

Ein SAS-Macro zur Berechnung des epidemiologischen Risikomaßes der Risk and Rate Advancement Period

Annette Pfahlberg, Tina van der Horst, Olaf Gefeller

Institut für Medizininformatik, Biometrie und Epidemiologie

Universität Erlangen-Nürnberg

Waldstr.6

91054 Erlangen

annette.pfahlberg@rzmail.uni-erlangen.de

Zusammenfassung

Die epidemiologische Analyse des Zusammenhangs zwischen der Existenz eines Expositions-faktors und dem Auftreten einer Erkrankung beschränkt sich zunehmend nicht mehr allein auf die Beantwortung der Frage, ob die Exposition als ätiologischer Faktor für die Erkrankung einzustufen ist. Die Quantifizierung der zeitlichen Dimension des Krankheitsgeschehens in Abhängigkeit von Einflüssen verschiedener Expositionen ist ein ebenso wichtiger Aspekt. Vielfach ist es weit bedeutsamer zu fragen, *wann* eine Krankheit expositionsbedingt auftritt, als *ob* sie überhaupt eintritt. Moderne Risikokonzepte in der Epidemiologie versuchen deshalb, die Zeitinformation über das Auftreten der Erkrankung geeignet in die Definition von Risikomaßen miteinzubeziehen. Der Ansatz der Risk and Rate Advancement Period (RAP) gehört zu diesen neueren Vorschlägen. Er quantifiziert die Stärke des Zusammenhangs zwischen Exposition und Erkrankung in Form von (Lebens-)Zeitspannen, um die das exponierte Kollektiv früher ($RAP > 0$) bzw. später ($RAP < 0$) dasselbe Erkrankungsrisiko (oder dieselbe Erkrankungsrate) aufweist wie die Nichtexponierten. Im Rahmen dieser Arbeit erfolgt die Vorstellung dieses epidemiologischen Ansatzes sowie des Problems der Intervallschätzung zur Konstruktion valider Konfidenzintervalle für die RAP. Für die Realisation dieses Ansatzes in der Praxis wurde eine flexible SAS-Macrosammlung entwickelt, welche sowohl die RAP aus verschiedenen verallgemeinerten linearen Modellen als auch die zugehörigen Konfidenzintervalle angibt.

Keywords: Epidemiologie, Intervallschätzung, Risikomaß, SAS–Macroimplementation.

1 Motivation

Die epidemiologische Analyse des Zusammenhangs zwischen der Existenz eines Expositionsfaktors und dem Auftreten einer Erkrankung beschränkt sich zunehmend nicht mehr allein auf die Beantwortung der Frage, ob die Exposition als ätiologischer Faktor für die Erkrankung einzustufen ist. Die Quantifizierung der zeitlichen Dimension des Krankheitsgeschehens in Abhängigkeit von Einflüssen verschiedener Expositionen ist ein ebenso wichtiger Aspekt der epidemiologischen Auseinandersetzung mit Expositions–Krankheits–Beziehungen geworden. Vielfach ist es weit bedeutsamer zu fragen, *wann* eine Krankheit expositionsbedingt auftritt, als *ob* sie überhaupt eintritt. Moderne Risikokonzepte in der Epidemiologie versuchen deshalb, die Zeitinformation über das Auftreten der Erkrankung geeignet in die Definition von Risikomaßen miteinzubeziehen.

Der bekannteste und älteste derartige Vorschlag ist das Konzept der *vorzeitig verlorenen Lebensjahre*, das bereits Ende der 40er Jahre vorgeschlagen wurde, um bei der Analyse von Todesursachenstatistiken nicht allein die Anzahl der Verstorbenen, sondern auch deren Alter zum Zeitpunkt des Todes zu berücksichtigen [1]. Die Maßzahl vorzeitig verlorener Lebensjahre wird bis heute meist auf der Datenbasis offizieller Gesundheitsstatistiken für Todesfälle geschätzt. Als epidemiologisches Risikokonzept für die Beschreibung expositionsspezifischer Erkrankungsrisiken ist es insbesondere bei chronischen Krankheiten nur eingeschränkt von Nutzen, da zu seiner validen Anwendung restriktive Voraussetzungen hinsichtlich des Expositionseffekts in der Population gegeben sein müssen [2].

