Anhang 9

Bias in ökologischen Studien bei nichtlinearen Risikomodellen

Bias in ökologischen Studien bei nichtlinearen Risikomodellen

J.C. Kaiser

GSF - Institut für Strahlenschutz,
Neuherberg, Deutschland

Dezember 2004

Linear-quadratisches Modell

Wenn die Dosis-Wirkungsbeziehung von einer einfachen linearen Form abweicht, ist zu erwarten, dass die Risikofaktoren mit zusammen gefassten Daten nur verzerrt geschätzt werden können, auch wenn keine Confounder wie z.B. Screening vorliegen. Zu dieser Fragestellung wurden Simulationsrechungen durchgeführt mit einer linear-quadratischen Dosis-Wirkungsbeziehung. Für eine Person *j* aus einer Ortschaft *i* und einer individuellen Schilddrüsendosis D_{ij} hat sie die Form

$$h_{ij} = h_0 + bD_{ij} + cD_{ij}^2 \qquad (1)$$

mit dem individuellen Risiko h_{ij}. In diesem Modell setzt sich das Gesamtrisiko zusammen aus dem konstanten Hintergrundrisiko h_0 und einem strahlenbedingten Risiko, das wiederum aus zwei Beträgen besteht, die linear und quadratisch von der individuellen Schilddrüsendosis D_{ij} abhängen. Die dosisabhängigen Beiträge werden mit den Koeffizienten *b* und *c* gewichtet.

Die Simulationen basieren auf den Abschätzungen der Ortsdosismittelwerte für 670 Ortschaften mit N_{pop} = 1002706 Personen aus der Geburtskohorte 1968-85 (Likhtarov et al. 2004). Die mittlere Dosis

$$\langle D \rangle = \frac{1}{N_{pop}} \sum_{ij} D_{ij} \qquad (2)$$

für diese Personengruppe beträgt 0,080 Gy. Die Verteilungsfunktion für die Dosis besitzt eine arithmetische Standardabweichung von σ = 0,304 Gy.

Für das Hintergrundrisiko h_0 wurde der Wert von 14,73 Fällen pro 10^6 PY angenommen. Es ist für alle Personen gleich. Für ein lineares Risikomodell würde ein EARPD β von 2,511 Fällen pro 10^4 PY Gy im Zeitraum ΔT = 1990-99 350 Neuerkrankungen vorhersagen. Diese Anzahl entspricht genau der Anzahl der tatsächlich registrierten Fälle. Sie soll auch mit dem linear-quadratischen Modell vorhergesagt werden.

Die Koeffizienten *b* und *c* werden dazu wie folgt ermittelt. Zunächst wird eine Abweichung $r = b/\beta$ vom EARPD β des linearen Modells festgelegt, die zugleich den linearen Koeffizienten *b* des linear-quadratischen Modells (1) bestimmt. Danach soll der quadratische Koeffizient *c* so angepasst werden, dass die Anzahl von 350 registrierten Neuerkrankungen vorhergesagt wird. Im Folgenden werden dazu zwei Methoden vorgestellt. Mit der Poisson-Approximation kann *c* sehr einfach analytisch bestimmt werden, wenn die Risikofunktion nur kleine, positive Werte annimmt. Wenn diese Vorbedingung nicht erfüllt ist, muss die Anpassung mit einer numerisch exakten Rechnung durchgeführt werden.

Die Anzahl *n* der zu erwartenden Neuerkrankungen

$$n = N_{pop}(1 - \langle S \rangle) \qquad (3)$$

ergibt sich, wenn man die mittlere Wahrscheinlichkeit $1-\langle S\rangle$, im Zeitraum ΔT an Schilddrüsenkrebs zu erkranken mit der Gesamtzahl der Person N_{pop} multipliziert. Der bevölkerungsbezogene Mittelwert der Wahrscheinlichkeit, *nicht* zu erkranken, ist definiert als

$$\langle S\rangle = \frac{1}{N_{pop}}\sum_{ij} S_{ij} \text{ mit der individuellen Wahrscheinlichkeit } S_{ij} = \exp\left(-\Delta T h_{ij}\right) \quad (4)$$

für eine Person j aus einer Ortschaft i.

