallsvariable mit  $\Upsilon_2 = 6$  handelt. Gemäß Abschnitt 9.1.3 kann der Parameter  $\Upsilon_1$  der zugrundeliegenden Betaverteilung berechnet werden. Die Dichtefunktion von  $\hat{C}$  wird zur Veranschaulichung in Abbildung 9.1 c) graphisch dargestellt.

### 9.1.11 Therapiekosten

Für Personen mit positivem Screeningergebnis fallen im Präventionsszenario Therapiekosten an. Die Therapiekosten pro behandelter Person werden in der Sensitivitätsanalyse simuliert. Für sie werden ein Erwartungswert von 400 Euro, ein Minimum von 300 Euro und ein Maximum von 500 Euro unterstellt. Weiter wird unterstellt, dass es sich bei diesen Kosten um eine lineare Transformation einer betaverteilten Zufallsvariable handelt. Um die Betaverteilung eindeutig festzulegen, wird zusätzlich der Parameter  $\Upsilon_2 = 6$  vorgegeben. Die Dichtefunktion der Therapiekosten pro behandelter Person wird in Abbildung 9.1 d) graphisch dargestellt. Um die durchschnittlichen Therapiekosten pro Person mit positivem Screeningergebnis zu ermitteln, müssen diese Kosten noch mit der geschätzten Compliance  $\hat{A}$  multipliziert werden, deren Ausprägung gemäß Abschnitt 9.1.10 festgelegt wird.

### 9.1.12 Screeningkosten

Auch die Kosten, die pro Risikoscreening mittels der Framingham-Risikogleichung entstehen, unterliegt gewissen Unsicherheiten. In der Monte-Carlo-Simulation werden pro Simulationsdurchlauf Werte zwischen 30 und 120 Euro generiert. Als Erwartungswert werden 70 Euro vorgegeben. Die Screeningkosten werden als lineare Transformation einer betaverteilten Zufallsvariablen modelliert. Damit die Betaver-

teilung eindeutig ist, wird der Parameter  $\Upsilon_2 = 6$  vorgegeben. Die Dichtefunktion der Screeningkosten wird in Abbildung 9.1 e) graphisch dargestellt.

### 9.1.13 $\widehat{RR}^{\text{Präv}}$

Aus der Meta-Analyse von Vreecer *et al.* (2003) wird für das Basisszenario ein relatives Risiko  $\widehat{RR}_{\text{Vreecer}}^{\text{Präv}}$  übernommen, das den potentiell möglichen Effekt primärer Prävention mit Statinen angibt. Für die Sensitivitätsanalyse wird angenommen, dass  $\log\left(\widehat{RR}_{\text{Vreecer}}^{\text{Präv}}\right)$  eine normalverteilte Zufallsvariable ist. Der Erwartungswert von  $\log\left(\widehat{RR}_{\text{Vreecer}}^{\text{Präv}}\right)$  wird mit seiner Realisation  $\log(0,67)$  geschätzt. Auch das 95 %-Konfidenzintervall von  $\log\left(\widehat{RR}_{\text{Vreecer}}^{\text{Präv}}\right)$  wird durch Logarithmieren der Konfidenzintervallgrenzen von  $\widehat{RR}_{\text{Vreecer}}^{\text{Präv}}$  erreicht. Unter der Annahme, dass es sich um ein symmetrisches Konfidenzintervall handelt, ergibt sich für die Varianz von  $\log\left(\widehat{RR}_{\text{Vreecer}}^{\text{Präv}}\right)$  der folgende Schätzer:

$$\widehat{\text{Var}}\left(\log\left(\widehat{RR}_{\text{Vreecer}}^{\text{Präv}}\right)\right) = \left(\frac{\log(0,77) - \log(0,58)}{2 \cdot 1,96}\right)^2 = 5,2 \cdot 10^{-3}. \quad (9.4)$$

Es werden neue Werte für  $\log\left(\widehat{RR}_{\text{Vreecer}}^{\text{Präv}}\right)$  simuliert. Der im Markovmodell verwendete Parameterschätzer  $\widehat{RR}^{\text{Präv}}$  berücksichtigt im Gegensatz zu  $\widehat{RR}_{\text{Vreecer}}^{\text{Präv}}$  die Compliance. Das Risiko von Personen, die keine Statine einnehmen, bleibt unverändert. Für sie ist das relative Risiko im Vergleich zum Szenario ohne Screening gleich 1. Wird mit der geschätzten Compliance  $\hat{A}$  zwischen diesen beiden Personengruppen gewichtet, so ergibt sich:

$$\widehat{RR}^{\text{Präv}} = \exp\left(\log\left(\widehat{RR}_{\text{Vreecer}}^{\text{Präv}}\right)\right) \cdot \hat{A} + (1 - \hat{A}). \quad (9.5)$$

### 9.1.14 ROC-Kurve

Die ROC-Kurve einer Kohorte unterliegt zwei Arten von Unsicherheiten: Erstens sind die Parameter der Framingham-Risikogleichung mit Unsicherheiten behaftet.

Zweitens handelt es sich bei der Population des Bundesgesundheits surveys um eine zufallsbehaftete Stichprobe. Da für die Parameter der Framingham-Risikogleichung in der Publikation von Anderson *et al.* (1991) weder Varianzen noch Konfidenzintervalle angegeben sind, werden diese Parameter in der Monte-Carlo-Simulation nicht variiert. Dass sich in der Stichprobe des Bundesgesundheits surveys eine andere Verteilung der Risikofaktoren hätte ergeben können, wird in der Monte-Carlo-Simulation durch das Bootstrap-Verfahren (Efron und Tibshirani 1993) eingebettet: Aus der jeweiligen im Bundesgesundheits survey realisierten Stichprobe wird eine erneute Stichprobe mit Zurücklegen gezogen. Der Stichprobenumfang wird dabei beibehalten. Zu der simulierten Stichprobe wird in Analogie zu Kapitel 7 die ROC-Kurve bestimmt.

### 9.1.15 Zahlungsbereitschaft

Die Zahlungsbereitschaft muss festgelegt werden, damit das „optimale“ Szenario bestimmt werden kann. Die Zahlungsbereitschaft fließt nicht als Zufallsvariable in die Monte-Carlo-Simulation ein. Wie in der Basiswertanalyse wird von einer maximalen Zahlungsbereitschaft von 50.000 Euro pro gewonnenem QALY ausgegangen.

## 9.2 Weitere Berechnungen

Für jede simulierte Parameterkombination wird das Markovmodell evaluiert. Sowohl für das Szenario ohne Prävention, als auch für die Präventionsszenarien der simulierten Framingham-ROC-Kurve\* werden die diskontierten Kosten und der diskontierte Nutzen bestimmt. Für jedes Präventionsszenario wird die inkrementelle Kosten-Nutzen-Relation im Vergleich zum Szenario ohne Prävention bestimmt. Alle berechneten Kosten-Nutzen-Relationen sind positiv. Das bedeutet,

---

\*In Analogie zu Kapitel 8 werden nur für eine endliche Zahl von Punkten auf der ROC-Kurve die diskontierten Kosten und Nutzen bestimmt.

dass zusätzlicher Nutzen immer mit zusätzlichen Kosten verbunden ist. Pro Durchlauf der Monte-Carlo-Simulation wird die niedrigste Kosten-Nutzen-Relation identifiziert. Diese Kosten-Nutzen-Relation kann als eine Zahlungsbereitschaft interpretiert werden, bei der gerade noch ein Präventionsszenario existiert, dessen zusätzliche Kosten in einem günstigen Verhältnis zu dem zusätzlichen Nutzen stehen. Liegt die tatsächliche Zahlungsbereitschaft unter diesem Wert, so sollte von einem Risikoscreening mit der Framingham-Risikogleichung abgeraten werden. Trifft die tatsächliche Zahlungsbereitschaft diesen Wert dagegen genau oder ist höher, so existiert ein Präventionsszenario, das im Vergleich zum Szenario ohne Prävention „lohnenswert“ ist. Aus allen Simulationsdurchläufen wird für jede Zahlungsbereitschaft die Wahrscheinlichkeit geschätzt, dass ein sich „lohnendes“ Präventionsszenario existiert. Der Punktschätzer dieser Wahrscheinlichkeit ist der Anteil der Simulationsdurchläufe, bei denen zur jeweiligen Zahlungsbereitschaft ein „lohnenswertes“ Präventionsszenario existiert. Da insgesamt nur 10 Simulationsdurchläufe durchgeführt werden, sind diese Punktschätzer mit einer relativ hohen Unsicherheit belastet. Um diese Unsicherheit abzuschätzen werden die Pearson-Clopper-Werte als 95 %-Konfidenzintervalle berechnet. Die Berechnung der Pearson-Clopper-Werte wird ausführlich von Hartung *et al.* (2005) beschrieben.

Des Weiteren wird für jede der sechs Kohorten ein Prognoseintervall für den Schätzer des „optimalen“ Trennpunktes ermittelt, mit dem die Kohorte bei einem Präventionsszenario in Personen mit hohem bzw. mit niedrigem Risiko unterteilt werden sollte. Die in den einzelnen Simulationsdurchläufen ermittelten Schätzer für den optimalen Trennpunkt werden logit-transformiert, da sie nur Werte zwischen 0 und 1 annehmen können. Den logit-transformierten Trennpunkten wird unterstellt, dass sie Realisationen einer normalverteilten Zufallsvariable sind. Der Erwartungswert und die Varianz dieser Normalverteilung werden geschätzt und ein beidseitiges 95 %-Prognoseintervall für den logit-transformierten Trennpunkt ermittelt. Die Prognoseintervallgrenzen und die Prognoseintervallmitte werden mit der Inversen der

logit-Funktion rücktransformiert. Es resultiert ein Prognoseintervall für Schätzer des „optimalen“ Trennpunktes zukünftiger Simulationsdurchläufe.