Im Vordergrund des Interesses steht deshalb hier ein von Brenner et al. [3] in die epidemiologische Literatur eingebrachtes alternatives Konzept zur Beschreibung des zuschreibbaren Risikos eines Expositionsfaktors für das Entstehen einer Erkrankung. Die Autoren entwickeln ein neues Maß zur Quantifizierung der durch den Expositionsfaktor bedingten Zeitspanne, um die ein kritischer Risikolevel für eine Erkrankung bei Exponierten früher als bei Nichtexponierten erreicht wird. Der Einfluss eines Expositionsfaktors auf die Entstehung einer Krankheit wird in diesem Konzept direkt in Lebenszeit ausgedrückt und somit transparent gemacht. Gerade für die Gesundheitsaufklärung in der Allgemeinbevölkerung ist die Art der Vermittlung epidemiologischer Erkenntnisse über das mit dem Vorhandensein einer Exposition verbundene Krankheitsrisiko von entscheidender Bedeutung. Hier eröffnet der anschauliche Ansatz von Brenner et al. [3] neue Möglichkeiten der Risikokommunikation, indem er bei-

spielsweise Aussagen der Art „die Exposition führt zu einer Vorverlagerung des Erkrankungsrisikos um 10 Jahre“ (oder etwas anders ausgedrückt: „Exponierte eines jeden Alters erreichen stets 10 Jahre vor gleichaltrigen Nichtexponierten dasselbe Risiko für den Eintritt der Krankheit“) erlaubt. Dem neuen Parameter geben die Autoren den Namen *Risk Advancement Period* bzw. *Rate Advancement Period*, falls die Information über das Auftreten der Krankheit in Form von Krankheitsraten vorliegt. In Ermangelung einer passenden deutschen Übersetzung wird auch im weiteren Verlauf der Arbeit der Originalausdruck verwendet oder durch die stets mit dem weiblichen Artikel benutzte Abkürzung *RAP* ersetzt. Während sich Brenner et al. [3] primär um eine epidemiologische Motivation dieser neuen Maßzahl bemühen und eine Diskussion der Nützlichkeit der *RAP* in praktischen epidemiologischen Studien führen, steht im Rahmen dieser Arbeit die Behandlung der statistischen Aspekte dieses neuen Risikokonzepts und der praktischen Umsetzung des Ansatzes in ein SAS-Macro im Vordergrund.

2 Definition und Herleitung der RAP

Der epidemiologische Risikoparameter *RAP* lässt sich formal durch folgende Definition beschreiben.

Definition: Risk Advancement Period

Betrachte die Situation einer dichotomen Krankheitsvariablen D , eines Expositionsfaktors E (kategorial oder stetig), einer Altersvariablen A (stetig) und einer Reihe weiterer Kovariablen C_i , $i = 1, \dots, m$, beliebigen Skalenniveaus, die in einem Kovariablenvektor $C := (C_1, \dots, C_m)'$ zusammengefasst seien.

Sei $R_T(E = e, A = a, C = c)$ das erwartete Risiko, ein Krankheitsereignis ($D = 1$) innerhalb von T Jahren nach Erreichen des Alters a zu entwickeln, unter der Bedingung, bis dahin gesund geblieben zu sein, in der Population der Personen mit Expositionsausprägung e und gegebenem Kovariablenvektor $c = (c_1, \dots, c_m)'$.

Für zwei Expositionsniveaus e_1 und e_0 lässt sich — für beliebiges Alter x , Kovariablenvektor c und Zeitspanne T — die Frage zur Klärung der zeitlichen Dimension des durch das Expositionsniveau e_1 im Vergleich zu e_0 bedingten „vorverlagerten“ Erkrankungsrisikos im Alter x wie folgt stellen:

„Für welches Alter y gilt, dass bei Expositionsniveau e_0 das erwartete mit y beginnende T -Jahres-Erkrankungsrisiko gleich hoch ist wie beim Expositionsniveau e_1 im Alter x ?“

Formal bedeutet das, man suche das kleinste y , welches die Gleichung

$$R_T(E = e_0, A = y, C = c) = R_T(E = e_1, A = x, C = c)$$

eindeutig löst.

Falls dieses y existiert, so definiert die Differenz $y - x$ die Risk Advancement Period ($RAP_{A,C}$) im Alter $A = x$ und vorgegebener Kovariablenstruktur $C = c$ für das Expositionsniveau $E = e_1$ im Vergleich zu $E = e_0$. Sollte y nicht existieren, so definiert man $RAP_{A,C} := \infty$.