1. Poisson-Approximation

Falls das Argument der Exponentialfunktion (4) $\Delta T h_{ij} << 1$, kann man sie annähern durch den Ausdruck $1-\Delta T h_{ij}$, der nun die Bedeutung einer Erkrankungswahrscheinlichkeit annimmt. Diese Wahrscheinlichkeit ist sehr klein und muss positiv bleiben. In dieser Näherung ergibt sich die Zahl der zu erwartenden Fälle aus

$$n' = N_{pop}\Delta T\langle h\rangle \text{ mit } \langle h\rangle = h_0 + b\langle D\rangle + c'\langle D^2\rangle. \quad (5)$$

Der Zahlenwerte für quadratischen Koeffizienten c' und damit auch für die Zahl der zu erwartenden Fälle n' können in der Poisson-Approximation von denen aus der exakten Berechnung abweichen. Deshalb werden sie durch eine gestrichene Notation unterschieden. Nun lässt sich der quadratische Koeffizient c' leicht bestimmen, wenn man fordert, dass die Zahl der zu erwartenden Neuerkrankungen im linearen und im linear-quadratischen Risikomodell gleich sein muss. Wenn die Risiken aus beiden Modellen hinreichend klein und positiv sind, ergibt sich aus dieser Forderung die Bestimmungsgleichung

$$h_0 + \beta\langle D\rangle = h_0 + r\beta\langle D\rangle + c'\langle D^2\rangle \text{ mit } r = \frac{b}{\beta}, \quad (6)$$

und der quadratische Koeffizient ergibt sich zu

$$c' = (1-r)\beta\frac{\langle D\rangle}{\sigma^2 + \langle D\rangle^2} \text{ mit } \langle D^2\rangle = \sigma^2 + \langle D\rangle^2. \quad (7)$$

Durch die Forderung nach der Vorhersage der gleichen Anzahl von zu erwartenden Neuerkrankungen haben die lineare und die linear-quadratische Dosis-Wirkungsbeziehung zwei Schnittpunkte. Der erste Schnittpunkt liegt bei der Dosis null, wo beide Beziehungen den Wert h_0 annehmen. Der zweite Schnittpunkt bei einer Dosis $D'_s > 0$. Nachdem die Koeffizienten der linear-quadratischen Dosis-Wirkungsbeziehung h_0, b und c' fest liegen, ergibt sich der Schnittpunkt zu

$$D_s' = \frac{\beta - b}{c'} = \frac{\sigma^2 + \langle D \rangle^2}{\langle D \rangle} = 1{,}25\,\mathrm{Gy}\,. \qquad (8)$$

Er wird unabhängig von den Koeffizienten allein bestimmt durch den Mittelwerte und die Standardabweichung der Dosisverteilung.

Die Koeffizienten b und c' für sechs verschiedene Abweichungen $r=b/\beta$ von der linearen Wirkungsbeziehung sind in Tabelle 1 zusammen gefasst. Für $r>1$ ergeben sich negative Werte für den quadratischen Koeffizienten c', um die Überschätzung der voraus gesagten Fälle auszugleichen, die durch den erhöhten Koeffizienten b verursacht wird. Allerdings wird die Zahl von 350 tatsächlich registrierten Neuerkrankungen immer übertroffen. Die Abweichung wird mit steigenden Werten für $r=b/\beta$ größer. Der Grund liegt darin, dass für negative quadratische Koeffizienten c' bei großen individuellen Dosen mit dem Risikomodell (1) in der Poisson-Approximation unzulässige Erkrankungswahrscheinlichkeiten größer als eins auftreten. Für $r<1$ ergeben sich positive Werte für c'. Für große Personendosen überwiegt jetzt der quadratische Term im Risikomodell (1), er kann nicht mehr als klein angesehen werden. Damit ist die Vorraussetzung für die Gültigkeit der Poisson-Approximation nicht mehr erfüllt. Als Folge der fehlenden Vorraussetzung liegt die Zahl n' der vorher gesagten Neuerkrankungen unter der Anzahl von 350 tatsächlich registrierten Neuerkrankungen. Im vorliegenden Beispiel wird der Betrag des quadratischen Koeffizienten c' wird in der Poisson-Approximation zu klein berechnet, weil nicht-lineare Effekte der Wirkungsbeziehung nicht berücksichtig werden können. Hierdurch entsteht das Potenzial für einen zusätzlichen Bias, der unabhängig von der Zusammenfassung von Ausgangsdaten entsteht.