### 9.3 Ergebnisse der Sensitivitätsanalyse

Ausgehend von einer maximalen Zahlungsbereitschaft von 50.000 Euro/QALY ist es sowohl bei 45- als auch bei 60-jährigen Männern und Frauen „lohenswert“ ein KHK-Risikoscreening mittels der Framingham-Risikogleichung durchzuführen. Aus Abbildung 9.2 ist zu erkennen, dass mit einer Wahrscheinlichkeit von 100 % (Punktschätzer) ein Präventionsszenario existiert, in dem der zusätzliche Nutzen in einem günstigen Verhältnis zu den zusätzlichen Kosten steht. Ob 30-jährige Männer und Frauen an einem Risikoscreening mittels der Framingham-Risikogleichung teilnehmen sollten ist grenzwertig. Bei einer Zahlungsbereitschaft von 50.000 Euro/QALY beträgt die Wahrscheinlichkeit, dass ein „lohenswertes“ Präventionsszenario existiert, bei Männern 70 % und bei Frauen 60 %.

Sollte eine konkrete Einführung eines Risikoscreenings seitens der Krankenkassen bevorstehen, so empfiehlt es sich, die Sensitivitätsanalyse mit genauer kalkulierten Behandlungskosten zu wiederholen. Eine höhere Anzahl an Simulationsdurchläufen wäre in diesem Fall gegebenenfalls sinnvoll. Im Rahmen dieser Arbeit ist die Anzahl der Simulationsdurchläufe jedoch ausreichend.

Die 95 %-Prognoseintervalle für den „optimalen“ Trennpunkt werden in Tabelle 9.1 dargestellt. Es ist zu erkennen, dass ihre Breite von Kohorte zu Kohorte variiert. Mit Hilfe der Prognoseintervalle kann abgeschätzt werden, wie genau die durch die Basiswertanalyse ermittelten Punktschätzer sind. Diese befinden sich ebenfalls in Tabelle 9.1. Zum Vergleich werden ihnen die Werte, die sich aus der Monte-Carlo-Simulation als Punktschätzer ergeben hätten, gegenübergestellt<sup>†</sup>. Die

---

<sup>†</sup>Sie entsprechen dem  $\text{logit}^{-1}$ -transformierten Mittelwert der simulierten  $\text{logit}$ -transformierten

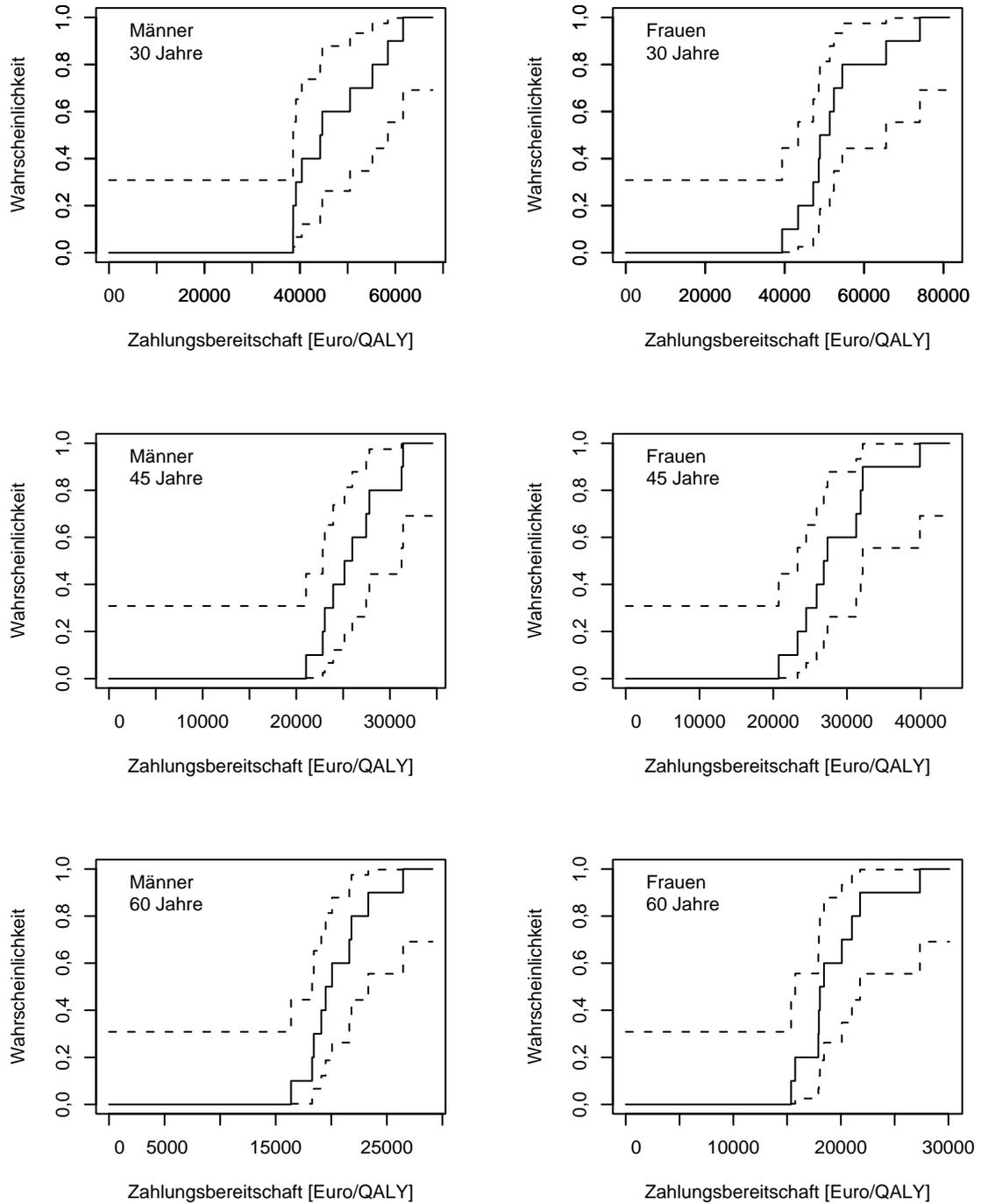


Abbildung 9.2: Wahrscheinlichkeit, dass ein Präventionsszenario existiert, in dem der zusätzliche Nutzen in einem „günstigen Verhältnis“ zu den zusätzlichen Kosten steht; Screening mittels Framingham-Risikogleichung

Alter	Geschlecht	$CP_{\text{basis}}^{\text{opt}}$	$CP_{\text{sens}}^{\text{opt}}$	Prognoseintervall	$SD(\text{logit}(CP^{\text{opt}}))$
30	männlich	0,03	0,01	(0,00; 0,11)	1,28
45	männlich	0,06	0,06	(0,03; 0,10)	0,31
60	männlich	0,04	0,04	(0,03; 0,07)	0,25
30	weiblich	0,00	0,00	(0,00; 0,03)	2,47
45	weiblich	0,02	0,03	(0,01; 0,05)	0,30
60	weiblich	0,04	0,03	(0,02; 0,07)	0,37

Tabelle 9.1: Kohortenspezifische optimale Trennpunkte, um Personen anhand des Framinghamrisikos in Personen mit erhöhtem und geringem Risiko zu unterteilen

durch die Monte-Carlo-Simulation ermittelten Punktschätzer unterscheiden sich kaum von denen des Basisszenarios.

# Kapitel 10

## Zusammenfassung

Die koronare Herzkrankheit (KHK) ist eine der häufigsten Todesursachen in Industrieländern (Andrawes *et al.* 2005). Außerdem führt die KHK zu einer geringeren Lebensqualität der Betroffenen (Torres *et al.* 2004). Durch primärpräventive Maßnahmen ist es möglich die Wahrscheinlichkeit an der KHK zu erkranken zu reduzieren (Hay und Sterling 2005). Aus diesem Grund wurden in den letzten Jahren verstärkt gesundheitsökonomische Analysen durchgeführt, in denen untersucht wurde, ob Primärprävention zur Vermeidung der KHK ökonomisch sinnvoll ist. Die bisherigen Analysen können zwar als Indiz für die steigende Relevanz von Evaluationen auf diesem Gebiet angesehen werden, decken das breite Spektrum spezieller Fragestellungen jedoch nicht ab. In dieser Arbeit wird ein Markovmodell entwickelt, mit dem untersucht werden kann, ob ein KHK-Risikoscreening mit anschließender Primärprävention für Personen mit erhöhter Erkrankungswahrscheinlichkeit ökonomisch sinnvoll ist. Primärprävention erfolgt hier mittels einer lebenslangen Gabe von Statinen. Die anschließende gesundheitsökonomische Analyse wird aus der Perspektive der Sozialversicherungen durchgeführt.

Mittels des Modells kann erstens beantwortet werden, ob es unter ökonomischen Gesichtspunkten sinnvoll ist, ein entsprechendes Risikoscreening durchzuführen. Zweitens kann beantwortet werden, wie der Trennpunkt gewählt werden soll, der

festgelegt, welche Personen Prävention erhalten und welche nicht. Drittens ist das Modell dazu geeignet, unterschiedliche Screeningmaßnahmen miteinander zu vergleichen.

Das Markovmodell wird so entwickelt, dass es der Realität so nahe wie möglich kommt. Dieses Modell wird schrittweise erstellt. Ausgehend von einem groben Modell werden die Zustände weiter unterteilt. Dabei werden vier Gleichungssysteme aufgestellt, die auf plausiblen Annahmen beruhen und mit einer Vielzahl von Daten gespeist werden. Die Gleichungssysteme bestehen aus jeweils bis zu 30 Gleichungen. Wegen ihrer Komplexität ist es erforderlich, sie mittels einer mathematischen Software (Maple, Version 9.01, Monagan *et al.* 2003) zu lösen. Es werden Lösungen für jede Kombination aus Screeningalter, Alter und Geschlecht bestimmt. Dabei ist es erforderlich, induktiv vorzugehen, da endogene Größen eines Alters als exogene Größen des darauf folgenden Alters in das Gleichungssystem einfließen.

Bei der Entwicklung des Gleichungssystems werden Funktionen hergeleitet, die gewisse theoretische Eigenschaften erfüllen sollen. Die Parameter dieser Funktionen werden anschließend mittels nichtlinearer Regression (NLR) aus empirischen Daten geschätzt. Bei der Quantifizierung der ins Modell einfließenden Größen werden weitere statistische Verfahren eingesetzt. So wird die Prävalenz der KHK alters- und geschlechtsspezifisch mit Hilfe einer logistischen Regression aus den Daten des Bundesgesundheits surveys geschätzt. Diverse generalisierte additive Modelle (GAMs) werden angepasst, um erwartete Kosten und erwartete Prävalenzen der Herzinsuffizienz zu schätzen. Einige dieser GAMs werden verwendet, um den Zusammenhang zwischen der KHK und der Herzinsuffizienz zu quantifizieren, bzw. die attributablen KHK-Gesundheitsausgaben zu bestimmen. Dabei kommt es zur mehrfachen Anwendung des Fehlerfortpflanzungsgesetzes. Es werden die dazu benötigten Ableitungen und Kovarianzmatrizen hergeleitet. Einige GAMs basieren auf aggregierten Daten des sozio-ökonomischen Panels (SOEP). Für diese

GAMs werden Gewichte hergeleitet, in die sowohl die Hochrechnungsfaktoren der SOEP-Daten sowie die Annahme, dass sich die Struktur der Varianz bei allen Personen der Stichprobe nicht unterscheidet, einfließen.

Um Screeningverfahren bezüglich ihrer Qualität beurteilen zu können, werden die Definitionen der Sensitivität und der Spezifität diagnostischer Tests auf diese übertragen. Am Beispiel von Framingham-10-Jahres-Risiken wird hergeleitet, wie ROC-Kurven für KHK-Risikoscores berechnet werden können. ROC-Kurven geben alle möglichen Kombinationen von Sensitivität und Spezifität wieder, die mit einem Screeningverfahren erreicht werden können. Weiter wird hergeleitet, wie mit Hilfe des erstellten Markovmodells Sensitivität und Spezifität berechnet werden können.

Das Markovmodell wird für sechs Kohorten von Personen gleichen Geschlechts und gleichen Alters zum Screeningzeitpunkt für unterschiedliche Parameterkombinationen evaluiert, so dass sich unterschiedliche Kombinationen von Sensitivität und Spezifität ergeben. Für die realisierten Kombinationen wird eine Delaunay-Triangulation bestimmt, d. h. es werden möglichst gleichwinklige Dreiecke im Koordinatensystem aus Sensitivität und (1 minus der) Spezifität bestimmt. Über diesen Dreiecken werden Kosten und Nutzen für beliebige Sensitivität-Spezifität-Kombinationen linear interpoliert. Ist die ROC-Kurve eines Screeningverfahrens bekannt, so können alle durch dieses Verfahren realisierbaren Kosten-Nutzen-Kombinationen ermittelt werden. Screeningkosten sind darin allerdings nicht enthalten und müssen separat ermittelt werden. Es wird eine Kosten-Nutzwert-Analyse durchgeführt. Sowohl Kosten als auch Nutzen werden mit einer Rate von 5% diskontiert. Die Stabilität der Ergebnisse wird im Zuge einer Sensitivitätsanalyse in Form einer Monte-Carlo-Simulation überprüft.

Als mögliche Szenarien werden in der Analyse das Szenario ohne Prävention sowie alle durch ein Framingham-Risikoscreening erreichbaren Präventionsszenarien

betrachtet. Ein KHK-Risikoscreening in Kombination mit einer lebenslangen Statintherapie für Personen mit erhöhtem Risiko führt in keinem der betrachteten Fälle zu Kosteneinsparungen. Für einige Kohorten stehen die zusätzlichen Kosten jedoch in einem günstigen Verhältnis zu dem zusätzlichen Nutzen (bei 45- und 60-jährigen Männern und Frauen, ausgehend von einer maximalen Zahlungsbereitschaft von 50.000 Euro pro gewonnenem qualitätsadjustierten Lebensjahr (QALY)). Bei 30-jährigen Männern und Frauen existiert im Basisszenario zwar kein Präventionsszenario, in dem die zusätzlichen Kosten in einem angemessenen Verhältnis zu dem gewonnenem Nutzen stehen, diese Aussage hat sich im Zuge der Monte-Carlo-Simulation jedoch als sensitiv, d. h. als ungewiss erwiesen.

# Anhang A

## Weitere Gleichungen

Einige Formeln der Gleichungssysteme aus Kapitel 4 sind vergleichsweise einfach und stören den Lesefluss. Der Vollständigkeit halber werden sie in diesem Anhang aufgeführt.

### A.1 Gleichungen des Ausgangsmodells

Exogen vorgegebene ursachenspezifische Sterberaten ( $p_{Li}$  und  $p_{Lt}$ ) müssen eingehalten werden:

$$p_{Li} = (1 - \pi(\text{Alter})) \cdot p_{gi} + \pi(\text{Alter}) \cdot p_{hi} \quad \text{und} \quad (\text{A.1})$$

$$p_{Lt} = (1 - \pi(\text{Alter})) \cdot p_{gt} + \pi(\text{Alter}) \cdot p_{ht} . \quad (\text{A.2})$$

Odds Ratios bezüglich ursachenspezifischer Sterbewahrscheinlichkeiten zwischen Personen mit und ohne KHK:

$$\omega_i = \frac{p_{gi}|_{\text{n. i. M.}}}{1 - p_{gi}|_{\text{n. i. M.}}} \cdot \frac{1 - p_{hi}|_{\text{n. i. M.}}}{p_{hi}|_{\text{n. i. M.}}} \quad \text{und} \quad (\text{A.3})$$

$$\omega_t = \frac{p_{gt}|_{\text{n. s. M.}}}{1 - p_{gt}|_{\text{n. s. M.}}} \cdot \frac{1 - p_{ht}|_{\text{n. s. M.}}}{p_{ht}|_{\text{n. s. M.}}} . \quad (\text{A.4})$$

Berechnung weiterer bedingter Übergangswahrscheinlichkeiten:

$$p_{hk}|_{\text{n. KHK-M.}} = \frac{p_{hk}}{1 - p_{hi} - p_{gt}} , \quad (\text{A.5})$$

$$p_{gi}|_{n. i. M.} = \frac{p_{gi}}{1 - p_{gk} - p_{gt}}, \quad (\text{A.6})$$

$$p_{hi}|_{n. i. M.} = \frac{p_{hi}}{1 - p_{hk} - p_{gt}}, \quad (\text{A.7})$$

$$p_{gt}|_{n. s. M.} = \frac{p_{gt}}{1 - p_{gk} - p_{gi}}, \quad (\text{A.8})$$

$$p_{ht}|_{n. s. M.} = \frac{p_{ht}}{1 - p_{hk} - p_{hi}}. \quad (\text{A.9})$$

Die Übergangswahrscheinlichkeiten eines jeweiligen Zustandes müssen 1 ergeben:

$$p_{gg} + p_{gh} + p_{gk} + p_{gi} + p_{gt} = 1, \quad (\text{A.10})$$

$$p_{hh} + p_{hk} + p_{hi} + p_{ht} = 1. \quad (\text{A.11})$$

Die Gesamtsterbewahrscheinlichkeit für Personen mit und ohne KHK ergeben sich mit

$$p_{gT} = p_{gk} + p_{gi} + p_{gt} \quad \text{und} \quad (\text{A.12})$$

$$p_{hT} = p_{hk} + p_{hi} + p_{ht}. \quad (\text{A.13})$$

## A.2 Gleichungen zur ersten Unterteilung der Personen mit KHK

Die zusätzlichen Odds Ratios, die als konstant angenommen werden, sind folgendermaßen definiert:

$$\text{OR}_{i. M.}^{g^0 h^R} := \frac{p_{g^0 i}|_{n. i. M.}}{1 - p_{g^0 i}|_{n. i. M.}} \cdot \frac{1 - p_{h^R i}|_{n. i. M.}}{p_{h^R i}|_{n. i. M.}} \quad \text{und} \quad (\text{A.14})$$

$$\text{OR}_{s. M.}^{g^0 h^R} := \frac{p_{g^0 t}|_{n. s. M.}}{1 - p_{g^0 t}|_{n. s. M.}} \cdot \frac{1 - p_{h^R t}|_{n. s. M.}}{p_{h^R t}|_{n. s. M.}}. \quad (\text{A.15})$$

Es müssen sich aus den neu entstehenden Übergangswahrscheinlichkeiten die des Ausgangsmodells ergeben. Dabei werden die Bevölkerungsanteile in den jeweiligen Zuständen (bzw. zusammengesetzten Zuständen) berücksichtigt.

$$p_{g^0 i} = a_{g^0}^g \cdot p_{gi} + a_{g^0}^{h^N} \cdot p_{h^N i}, \quad (\text{A.16})$$

### 177A.3 Gleichungen zur Unterteilung der Gesunden nach Erkrankungsrisiko

$$p_{g^0t} = a_{g^0}^g \cdot p_{gt} + a_{g^0}^{h^N} \cdot p_{h^Nt}, \quad (\text{A.17})$$

$$p_{hi} = a_h^{h^N} \cdot p_{h^Ni} + a_h^{h^R} \cdot p_{h^Ri} \quad \text{und} \quad (\text{A.18})$$

$$p_{ht} = a_h^{h^N} \cdot p_{h^Nt} + a_h^{h^R} \cdot p_{h^Rt}. \quad (\text{A.19})$$

Wahrscheinlichkeiten der neuen Zustände müssen sich zu 1 addieren:

$$p_{h^N h^N} + p_{h^N k} + p_{h^N i} + p_{h^N t} = 1, \quad (\text{A.20})$$

$$p_{h^R h^R} + p_{h^R k} + p_{h^R i} + p_{h^R t} = 1. \quad (\text{A.21})$$

Die drei Teilmortalitäten setzen sich zur Gesamtmortalität zusammen:

$$p_{h^N T} = p_{h^N k} + p_{h^N i} + p_{h^N t}, \quad (\text{A.22})$$

$$p_{h^R T} = p_{h^R k} + p_{h^R i} + p_{h^R t}. \quad (\text{A.23})$$

Es gelten die Regeln für bedingte Wahrscheinlichkeiten:

$$p_{g^0t} \Big|_{\text{n. s. M.}} := \frac{p_{g^0t}}{1 - p_{g^0k} - p_{g^0i}}, \quad (\text{A.24})$$

$$p_{h^Rt} \Big|_{\text{n. s. M.}} := \frac{p_{h^Rt}}{1 - p_{h^Rk} - p_{h^Ri}}, \quad (\text{A.25})$$

$$p_{g^0k} \Big|_{\text{n. KHK-M.}} := \frac{p_{g^0k}}{1 - p_{g^0t} - p_{g^0i}}, \quad (\text{A.26})$$

$$p_{h^Rk} \Big|_{\text{n. KHK-M.}} := \frac{p_{h^Rk}}{1 - p_{h^Rt} - p_{h^Ri}}, \quad (\text{A.27})$$

$$p_{g^0i} \Big|_{\text{n. i. M.}} := \frac{p_{g^0i}}{1 - p_{g^0k} - p_{g^0t}} \quad \text{und} \quad (\text{A.28})$$

$$p_{h^Ri} \Big|_{\text{n. i. M.}} := \frac{p_{h^Ri}}{1 - p_{h^Rk} - p_{h^Rt}}. \quad (\text{A.29})$$