2. Exakte Berechnung

Die Anzahl der zu erwartenden Neuerkrankungen n wird mit der Gleichung (3) numerisch exakt berechnet. Die darin enthaltene mittlere bevölkerungsbezogene Wahrscheinlichkeit $\langle S \rangle$, im Zeitraum 1990-99 nicht an Schilddrüsenkrebs zu erkranken, kommt aus der Summation der individuellen Wahrscheinlichkeiten nach Gleichung (4). Um die Summation ausführen zu können, werden zuerst die individuellen Schilddrüsendosen für 1002706 Personen simuliert.

Damit die zu erwartenden Neuerkrankungen n den beobachteten Wert 350 annehmen, wird nun bei festem Hintergrundrisiko h_0 und bei festem linearen Koeffizienten b der quadratische Koeffizient c angepasst. Im vorliegenden Beispiel geschah dies durch die numerische Lösung der Gleichung $350-n(c)=0$ mit der Funktion *rtbis ( )* aus den Numerical Recipes (Teukolsky et al. 1992). Diese Gleichung wurde gelöst für 100 simulierte Datensätze mit 1002706 Personendosen. Dann wurde der Mittelwert aus den 100 numerisch berechneten Werten des quadratischen Koeffizienten c gebildet.

Falls die Abweichung des linearen Koeffizienten $r=b/\beta>1$, kann für große Personendosen die Risikofunktion Werte unterhalb des Hintergrundrisikos h_0 annehmen oder sogar negativ werden. Dieses unrealistische Verhalten wurde korrigiert, indem für Risikowerte kleiner als h_0 der Wert für h_0 benutzt wurde (Abbildung 2). In der Simulation wurde diese Korrektur für r=1,1, 1,2 und 1,5 bei ca. 100, 600 und 5000 Personen angewandt.

In Tabelle 1 stehen die Mittelwerte aus 100 Simulationsläufen für den quadratischen Koeffizienten c bei sechs verschiedenen Abweichungen $r=b/\beta$. Der Wert für c ist betrags-

mäßig immer größer als der Wert für c' aus der Poisson-Approximation. Deshalb fällt der Schnittpunkt D_s von linearer und linear-quadratischer Dosiswirkungsbeziehung auch immer unter den Schnittpunkt $D_s^{'} = 1{,}25\,\text{Gy}$ der Poisson-Approximation. Er ist jetzt auch von der Wahl der Koeffizienten abhängig, was bei der Poisson-Approximation nicht der Fall war. Die Abbildungen 1 und 2 zeigen die Risikofunktionen getrennt für $r=b/\beta<1$ und $r=b/\beta>1$ mit den quadratischen Koeffizienten aus der numerisch exakten Berechnung.

Tabelle 1: Linearer Koeffizient b und quadratischer Koeffizient *c´* aus der Poisson-Approximation bzw. *c* aus der exakten Berechnung für die linear-quadratische Dosis-Wirkungsbeziehung, sowie die Anzahl der vorhergesagten Neuerkrankungen *n´* aus Gleichung (4) bzw. *n* aus Gleichung (2) im Zeitraum 1990-99, lineare und linear-quadratische Dosis-Wirkungsbeziehung schneiden sich bei der Dosis D_s

r	b	*c´*	*n´*	*c*	*n*	*c´/c*	D_s
$=b/\beta$	[10^{-6} PY^{-1}]	[10^{-5} PY^{-1} Gy^{-2}]		[10^{-5} PY^{-1} Gy^{-2}]			Gy
0,5	1,26	10,06	332	12,26	350	0,82	1,02
0,8	2,01	4,05	344	4,72	350	0,86	1,06
0,9	2,26	2,01	348	2,30	350	0,87	1,09
1,1	2,76	-1,64	354	-2,60	350	0,63	0,97
1,2	3,01	-3,38	358	-6,74	350	0,50	0,75
1,5	3,77	-8,19	369	-28,73	350	0,29	0,44