## A.3 Gleichungen zur Unterteilung der Gesunden nach Erkrankungsrisiko

Die bedingten Wahrscheinlichkeiten lassen sich folgendermaßen berechnen:

$$(p_{g^+h^N} + p_{g^+k}) \Big|_{\text{n. KHK-M.}} = \frac{p_{g^+h^N} + p_{g^+k}}{1 - p_{g^+i} - p_{g^+t}}, \quad (\text{A.30})$$

$$(p_{g^{-h^N}} + p_{g^{-k}}) \Big|_{\text{n. KHK-M.}} = \frac{p_{g^{-h^N}} + p_{g^{-k}}}{1 - p_{g^{-i}} - p_{g^{-t}}}, \quad (\text{A.31})$$

$$p_{g^{+i}} \Big|_{\text{n. i. M.}} = \frac{p_{g^{+i}}}{1 - p_{g^{+k}} - p_{g^{+t}}}, \quad (\text{A.32})$$

$$p_{g^{-i}} \Big|_{\text{n. i. M.}} = \frac{p_{g^{-i}}}{1 - p_{g^{-k}} - p_{g^{-t}}}, \quad (\text{A.33})$$

$$p_{g^{+t}} \Big|_{\text{n. s. M.}} = \frac{p_{g^{+t}}}{1 - p_{g^{+k}} - p_{g^{+i}}} \quad \text{und} \quad (\text{A.34})$$

$$p_{g^{-t}} \Big|_{\text{n. s. M.}} = \frac{p_{g^{-t}}}{1 - p_{g^{-k}} - p_{g^{-i}}}. \quad (\text{A.35})$$

Die Übergangswahrscheinlichkeiten der neuen Zustände  $g^+$  und  $g^-$  müssen, gewichtet mit den Anteilen der Personen, die sich in den Zuständen befinden, die Übergangswahrscheinlichkeiten von  $g$  ergeben:

$$a_g^{g^+} \cdot p_{g^{+h^N}} + a_g^{g^-} \cdot p_{g^{-h^N}} = p_{gh^N}, \quad (\text{A.36})$$

$$a_g^{g^+} \cdot p_{g^{+k}} + a_g^{g^-} \cdot p_{g^{-k}} = p_{gk}, \quad (\text{A.37})$$

$$a_g^{g^+} \cdot p_{g^{+i}} + a_g^{g^-} \cdot p_{g^{-i}} = p_{gi} \quad \text{und} \quad (\text{A.38})$$

$$a_g^{g^+} \cdot p_{g^{+t}} + a_g^{g^-} \cdot p_{g^{-t}} = p_{gt}. \quad (\text{A.39})$$

Die drei Mortalitätsarten addieren sich zur Gesamtmortalität.

$$p_{g^{+T}} = p_{g^{+k}} + p_{g^{+i}} + p_{g^{+t}}, \quad (\text{A.40})$$

$$p_{g^{-T}} = p_{g^{-k}} + p_{g^{-i}} + p_{g^{-t}}. \quad (\text{A.41})$$

Die Wahrscheinlichkeiten eines Zustandes addieren sich zu 1.

$$p_{g^{+T}} + p_{g^{+h^N}} + p_{g^{+g^+}} = 1, \quad (\text{A.42})$$

$$p_{g^{-T}} + p_{g^{-h^N}} + p_{g^{-g^-}} = 1. \quad (\text{A.43})$$

## A.4 Gleichungen zur zweiten Unterteilung der Personen mit KHK

Die Übergangswahrscheinlichkeiten der Zustände  $h^+$  und  $h^-$  müssen die Übergangswahrscheinlichkeiten des Zustands  $h^N$  ergeben.

$$a_{h^N}^{h^+} \cdot p_{h^+k} + a_{h^N}^{h^-} \cdot p_{h^-k} = p_{h^Nk}, \quad (\text{A.44})$$

$$a_{h^N}^{h^+} \cdot p_{h^+i} + a_{h^N}^{h^-} \cdot p_{h^-i} = p_{h^Ni}, \quad (\text{A.45})$$

$$a_{h^N}^{h^+} \cdot p_{h^+t} + a_{h^N}^{h^-} \cdot p_{h^-t} = p_{h^Nt}. \quad (\text{A.46})$$

Es wird unterstellt, dass die Odds Ratios der bedingten Mortalitätsraten zwischen  $g^+$  zu  $g^-$  und  $h^+$  zu  $h^-$  identisch sind:

$$\text{OR}_{\text{KHK-M.}}^{g^+,g^-} \stackrel{!}{=} \text{OR}_{\text{KHK-M.}}^{h^+,h^-}, \quad (\text{A.47})$$

$$\text{OR}_{\text{i. M.}}^{g^+,g^-} \stackrel{!}{=} \text{OR}_{\text{i. M.}}^{h^+,h^-} \quad \text{und} \quad (\text{A.48})$$

$$\text{OR}_{\text{s. M.}}^{g^+,g^-} \stackrel{!}{=} \text{OR}_{\text{s. M.}}^{h^+,h^-} \quad (\text{A.49})$$

mit

$$\text{OR}_{\text{KHK-M.}}^{g^+,g^-} := \frac{p_{g^+k}|_{\text{n. KHK-M.}}}{1 - p_{g^+k}|_{\text{n. KHK-M.}}} \cdot \frac{1 - p_{g^-k}|_{\text{n. KHK-M.}}}{p_{g^-k}|_{\text{n. KHK-M.}}}, \quad (\text{A.50})$$

$$\text{OR}_{\text{i. M.}}^{g^+,g^-} := \frac{p_{g^+i}|_{\text{n. i. M.}}}{1 - p_{g^+i}|_{\text{n. i. M.}}} \cdot \frac{1 - p_{g^-i}|_{\text{n. i. M.}}}{p_{g^-i}|_{\text{n. i. M.}}}, \quad (\text{A.51})$$

$$\text{OR}_{\text{s. M.}}^{g^+,g^-} := \frac{p_{g^+t}|_{\text{n. s. M.}}}{1 - p_{g^+t}|_{\text{n. s. M.}}} \cdot \frac{1 - p_{g^-t}|_{\text{n. s. M.}}}{p_{g^-t}|_{\text{n. s. M.}}}, \quad (\text{A.52})$$

und

$$\text{OR}_{\text{KHK-M.}}^{h^+,h^-} := \frac{p_{h^+k}|_{\text{n. KHK-M.}}}{1 - p_{h^+k}|_{\text{n. KHK-M.}}} \cdot \frac{1 - p_{h^-k}|_{\text{n. KHK-M.}}}{p_{h^-k}|_{\text{n. KHK-M.}}}, \quad (\text{A.53})$$

$$\text{OR}_{\text{i. M.}}^{h^+,h^-} := \frac{p_{h^+i}|_{\text{n. i. M.}}}{1 - p_{h^+i}|_{\text{n. i. M.}}} \cdot \frac{1 - p_{h^-i}|_{\text{n. i. M.}}}{p_{h^-i}|_{\text{n. i. M.}}}, \quad (\text{A.54})$$

$$\text{OR}_{\text{s. M.}}^{h^+,h^-} := \frac{p_{h^+t}|_{\text{n. s. M.}}}{1 - p_{h^+t}|_{\text{n. s. M.}}} \cdot \frac{1 - p_{h^-t}|_{\text{n. s. M.}}}{p_{h^-t}|_{\text{n. s. M.}}}. \quad (\text{A.55})$$

Vier von den bedingten Wahrscheinlichkeiten wurden bereits in Abschnitt 4.4 definiert (Formeln A.32 bis A.35). Für die bedingte Wahrscheinlichkeit eines fatalen koronaren Ereignisses gelten folgende Gleichungen:

$$p_{g^+k} |_{\text{n. KHK-M.}} = \frac{p_{g^+k}}{1 - p_{g^+t} - p_{g^+i}}, \quad (\text{A.56})$$

$$p_{g^-k} |_{\text{n. KHK-M.}} = \frac{p_{g^-k}}{1 - p_{g^-t} - p_{g^-i}}. \quad (\text{A.57})$$

Für die übrigen bedingten Wahrscheinlichkeiten gilt analog:

$$p_{h^+k} |_{\text{n. KHK-M.}} = \frac{p_{h^+k}}{1 - p_{h^+t} - p_{h^+i}}, \quad (\text{A.58})$$

$$p_{h^-k} |_{\text{n. KHK-M.}} = \frac{p_{h^-k}}{1 - p_{h^-t} - p_{h^-i}}, \quad (\text{A.59})$$

$$p_{h^+i} |_{\text{n. i. M.}} = \frac{p_{h^+i}}{1 - p_{h^+t} - p_{h^+k}}, \quad (\text{A.60})$$

$$p_{h^-i} |_{\text{n. i. M.}} = \frac{p_{h^-i}}{1 - p_{h^-t} - p_{h^-k}}, \quad (\text{A.61})$$

$$p_{h^+t} |_{\text{n. s. M.}} = \frac{p_{h^+t}}{1 - p_{h^+k} - p_{h^+i}}, \quad (\text{A.62})$$

$$p_{h^-t} |_{\text{n. s. M.}} = \frac{p_{h^-t}}{1 - p_{h^-k} - p_{h^-i}}. \quad (\text{A.63})$$