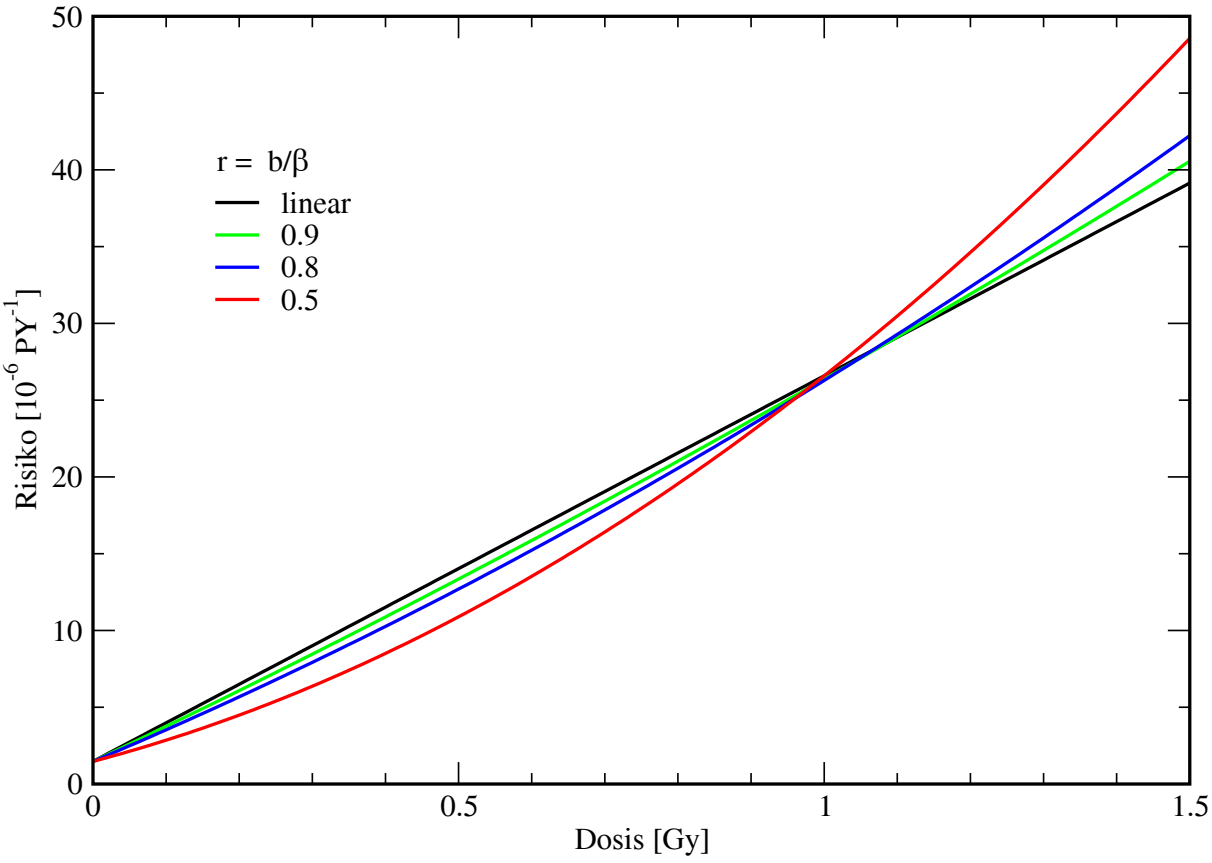

Abbildung 1: Lineare und linear-quadratische Risikofunktionen für verschiedene Abweichungen $r=b/\beta<1$ und exakt berechnete quadratische Koeffizienten *c*