Die Summe der Übergangswahrscheinlichkeiten eines Zustandes muss 1 ergeben:

$$p_{h^+h^+} + p_{h^+k} + p_{h^+i} + p_{h^+t} = 1, \quad (\text{A.64})$$

$$p_{h^-h^-} + p_{h^-k} + p_{h^-i} + p_{h^-t} = 1. \quad (\text{A.65})$$

Die Gesamtsterbewahrscheinlichkeit setzt sich aus den Teilsterbewahrscheinlichkeiten zusammen:

$$p_{h^+T} = p_{h^+k} + p_{h^+i} + p_{h^+t}, \quad (\text{A.66})$$

$$p_{h^-T} = p_{h^-k} + p_{h^-i} + p_{h^-t}. \quad (\text{A.67})$$

## A.5 Gleichungen zur Primärprävention

Bedingte Wahrscheinlichkeiten:

$$p_{g^{V+t}} \Big|_{\text{n. s. M.}} = \frac{p_{g^{V+t}}}{1 - p_{g^{V+i}} - p_{g^{V+k}}}, \quad (\text{A.68})$$

$$p_{h+t} \Big|_{\text{n. s. M.}} = \frac{p_{h+t}}{1 - p_{h+i} - p_{h+k}}, \quad (\text{A.69})$$

$$p_{g^{V+i}} \Big|_{\text{n. i. M.}} = \frac{p_{g^{V+i}}}{1 - p_{g^{V+k}} - p_{g^{V+t}}}, \quad (\text{A.70})$$

$$p_{h+i} \Big|_{\text{n. i. M.}} = \frac{p_{h+i}}{1 - p_{h+k} - p_{h+t}}, \quad (\text{A.71})$$

$$p_{g^{+h+}}^{\text{Präv}} \Big|_{\text{n. KHK.-M.}} = \frac{p_{g^{+h+}}^{\text{Präv}}}{1 - p_{g^{+i}}^{\text{Präv}} - p_{g^{+t}}^{\text{Präv}}}, \quad (\text{A.72})$$

$$p_{g^{+k}}^{\text{Präv}} \Big|_{\text{n. KHK.-M.}} = \frac{p_{g^{+k}}^{\text{Präv}}}{1 - p_{g^{+i}}^{\text{Präv}} - p_{g^{+t}}^{\text{Präv}}}, \quad (\text{A.73})$$

$$p_{g^{V+h+}} \Big|_{\text{n. KHK.-M.}} = \frac{p_{g^{V+h+}}}{1 - p_{g^{V+i}} - p_{g^{V+t}}}, \quad (\text{A.74})$$

$$p_{g^{V+k}} \Big|_{\text{n. KHK.-M.}} = \frac{p_{g^{V+k}}}{1 - p_{g^{V+i}} - p_{g^{V+t}}}, \quad (\text{A.75})$$

$$\left( p_{g^{+h+}}^{\text{Präv}} + p_{g^{+k}}^{\text{Präv}} \right) \Big|_{\text{n. KHK.-M.}} = \frac{p_{g^{+h+}}^{\text{Präv}} + p_{g^{+k}}^{\text{Präv}}}{1 - p_{g^{+i}}^{\text{Präv}} - p_{g^{+t}}^{\text{Präv}}}, \quad (\text{A.76})$$

$$(p_{g^{+h+}} + p_{g^{+k}}) \Big|_{\text{n. KHK.-M.}} = \frac{p_{g^{+h+}} + p_{g^{+k}}}{1 - p_{g^{+i}} - p_{g^{+t}}}, \quad (\text{A.77})$$

$$p_{g^{+t}}^{\text{Präv}} \Big|_{\text{n. s. M.}} = \frac{p_{g^{+t}}^{\text{Präv}}}{1 - p_{g^{+k}}^{\text{Präv}} - p_{g^{+i}}^{\text{Präv}}}, \quad (\text{A.78})$$

$$p_{g^{+t}} \Big|_{\text{n. s. M.}} = \frac{p_{g^{+t}}}{1 - p_{g^{+k}} - p_{g^{+i}}}, \quad (\text{A.79})$$

$$p_{g^{+i}}^{\text{Präv}} \Big|_{\text{n. i. M.}} = \frac{p_{g^{+i}}^{\text{Präv}}}{1 - p_{g^{+k}}^{\text{Präv}} - p_{g^{+t}}^{\text{Präv}}} \quad \text{und} \quad (\text{A.80})$$

$$p_{g^{+i}} \Big|_{\text{n. i. M.}} = \frac{p_{g^{+i}}}{1 - p_{g^{+k}} - p_{g^{+t}}}. \quad (\text{A.81})$$

$$(\text{A.82})$$

Die Übergangswahrscheinlichkeiten eines Zustandes addieren sich zu 1:

$$p_{g+g^+}^{\text{Präv}} + p_{g+g^{V+}}^{\text{Präv}} + p_{g+h^+}^{\text{Präv}} + p_{g+k}^{\text{Präv}} + p_{g+i}^{\text{Präv}} + p_{g+t}^{\text{Präv}} = 1 \quad \text{sowie} \quad (\text{A.83})$$

$$p_{g^{V+}g^{V+}} + p_{g^{V+h^+}} + p_{g^{V+k}} + p_{g^{V+i}} + p_{g^{V+t}} = 1. \quad (\text{A.84})$$

Die ursachenspezifischen Sterbewahrscheinlichkeiten ergeben die Gesamtsterbewahrscheinlichkeit:

$$p_{g+k}^{\text{Präv}} + p_{g+i}^{\text{Präv}} + p_{g+t}^{\text{Präv}} = p_{g+T}^{\text{Präv}} \quad \text{sowie} \quad (\text{A.85})$$

$$p_{g^{V+k}} + p_{g^{V+i}} + p_{g^{V+t}} = p_{g^{V+T}}. \quad (\text{A.86})$$

# Anhang B

## Tabellen

In Anhang B befinden sich Tabellen, die auf den Krankenkassendaten aus Abschnitt 5.1.3 basieren. Der Krankenkassendatensatz unterscheidet sich von den SOEP-Daten und den Daten des Bundesgesundheits surveys u. a. derart, dass er nicht der Öffentlichkeit für wissenschaftliche Zwecke zur Verfügung steht. Mit Hilfe der hier veröffentlichten Tabellen B.1 bis B.9 ist es möglich, unter Hinzunahme weiterer (öffentlich zugänglicher) Quellen die Analyseergebnisse dieser Dissertation zu reproduzieren.

	Alter	$\hat{\delta}_{\text{HI}}$ (Männer)	$SD(\hat{\delta}_{\text{HI}})$ (Männer)	$\hat{\delta}_{\text{HI}}$ (Frauen)	$SD(\hat{\delta}_{\text{HI}})$ (Frauen)
1	30	0,37192	5,386e-04	0,39660	5,232e-04
2	31	0,37000	4,432e-04	0,38032	4,694e-04
3	32	0,35780	4,038e-04	0,36514	4,091e-04
4	33	0,34185	3,505e-04	0,35471	3,581e-04
5	34	0,32996	2,894e-04	0,34944	3,033e-04
6	35	0,33494	2,464e-04	0,33993	2,657e-04
7	36	0,31693	2,048e-04	0,34339	2,229e-04
8	37	0,31503	1,796e-04	0,31203	1,910e-04
9	38	0,29969	1,515e-04	0,31330	1,652e-04
10	39	0,29447	1,291e-04	0,30766	1,405e-04
11	40	0,28730	1,092e-04	0,29782	1,182e-04
12	41	0,27774	9,405e-05	0,29472	1,033e-04
13	42	0,26870	7,998e-05	0,28269	8,665e-05
14	43	0,25874	6,782e-05	0,26799	7,317e-05
15	44	0,24694	5,651e-05	0,26609	6,302e-05
16	45	0,23920	4,864e-05	0,25627	5,378e-05
17	46	0,23247	4,224e-05	0,24617	4,612e-05
18	47	0,22212	3,627e-05	0,23834	3,996e-05
19	48	0,21684	3,215e-05	0,22828	3,474e-05
20	49	0,20758	2,782e-05	0,22348	3,119e-05
21	50	0,20044	2,464e-05	0,21455	2,709e-05
22	51	0,19276	2,171e-05	0,20803	2,412e-05
23	52	0,18585	1,903e-05	0,20262	2,158e-05
24	53	0,18259	1,720e-05	0,19605	1,900e-05
25	54	0,17735	1,512e-05	0,19065	1,680e-05
26	55	0,17276	1,320e-05	0,18748	1,483e-05
27	56	0,16780	1,142e-05	0,18270	1,294e-05
28	57	0,16396	1,012e-05	0,17872	1,144e-05
29	58	0,16088	9,030e-06	0,17629	1,032e-05
30	59	0,15832	8,191e-06	0,17250	9,240e-06
31	60	0,15542	7,488e-06	0,16752	8,317e-06

Tabelle B.1: Schätzer für den Anteil  $\delta_{\text{HI}}$  um den die Herzinsuffizienz verringert werden kann, wenn bei Personen mit KHK die KHK vermieden wird, sowie der Standardfehler des Schätzers – bei 30- bis 60-jährigen Männern und Frauen

	Alter	$\hat{\delta}_{\text{HI}}$ (Männer)	$SD(\hat{\delta}_{\text{HI}})$ (Männer)	$\hat{\delta}_{\text{HI}}$ (Frauen)	$SD(\hat{\delta}_{\text{HI}})$ (Frauen)
32	61	0,15334	6,873e-06	0,16517	7,586e-06
33	62	0,15135	6,343e-06	0,16223	6,915e-06
34	63	0,14919	5,796e-06	0,15965	6,315e-06
35	64	0,14666	5,300e-06	0,15623	5,677e-06
36	65	0,14682	4,980e-06	0,15416	5,224e-06
37	66	0,14548	4,680e-06	0,15162	4,819e-06
38	67	0,14604	4,556e-06	0,14940	4,523e-06
39	68	0,14600	4,422e-06	0,14850	4,342e-06
40	69	0,14501	4,296e-06	0,14574	4,096e-06
41	70	0,14446	4,152e-06	0,14427	3,895e-06
42	71	0,14500	4,039e-06	0,14285	3,676e-06
43	72	0,14425	3,849e-06	0,14009	3,387e-06
44	73	0,14499	3,712e-06	0,13807	3,122e-06
45	74	0,14420	3,566e-06	0,13574	2,889e-06
46	75	0,14266	3,520e-06	0,13241	2,731e-06
47	76	0,14173	3,590e-06	0,12970	2,693e-06
48	77	0,13980	3,715e-06	0,12729	2,733e-06
49	78	0,13578	3,833e-06	0,12333	2,782e-06
50	79	0,13509	4,082e-06	0,12012	2,866e-06
51	80	0,13032	4,280e-06	0,11545	2,925e-06
52	81	0,12901	4,809e-06	0,11213	3,056e-06
53	82	0,12728	5,176e-06	0,10961	3,126e-06
54	83	0,12267	5,108e-06	0,10679	3,073e-06
55	84	0,12038	5,030e-06	0,10259	2,914e-06
56	85	0,11511	4,997e-06	0,09917	2,915e-06
57	86	0,11334	5,656e-06	0,09362	3,133e-06
58	87	0,11215	6,918e-06	0,09184	3,723e-06
59	88	0,10527	8,273e-06	0,08868	4,508e-06
60	89	0,10685	1,099e-05	0,08524	5,455e-06
61	90	0,10525	1,063e-05	0,08742	6,063e-06

Tabelle B.2: Schätzer für den Anteil  $\delta_{\text{HI}}$  um den die Herzinsuffizienz verringert werden kann, wenn bei Personen mit KHK die KHK vermieden wird, sowie der Standardfehler des Schätzers – bei 61- bis 90-jährigen Männern und Frauen

	Alter	$\hat{Q}_1$ (Männer)	$SD(\hat{Q}_1)$ (Männer)	$\hat{Q}_1$ (Frauen)	$SD(\hat{Q}_1)$ (Frauen)
1	30	7,377	0,7897	6,152	0,3979
2	31	7,598	0,5251	6,834	0,4646
3	32	6,493	0,3835	6,648	0,4910
4	33	6,463	0,3659	6,838	0,3407
5	34	8,035	0,4155	6,476	0,2618
6	35	7,232	0,2563	6,159	0,2032
7	36	7,345	0,2285	6,059	0,1416
8	37	6,834	0,1612	6,649	0,1817
9	38	7,231	0,1617	6,434	0,1302
10	39	7,111	0,1313	5,983	0,1043
11	40	6,392	0,0839	6,137	0,0903
12	41	6,562	0,0765	5,716	0,0682
13	42	6,592	0,0706	5,922	0,0668
14	43	6,422	0,0565	5,821	0,0530
15	44	6,452	0,0475	5,484	0,0366
16	45	6,125	0,0380	5,381	0,0316
17	46	6,156	0,0329	5,478	0,0296
18	47	5,908	0,0276	5,278	0,0230
19	48	5,714	0,0220	5,243	0,0203
20	49	5,543	0,0179	4,978	0,0153
21	50	5,348	0,0148	4,755	0,0124
22	51	5,170	0,0125	4,660	0,0106
23	52	4,936	0,0105	4,442	0,0087
24	53	4,711	0,0089	4,311	0,0077
25	54	4,500	0,0075	4,095	0,0064
26	55	4,343	0,0060	3,919	0,0051
27	56	4,151	0,0048	3,812	0,0042
28	57	4,005	0,0041	3,616	0,0034
29	58	3,877	0,0034	3,496	0,0029
30	59	3,708	0,0029	3,375	0,0024
31	60	3,586	0,0026	3,454	0,0025

Tabelle B.3: Schätzer für den Quotienten  $Q_1$  – bei 30- bis 60-jährigen Männern und Frauen

	Alter	$\hat{Q}_1$ (Männer)	SD( $\hat{Q}_1$ ) (Männer)	$\hat{Q}_1$ (Frauen)	SD( $\hat{Q}_1$ ) (Frauen)	
	32	61	3,301	0,0022	3,055	0,0018
	33	62	3,213	0,0020	2,945	0,0017
	34	63	3,102	0,0019	2,870	0,0016
	35	64	3,018	0,0018	2,828	0,0015
	36	65	2,996	0,0017	2,812	0,0015
	37	66	2,819	0,0015	2,690	0,0013
	38	67	2,773	0,0015	2,651	0,0013
	39	68	2,691	0,0015	2,615	0,0012
	40	69	2,630	0,0015	2,574	0,0012
	41	70	2,572	0,0014	2,514	0,0011
	42	71	2,493	0,0014	2,460	0,0010
	43	72	2,394	0,0013	2,381	0,0009
	44	73	2,323	0,0012	2,327	0,0008
	45	74	2,249	0,0011	2,252	0,0007
	46	75	2,197	0,0010	2,196	0,0006
	47	76	2,101	0,0010	2,108	0,0006
	48	77	2,048	0,0011	2,041	0,0005
	49	78	1,993	0,0012	1,983	0,0006
	50	79	1,921	0,0012	1,908	0,0006
	51	80	1,866	0,0013	1,861	0,0006
	52	81	1,826	0,0014	1,826	0,0007
	53	82	1,816	0,0015	1,794	0,0007
	54	83	1,769	0,0014	1,755	0,0006
	55	84	1,771	0,0014	1,734	0,0006
	56	85	1,782	0,0015	1,734	0,0006
	57	86	1,761	0,0018	1,709	0,0006
	58	87	1,756	0,0023	1,707	0,0007
	59	88	1,759	0,0032	1,715	0,0010
	60	89	1,782	0,0045	1,714	0,0013
	61	90	1,801	0,0060	1,740	0,0018

Tabelle B.4: Schätzer für den Quotienten  $Q_1$  – bei 61- bis 90-jährigen Männern und Frauen

	Alter	$\hat{Q}_2$ (Männer)	$SD(\hat{Q}_2)$ (Männer)	$\hat{Q}_2$ (Frauen)	$SD(\hat{Q}_2)$ (Frauen)
1	30	0,6864	3e-04	0,6867	3e-04
2	31	0,6879	2e-04	0,6915	3e-04
3	32	0,6855	2e-04	0,6655	3e-04
4	33	0,6824	2e-04	0,6910	2e-04
5	34	0,6843	2e-04	0,6709	2e-04
6	35	0,6838	2e-04	0,6696	2e-04
7	36	0,6774	1e-04	0,6746	1e-04
8	37	0,6738	1e-04	0,6614	2e-04
9	38	0,6658	1e-04	0,6612	1e-04
10	39	0,6653	1e-04	0,6540	1e-04
11	40	0,6575	1e-04	0,6519	1e-04
12	41	0,6531	1e-04	0,6491	1e-04
13	42	0,6474	1e-04	0,6370	1e-04
14	43	0,6484	1e-04	0,6318	1e-04
15	44	0,6426	1e-04	0,6401	1e-04
16	45	0,6398	1e-04	0,6341	1e-04
17	46	0,6382	1e-04	0,6294	1e-04
18	47	0,6304	1e-04	0,6291	1e-04
19	48	0,6281	0e+00	0,6207	0e+00
20	49	0,6224	0e+00	0,6164	0e+00
21	50	0,6151	0e+00	0,6124	0e+00
22	51	0,6061	0e+00	0,6052	0e+00
23	52	0,5975	0e+00	0,5949	0e+00
24	53	0,5867	0e+00	0,5842	0e+00
25	54	0,5754	0e+00	0,5721	0e+00
26	55	0,5622	0e+00	0,5591	0e+00
27	56	0,5468	0e+00	0,5443	0e+00
28	57	0,5322	0e+00	0,5314	0e+00
29	58	0,5195	0e+00	0,5191	0e+00
30	59	0,5073	0e+00	0,5079	0e+00
31	60	0,4943	0e+00	0,4917	0e+00

Tabelle B.5: Schätzer für den Quotienten  $Q_2$  – bei 30- bis 60-jährigen Männern und Frauen

	Alter	$\hat{Q}_2$ (Männer)	SD( $\hat{Q}_2$ ) (Männer)	$\hat{Q}_2$ (Frauen)	SD( $\hat{Q}_2$ ) (Frauen)	
	32	61	0,4840	0e+00	0,4831	0e+00
	33	62	0,4727	0e+00	0,4738	0e+00
	34	63	0,4639	0e+00	0,4648	0e+00
	35	64	0,4544	0e+00	0,4579	0e+00
	36	65	0,4491	0e+00	0,4548	0e+00
	37	66	0,4456	0e+00	0,4504	0e+00
	38	67	0,4400	0e+00	0,4442	0e+00
	39	68	0,4373	0e+00	0,4421	0e+00
	40	69	0,4309	0e+00	0,4360	0e+00
	41	70	0,4274	0e+00	0,4322	0e+00
	42	71	0,4211	0e+00	0,4246	0e+00
	43	72	0,4151	0e+00	0,4165	0e+00
	44	73	0,4050	0e+00	0,4072	0e+00
	45	74	0,3932	0e+00	0,3958	0e+00
	46	75	0,3800	0e+00	0,3818	0e+00
	47	76	0,3677	0e+00	0,3690	0e+00
	48	77	0,3510	0e+00	0,3539	0e+00
	49	78	0,3369	0e+00	0,3379	0e+00
	50	79	0,3241	0e+00	0,3223	0e+00
	51	80	0,3122	0e+00	0,3116	0e+00
	52	81	0,2999	1e-04	0,2984	1e-04
	53	82	0,2904	1e-04	0,2887	1e-04
	54	83	0,2847	1e-04	0,2822	1e-04
	55	84	0,2766	1e-04	0,2751	1e-04
	56	85	0,2758	1e-04	0,2746	1e-04
	57	86	0,2755	1e-04	0,2745	1e-04
	58	87	0,2808	1e-04	0,2761	1e-04
	59	88	0,2833	1e-04	0,2778	1e-04
	60	89	0,2863	2e-04	0,2825	1e-04
	61	90	0,2935	2e-04	0,2906	2e-04

Tabelle B.6: Schätzer für den Quotienten  $Q_2$  – bei 61- bis 90-jährigen Männern und Frauen