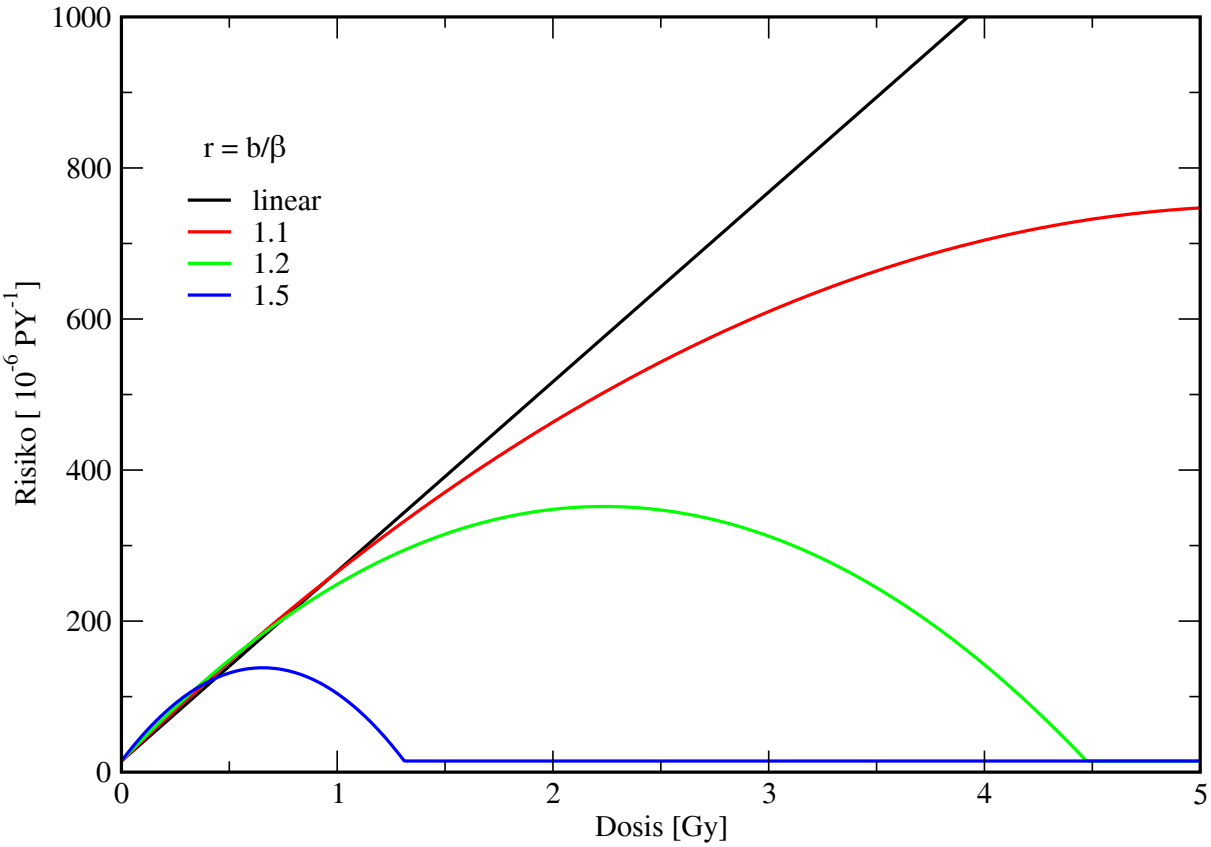

Abbildung 2: Lineare und linear-quadratische Risikofunktionen für verschiedene Abweichungen *r=b/β* >1 und exakt berechnete quadratische Koeffizienten *c*, bei den linear-quadratischen Funktionen wurde für Werte kleiner als das Hintergrundrisiko h_0 der Wert für h_0 angenommen

3. Simulationsergebnisse

Die Simulationen wurden mit den in der Einleitung vorgestellten Ausgangsdaten durchgeführt. Jeder Person wurde individuell eine Dosis und ein Gesundheitszustand (an Schilddrüsenkrebs erkrankt oder nicht) zugeordnet. Die Krebsfälle wurden mit der Risikofunktion (1) erzeugt. Wie oben gezeigt, kann sie jedoch nicht direkt als Erkrankungswahrscheinlichkeit interpretiert werden, sondern es muss die exakte Erkankungswahrscheinlichkeit $1-\exp\left(-\Delta Th_{ij}\right)$ verwendet werden. Einer Person j aus einer Ortschaft i wird ein Fall zugeordnet, falls eine für sie gezogene, zwischen 0 und 1 gleichverteilte Zufallszahl größer ist als diese exakte Erkankungswahrscheinlichkeit. sich ein Im Mittel über 100 Simulationsläufe traten 350 Fälle in 1002706 Personen im Zeitraum von 10 Jahren auf.