---

Alter	Männer	Frauen
1	1,017e-04	1,736e-04
2	8,963e-05	1,309e-04
3	5,983e-05	1,355e-04
4	8,292e-05	3,877e-05
5	9,774e-05	3,749e-05
6	5,438e-05	8,595e-05
7	9,113e-05	3,317e-05
8	1,736e-05	1,692e-05
9	2,603e-05	3,984e-05
10	4,584e-05	1,876e-05
11	4,679e-05	7,830e-05
12	3,584e-05	8,272e-05
13	7,989e-05	4,092e-05
14	6,627e-05	1,496e-04
15	6,666e-05	1,302e-04
16	9,954e-05	1,706e-04
17	1,306e-04	2,964e-04
18	1,857e-04	2,378e-04
19	1,863e-04	3,271e-04
20	1,988e-04	3,804e-04
21	3,355e-04	2,846e-04
22	2,069e-04	3,051e-04
23	4,072e-04	4,373e-04
24	3,149e-04	3,775e-04
25	3,439e-04	4,080e-04
26	3,544e-04	4,434e-04
27	4,532e-04	5,213e-04
28	5,509e-04	5,400e-04
29	7,052e-04	6,461e-04
30	8,397e-04	6,092e-04

---

Tabelle B.7: Prävalenz der KHK, geschätzt aus den RSA-Krankenkassendaten (1- bis 30-jährige Männer und Frauen)

---

Alter	Männer	Frauen
31	9,602e-04	6,495e-04
32	1,211e-03	7,830e-04
33	1,127e-03	8,838e-04
34	1,425e-03	1,074e-03
35	1,608e-03	9,682e-04
36	2,368e-03	1,423e-03
37	2,656e-03	1,464e-03
38	3,166e-03	1,746e-03
39	3,801e-03	1,978e-03
40	4,937e-03	2,601e-03
41	5,492e-03	2,698e-03
42	6,472e-03	3,386e-03
43	7,576e-03	4,000e-03
44	9,794e-03	4,847e-03
45	1,097e-02	5,616e-03
46	1,290e-02	6,681e-03
47	1,462e-02	7,571e-03
48	1,692e-02	8,512e-03
49	2,030e-02	9,968e-03
50	2,304e-02	1,118e-02
51	2,540e-02	1,284e-02
52	2,972e-02	1,506e-02
53	3,315e-02	1,727e-02
54	3,741e-02	1,990e-02
55	4,221e-02	2,391e-02
56	4,852e-02	2,634e-02
57	5,177e-02	2,988e-02
58	5,678e-02	3,313e-02
59	6,210e-02	3,641e-02
60	6,575e-02	3,734e-02

---

Tabelle B.8: Prävalenz der KHK, geschätzt aus den RSA-Krankenkassendaten (31- bis 60-jährige Männer und Frauen)

---

Alter	Männer	Frauen
61	7,887e-02	5,240e-02
62	8,806e-02	5,980e-02
63	9,663e-02	6,782e-02
64	1,083e-01	7,372e-02
65	1,140e-01	7,725e-02
66	1,304e-01	9,006e-02
67	1,440e-01	1,018e-01
68	1,572e-01	1,110e-01
69	1,695e-01	1,195e-01
70	1,835e-01	1,286e-01
71	1,976e-01	1,412e-01
72	2,102e-01	1,525e-01
73	2,212e-01	1,648e-01
74	2,315e-01	1,762e-01
75	2,454e-01	1,873e-01
76	2,507e-01	1,961e-01
77	2,633e-01	2,117e-01
78	2,735e-01	2,266e-01
79	2,848e-01	2,362e-01
80	2,899e-01	2,494e-01
81	2,965e-01	2,586e-01
82	3,126e-01	2,721e-01
83	3,156e-01	2,808e-01
84	3,257e-01	2,873e-01
85	3,228e-01	2,909e-01
86	3,241e-01	2,896e-01
87	3,181e-01	2,922e-01
88	3,267e-01	2,921e-01
89	3,176e-01	2,889e-01
90	3,040e-01	2,751e-01

---

Tabelle B.9: Prävalenz der KHK, geschätzt aus den RSA-Krankenkassendaten (61- bis 90-jährige Männer und Frauen)

# Anhang C

## Programmierung

Die im Rahmen dieser Dissertation geschriebenen Computerprogramme sind leider zu umfangreich, um hier abgedruckt werden zu können. Sie sind jedoch nach Absprache beim Autor einsehbar ([b.stollenwerk@gmx.de](mailto:b.stollenwerk@gmx.de)).

Der größte Teil der Programmierung erfolgte mit der statistischen Software R (Version 2.1.0) (R Development Core Team 2005). Die umfangreichen Gleichungssysteme aus Kapitel 4 wurden mit Hilfe der mathematischen Software MAPLE (Version 9.01) gelöst. Bevor die SOEP-Daten und die Krankenkassendaten weiter mit R analysiert wurden, wurden sie mit SAS (Version 9.1, deutsch) aufbereitet und aggregiert.

# Literatur

Abramowitz, M. und I. A. Stegun (1972). *Handbook of Mathematical Functions*. New York: Dover.

Anderson, K. M., P. M. Odell, P. W. F. Wilson und W. B. Kannel (1990). Cardiovascular disease risk profiles. *American Heart Journal* 121, 293–298.

Anderson, K. M., P. W. F. Wilson, P. M. Odell und W. B. Kannel (1991). An updated coronary risk profile: A statement for health professionals. *Circulation* 83, 356–362.

Andrawes, W. F., C. Bussy und J. Belmin (2005). Prevention of cardiovascular events in elderly people. *Drugs and Aging* 22(10), 859–876.

Assmann, G., P. Cullen und H. Schulte (2002). Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation* 105, 310–315.

Bates, D. M. und D. G. Watts (1988). *Nonlinear Regression Analysis and Its Applications*. New York: John Wiley & Sons.

Breyer, F. und P. Zweifel (1997). *Gesundheitsökonomie*. Berlin: Springer.

Bundesversicherungsamt (2005). Verordnung über das Verfahren zum Risikostrukturausgleich in der gesetzlichen Krankenversicherung (Risikostruktur-Ausgleichsverordnung – RSAV). (Stand: 18.1.2005).

Dahlof, B., R. B. Devereux, S. E. Kjeldsen, S. Julius, G. Beevers, U. Faire, F. Fyhr-

- quist, H. Ibsen, K. Kristiansson, O. Lederballe-Pedersen, L. H. Lindholm, M. S. Nieminen, P. Omvik, S. Oparil und H. Wedel (2002). Cardiovascular morbidity and mortality in the losartan intervention for endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. *The Lancet* 359, 995–1003.
- Daviglus, M. L., J. Stamler, A. Pirzada, L. L. Yan, D. B. Garside, K. Liu, R. Wang, A. R. Dyer, D. M. Lloyd-Jones und P. Greenland (2004). Favorable cardiovascular risk profile in young women and long-term risk of cardiovascular and all-cause mortality. *Journal of the American Medical Association* 292(13), 1588–1592.
- Delaunay, B. (1934). Sur la sphère vide. *Izvestia Akademia Nauk SSSR* 7, 793–800.
- Delea, T. E., T. A. Jacobson, P. W. J. C. Serruys, J. S. Edelsberg und G. Oster (2005). Cost-effectiveness of Fluvastatin following successful first percutaneous coronary intervention. *The Annals of Pharmacotherapy* 39, 610–616.
- Deutsches Institut für Wirtschaftsforschung (2006). Das Sozio-oekonomische Panel. <http://www.diw.de/deutsch/sop/uebersicht/index.html>. (Abfrage: Februar 2006).
- Dobson, A. J. (2002). *An Introduction to Generalized Linear Models* (2. Auflage). Boca Raton, London, New York: Chapman & Hall/CRC.
- Drummond, M. F. und T. O. Jefferson (1996). Guidelines for authors and peer reviewers of economic submissions to the BMJ. *British Medical Journal* 313, 275–283.
- Drummond, M. F., M. J. Sculpher, G. W. Torrance, B. J. O’Brien und G. L. Stoddart (2005). *Methods for the Economic Evaluation of Health Care Programmes* (3. Auflage). Oxford: Oxford University Press.
- Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in solobev spaces. In W. Schemp und K. Zeller (Hrsg.), *Construction Theory of Functions*

- of Several Variables*, Seiten 85–100. Berlin: Springer.
- Efron, B. und R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability 57. New York, London: Chapman & Hall.
- Gebhardt, A. (2005). *The tripack Package – Triangulation of irregularly spaced data*. Fortran code by R. J. Renka. R functions by Albrecht Gebhardt. Version 1.2-10. <http://cran.r-project.org>.
- Gold, M. R., J. E. Siegel, L. B. Russell und M. C. Weinstein (Hrsg.) (1996). *Cost-Effectiveness in Health and Medicine*. New York, Oxford: Oxford University Press.
- Goldman, L., A. M. Garber, S. A. Grover und M. A. Hlatky (1996). 27th Bethesda conference: matching the intensity of risk factor management with the hazard for coronary disease events. task force 6. cost effectiveness of assessment and management of risk factors. *Journal of the American College of Cardiology* 27(5), 1020–30.
- Golub, G. H., M. Heath und G. Wahba (1979). Generalized cross validation as a method for choosing a good ridge parameter. *Technometrics* 21(2), 215–223.
- Green, P. J. und B. W. Silverman (1994). *Nonparametric Regression and Generalized Linear*. London, New York: Chapman & Hall.
- Hafner, R. (1989). *Wahrscheinlichkeitsrechnung und Statistik*. Wien: Springer-Verlag.
- Hartung, J. und B. Elpelt (1999). *Multivariate Statistik: Lehr- und Handbuch der angewandten Statistik* (6. Auflage). München, Wien: Oldenbourg.
- Hartung, J., B. Elpelt und K.-H. Klösener (2005). *Statistik: Lehr- und Handbuch der angewandten Statistik* (14. Auflage). München, Wien: Oldenbourg.
- Hastie, T. J. und R. J. Tibshirani (1990). *Generalized Additive Models*