Zur Poissonregression wurden die Personen in 670 Ortschaften gruppiert und die arithmetischen Mittelwerte der Ortsdosis gebildet. Für eine Ortschaft i mit N_i Personen aus der Risikogruppe ist dieser Mittelwert

$$\overline{D}_i = \frac{1}{N_i}\sum_j D_{ij} \,. \qquad (9)$$

Zwei verschiedene Risikomodelle wurden in der Regression verwandt, die sich in der Behandlung des quadratischen Terms unterscheiden. Ist nur der Ortsdosismittelwert $\overline{D}_i$

bekannt aus einer Dosisabschätzung, muss für die linear-quadratische Risikofunktion die Beziehung

$$h_i = h_{0,eco} + b_{eco}\overline{D}_i + c_{eco}\left(\overline{D}_i\right)^2 \qquad (10)$$

angenommen werden. Wenn sowohl der Mittelwert $\overline{D}_i$ als auch die arithmetische Standardabweichung σ_i für eine Ortschaft bekannt sind, kann der Mittelwert der quadratischen Ortsdosis $\overline{D_i^2} = \sigma_i^2 + \left(\overline{D}_i\right)^2$ berechnet werden. Er wird in die ortsbezogene Risikofunktion

$$\overline{h}_i = h_{0,eco} + b_{eco}\overline{D}_i + c_{eco}\overline{D_i^2} \qquad (11)$$

eingesetzt.

Für sechs Werte der Abweichung $r=b/\beta$ wurden Simulationsläufe durchgeführt. In den Tabellen 2 und 3 sind die Punktschätzer für die Koeffizienten der linear-quadratischen Risikofunktionen (10) und (11) dargestellt. Um die Genauigkeit der Punktschätzer zu erhöhen, sind jeweils die Mittelwerte aus 100 Simulationsläufen angegeben. Die Konfidenzintervalle für 95 % wurden unter der Annahme berechnet, dass die Poisson-Deviance nahe dem Minimum eine Parabelform besitzt. Daher liegen sie symmetrisch um die Punktschätzer. In jedem Simulationslauf wurde ein Konfidenzintervall berechnet, in den Tabellen stehen die dazugehörigen Mittelwerte aus 100 Simulationsläufen.

Die Punktschätzer für das Hintergrundrisiko $h_{0,eco}$ und den linearen Koeffizienten b_{eco} sind fast identisch für beide Risikofunktionen (10) und (11). Für $r<1$ wird $h_{0,eco}$ unterschätzt und b überschätzt, für r>1 verläuft der Trend umgekehrt. Wenn die Abweichung r zwischen 0,8 und 1,2 liegt, beträgt der Bias ca. 10 Prozent. Für $r=0{,}5$ und $r=1{,}5$ steigt der Bias merkbar an bis zu einem Faktor 1,3 für $h_{0,eco}$ und einem Faktor 1,6 für b.

Die Punktschätzer für den quadratischen Koeffizienten c_{eco} unterscheiden sich deutlich bei den Risikofunktionen (10) und (11). Wenn man den quadratischen Beitrag mit $\left(\overline{D}_i\right)^2$ gewichtet (10), übersteigt der Betrag des geschätzten quadratischen Koeffizienten c_{eco} jeweils den Koeffizienten, der mit der Risikofunktion (11) ermittelt wird. Der Grund liegt darin, dass immer die Ungleichung $\left(\overline{D}_i\right)^2 \le \overline{D_i^2}$ gilt. Durch das Größenverhältnis der quadratischen Koeffizienten wird gewährleistet, dass der quadratische Beitrag der Risiko-funktionen (10) und (11) im Mittel gleich ist.