- (1. Auflage), Band 43 aus *Monographs on Statistics and Applied Probability*. London, New York: Chapman & Hall.
- Hay, J. W. und K. L. Sterling (2005). Cost effectiveness of treating low HDL-cholesterol in the primary prevention of coronary heart disease. *Pharmacoeconomics* 23(2), 133–141.
- Hayashino, Y., S. Nagata-Kobayashi, T. Morimoto, K. Maeda, T. Shimbo und T. Fukui (2004). Cost-effectiveness of screening for coronary artery disease in asymptomatic patients with type 2 diabetes and additional atherogenic risk factors. *Journal of General Internal Medicine* 19, 1181–1191.
- Hoppe, U. C., M. Böhm, R. Dietz, P. Hanrath, H. K. Kroemer, A. Osterspey, A. A. Schmaltz und E. Erdmann (2005). Leitlinien zur Therapie der chronischen Herzinsuffizienz. *Leitlinien zur Therapie der chronischen Herzinsuffizienz* 94(8), 488–509.
- Hu, F. B., E. Cho, K. M. Rexrode, C. M. Albert und J. E. Manson (2003). Fish and long-chain omega-3 fatty acid intake and risk of coronary heart disease and total mortality in diabetic women. *Circulation* 107(14), 1852–1857.
- Jousilahti, P., E. Vartiainen, J. Tuomilehto und P. Puska (1999). Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14 786 middle-aged men and women in Finland. *Circulation* 99(9), 1165–1172.
- Kattainen, A., A. Reunanen, S. Koskinen, T. Martelin, P. Knekt und A. Aromaa (2004). Disability predicted mortality in men but not women with coronary heart disease. *Journal of Clinical Epidemiology* 57, 513–521.
- Kleinbaum, D. G. (1994). *Logistic Regression. A Self-Learning Text*. Statistics in the Health Sciences. New York, Berlin, Heidelberg: Springer.
- Klever-Deichert, G., B. Hinzpeter, G. Wendland und K. Lauterbach (2000). Kosten-Nutzen-Analyse einer evidenzbasierten Sekundärprävention koronarer

- Herzkrankheiten durch Statine. *Medizinische Klinik* 95(6), 305–313.
- Klever-Deichert, G., B. Stollenwerk, A. Gerber, G. Wendland und K. Lauterbach (2006). Gesundheitsökonomie in der kardiovaskulären Prävention mit Lipidsenkern. In P. Schwandt, W. O. Richter und K. G. Parhofer (Hrsg.), *Handbuch der Fettstoffwechselstörungen*. Schattauer.
- Lange, A. P. und T. D. Szucs (2004). Kosteneffektivität von Atorvastatin bei der Behandlung des akuten Koronarsyndroms in Deutschland anhand der MIRACL-Studie. *Medizinische Klinik* 99(9), 500–505.
- Lauterbach, K. W., M. Lungen, B. Stollenwerk, A. Gerber und G. Klever-Deichert (2005). Auswirkungen einer Bürgerversicherung in der Pflegeversicherung. *Gesundheitsökonomie und Qualitätsmanagement* 10, 221–230.
- Lauterbach, K. W., E. Wille, M. Lungen, S. Stock, G. Wendland, H. Cischinsky und S. Resch (2001). Modell eines fairen Wettbewerbs durch den Risikostrukturausgleich. Technischer Bericht, Institut für Gesundheitsökonomie und Klinische Epidemiologie (IGKE) der Universität zu Köln und Lehrstuhl für Volkswirtschaftslehre der Universität Mannheim, Köln, Mannheim. Gutachten im Auftrag des Verbandes der Angestellten-Krankenkassen e. V. (VdAK), des Arbeiter-Ersatzkassen-Verbandes e. V. (AEV), dem AOK-Bundesverband (AOK-BV) und dem IKK-Bundesverband (IKK-BV).
- Lexis, W. (1879). Über die Theorie der Stabilität statistischer Reichen. *Jahrbücher für Nationalökonomie und Statistik* 32, 60–98.
- Lindgren, P., M. Buxton, T. Kahan, N. R. Poulter, B. Dahlöf, P. S. Sever, H. Wedel und B. Jönsson (2005). Cost-effectiveness of atorvastatin for the prevention of coronary and stroke events: an economic analysis of the anglo-scandinavian cardiac outcomes trial – lipid-lowering arm (ASCOT-LLA). *The European Society of Cardiology* 12(1), 29–36.
- Löllgen, H. (2003). Primärprävention kardialer Erkrankungen. Stellenwert der

- körperlichen Aktivität. *Deutsches Ärzteblatt* 100(15), A987–996.
- Lowe, L. P., P. Greenland, K. J. Ruth, A. R. Dyer, R. Stamler und J. Stamler (1998). Impact of major cardiovascular disease risk factors, particularly in combination on 22-year mortality in women and men. *Archives of Internal Medicine* 158, 2007–2014.
- Maclaine, G. und H. Patel (2001). A cost-effectiveness model of alternative statins to achieve target LDL-cholesterol levels. *International Journal of Clinical Practice* 55(4), 243–249.
- McCullagh, P. und J. A. Nelder (1989). *Generalized Linear Models* (2. Auflage). New York: Chapman & Hall.
- McMurray, J. J. und S. Stewart (2000). Epidemiology, aetiology, and prognosis of heart failure. *Heart* 83, 596–602.
- Monagan, M. B., K. O. Geddes, K. M. Heal, G. Labahn, S. M. Vorkoetter, J. McCarron und P. DeMarco (2003). *Maple 9: Introductory Programming Guide*. Waterloo Maple Inc.: Maplesoft. ISBN 1-894511-43-3.
- O’Malley, P. G., B. A. Greenberg und A. J. Taylor (2004). Cost-effectiveness of using electron beam computed tomography to identify patients at risk for clinical coronary artery disease. *American Heart Journal* 148(1), 106–113.
- Petitti, D. B. (1994). *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis*, Band 24 aus *Monographs in Epidemiology and Biostatistics*. New York, Oxford: Oxford University Press.
- Pschyrembel, W. (1990). *Klinisches Wörterbuch*. Berlin: de Gruyter.
- R Development Core Team (2005). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Renka, R. J. (1996). Algorithm 751: TRIPACK: a constrained two-dimensional

- Delaunay triangulation package. *ACM Transactions on Mathematical Software* 22, 1–8.
- Robbins, J., J. C. Nelson, P. M. Rautaharju und J. S. Gottdiener (2003). The association between the length of the QT interval and mortality in the cardiovascular health study. *The American Journal of Medicine* 115, 689–694.
- Rosengren, A., A. Dotevall, H. Eriksson und L. Wilhelmsen (2001). Optimal risk factors in the population: prognosis, prevalence, and secular trends. *European Heart Journal* 22, 136–144.
- Rosengren, A., L. Wilhelmsen, M. Hagman und H. Wedel (1998). Natural history of myocardial infarction and angina pectoris in a general population sample of middle-aged men: a 16-year follow-up of the Primary Prevention Study, Göteborg, Sweden. *Journal of Internal Medicine* 244, 495–505.
- Schumacher, M. und G. Schulgen (2002). *Methodik klinischer Studien*. Berlin, Heidelberg: Springer.
- Schwabe, U. und D. Paffrath (2004). *Arzneiverordnungs-Report 2004: Aktuelle Daten, Kosten, Trends und Kommentare*. Springer-Verlag.
- Sonnenberg, F. A. und J. R. Beck (1993). Markov models in medical decision making: A practical guide. *Medical Decision Making* 13, 322–338.
- Stamler, J., R. Stamler, J. D. Neaton, D. Wentworth, M. L. Daviglus, D. Garside, A. R. Dyer, K. Liu und P. Greenland (1999). Low risk-factor profile and long-term cardiovascular and noncardiovascular mortality and life expectancy. *Journal of the American Medical Association* 282(21), 2012–2018.
- Statistisches Bundesamt (2005). *Gesundheitswesen - Todesursachen in Deutschland 2003*. Wiesbaden.
- Stolzenberg, H. (2000). Bundes-Gesundheitssurvey 1998: Public use file BGS98. Technischer Bericht, Robert Koch Institut, Berlin. Dokumentation des Datensatzes.

- Stolzenberg, H. (2002). *Gesundheitssurveys in Deutschland*. Berlin: RKI.
- Szucs, T. D., G. Klose und R. Düsing (2004). Kosteneffektivität von Atorvastatin zur Prävention der koronaren Herzkrankheit. *Deutsche Medizinische Wochenschrift* 129, 1420–1424.
- Torres, M. S., S. M. Calderón, I. R. Díaz, A. B. Chacón, F. L. Fernández und I. F. Martínez (2004). Health-related quality of life in coronary heart disease compared to norms in Spanish population. *Quality of Life Research* 13, 1401–1407.
- Venables, W. N. und B. D. Ripley (2002). *Modern Applied Statistics with S* (4. Auflage). New York, Berlin, Heidelberg: Springer.
- Vrečer, M., S. Turk, J. Drinovec und A. Mrhar (2003). Use of statins in primary and secondary prevention of coronary heart disease and ischemic stroke: Meta-analysis of randomized trials. *International Journal of Clinical Pharmacology and Therapeutics* 41(12), 567–577.
- Wahba, G. (1990). Spline models for observational data. Philadelphia: SIAM.
- Wannamethee, S. G., A. G. Shaper und L. Lennon (2004). Cardiovascular disease incidence and mortality in older men with diabetes and in men with coronary heart disease. *Heart* 90, 1398–1403.
- Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61(3), 439–447.
- Weinstein, M. C., P. G. Coxson, L. W. Williams, T. M. Pass, W. B. Stason und L. Goldman (1987). Forecasting coronary heart disease incidence, mortality, and cost: The coronary heart disease policy model. *American Journal of Public Health* 77(11), 1417–1426.
- Wendland, G. (2002). *Kosteneffektivitätsanalyse der Statintherapie zur Primärprävention der koronaren Herzkrankheit in Deutschland*. Dissertation, Universität zu Köln, medizinische Fakultät.

- 
- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society B* 65(1), 95–114.
- Wood, S. N. (2006). *Generalized Additive Models: an introduction with R*. Boca Raton, London: Chapman & Hall/CRC.