Die geschätzten quadratischen Koeffizienten besitzen große Unsicherheiten und sind für $r=0{,}9$ und $r=1{,}1$ nicht signifikant verschieden von Null. Der Bias ist meist sehr groß, der wahre quadratische Koeffizient c wurde bis zu einem Faktor 3 unterschätzt und bis zu einem Faktor zwei überschätzt.

In der Bewertung muss man sagen, dass die Koeffizienten einer linear-quadratischen Dosis-wirkungsbeziehung in der ökologischen Regression nicht genau bestimmt werden können. Besonders schlecht gelingt die Schätzung des quadratischen Koeffizienten, auch wenn man den Ortsmittelwert der quadratischen Dosis zur Schätzung benutzt. Die vorliegenden

Rechnungen wurden ohne die Anwesenheit von Confoundern, wie z.B. Screening durchgeführt. Es steht zu erwarten, dass durch den Screening-Effekt die Punktschätzer der Koeffizienten noch weiter verzerrt werden. Sollte eine nicht-lineare Dosiswirkungsbeziehung für das Schilddrüsenkrebsrisiko nach Tschernobyl vorliegen, können deren Parameter mit einer ökologischen Regression nicht verlässlich bestimmt werden.

Tabelle 2: Koeffizienten aus der ökologischen Poissonregression mit der linear-quadratischen Risikofunktion (9) mit Konfidenzintervallen für 95 % aus der parabolischen Approximation der Poisson-Deviance

r	$h_{0,eco}$	bias	b_{eco}	bias	c_{eco}	bias
$=b/\beta$	$[10^{-6}\ PY^{-1}]$	$h_{0,eco}/h_0$	$[10^{-4}\ PY^{-1}\ Gy^{-1}]$	b_{eco}/b	$[10^{-5}\ PY^{-1}\ Gy^{-2}]$	c_{eco}/c
0,5	12,5±5,5	0,85	1,78±0,92	1,42	22,2±8,3	1,97
0,8	13,8±5,6	0,94	2,21±0,88	1,10	9,17±6,46	1,94
0,9	14,1±5,5	0,96	2,37±0,86	1,05	4,55±5,52	1,77
1,1	15,8±5,1	1,07	2,54±0,67	0,92	-3,47±4,01	1,34
1,2	16,6±5,3	1,13	2,59±0,70	0,86	-7,46±4,05	1,11
1,5	19,5±5,7	1,32	2,50±0,80	0,66	-19,4±9,3	0,67

Tabelle 3: Koeffizienten aus der ökologischen Poissonregression mit der linear-quadratischen Risikofunktion (10) mit Konfidenzintervallen für 95 % aus der parabolischen Approximation der Poisson-Deviance

r	$h_{0,eco}$	bias	b_{eco}	bias	c_{eco}	bias
$=b/\beta$	$[10^{-6}\ PY^{-1}]$	$h_{0,eco}/h_0$	$[10^{-4}\ PY^{-1}\ Gy^{-1}]$	b_{eco}/b	$[10^{-5}\ PY^{-1}\ Gy^{-2}]$	c_{eco}/c
0,5	11,6±5,4	0,79	1,95±0,91	1,55	7,74±3,04	0,63
0,8	13,6±5,5	0,92	2,25±0,87	1,12	3,42±2,42	0,72
0,9	14,2±5,4	0,96	2,38±0,83	1,05	1,64±1,96	0,71
1,1	15,7±5,0	1,07	2,57±0,64	0,93	-1,62±8,51	0,62
1,2	16,7±5,0	1,13	2,60±0,62	0,86	-3,33±1,45	0,49
1,5	19,8±5,3	1,34	2,52±0,69	0,67	-8,70±3,41	0,30

Literatur

Likhtarov I, Kovgan L, Vavilov S, Chepurny M, Bouville A, Luckyanov N, Jacob P, Voillequé P und Voigt G. (2004), Post-Chernobyl thyroid doses in Ukraine. Report I: Estimation of thyroid doses, submitted to Radiation Research

Press WH, Flannery BP, Teukolski SA und Vetterling WT (1992), Numerical Recipes in C (2^{nd} edn), Cambridge University Press: Cambridge, MA