Die Eindeutigkeit der RAP -Definition ist garantiert durch die Beschränkung auf mit zunehmendem Alter monoton ansteigende Risikofunktionen in den Anwendungen des Konzeptes. Ein Wegfall dieser Restriktion kommt somit auch aus formalen Gründen nicht in Betracht. Falls die $RAP_{A,C}$ unabhängig vom Ausgangsalter A und/oder der Kovariablenstruktur C ist, so lässt man der Übersichtlichkeit halber den jeweiligen Index weg.

Eine graphische Illustration der obigen Definition bietet Abbildung 1, in der hypothetische altersabhängige Kurven des erwarteten T -Jahres-Erkrankungsrisikos für den Expositionsniveau $E = e_1$ und den Expositionsniveau $E = e_0$ abgebildet sind. Für die beiden Altersstufen $A = 30$ und $A = 45$ ist die jeweils zugehörige $RAP = 10$ als gestrichelte Linie eingetragen. Im abgebildeten Beispiel sind die beiden Risikokurven so gewählt, dass für beliebiges Alter die zugehörige RAP jeweils 10 beträgt. Diese Konstanz der RAP ist jedoch nicht zwingend. Wie in obiger Definition aus der Indizierung ersichtlich, kann die RAP sowohl mit dem Alter als auch mit weiteren Kovariablen variieren.

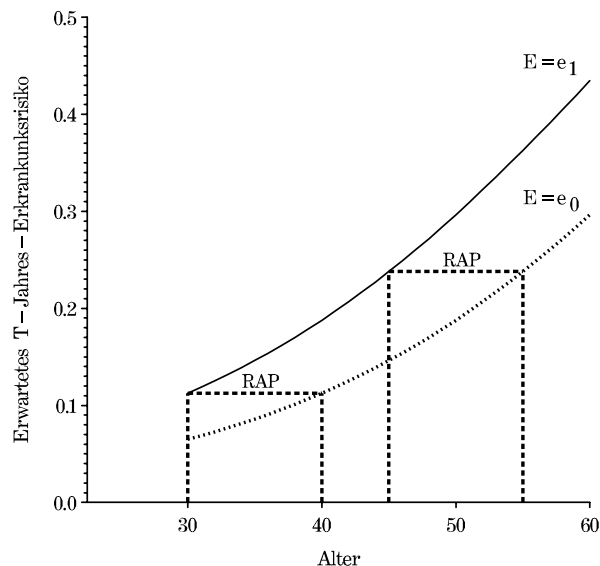


Abbildung 1: Graphische Illustration der RAP

Völlig analog lässt sich die obige Definition der *RAP* wiederholen, falls die epidemiologische Information über den Zusammenhang zwischen Exposition und Krankheit in Form zeitbezogener Erkrankungsraten (anstatt kumulativ auf eine Zeitperiode bezogene Erkrankungsrisiken) vorliegt. Formal ändert sich durch den Übergang von „Risiken“ zu „Raten“ nichts.

Die allgemeine Definition der *RAP* eröffnet noch keinen direkten Zugang zu ihrer konkreten Ermittlung. Dies soll hier im Kontext der Modellierung von Erkrankungsrisiken bzw. -raten in Abhängigkeit von Expositions-, Alter- und Kovariableneinflüssen nachgeholt werden. Die Angabe der interessierenden Expositions-Krankheits-Beziehung erfolgt in epidemiologischen Beobachtungsstudien üblicherweise mit Hilfe eines verallgemeinerten linearen Regressionsmodell der Form

$$f(R(E = e, A = a, C = c)) = \beta_0 + \beta_1 \cdot e + \beta_2 \cdot a + \sum_{i=1}^m \beta_{i+2} \cdot c_i, \quad (1)$$

wobei $R(\cdot)$ entweder das Erkrankungsrisiko oder die Erkrankungsrate repräsentiert, $f(\cdot)$ die streng monotone Link-Funktion des jeweiligen Modells bezeichnet und $\beta := (\beta_0, \beta_1, \dots, \beta_{m+2})'$ der Vektor der Modellparameter ist [4]. Für die Herleitung der *RAP* spielt der spezifische Modelltyp keine Rolle und braucht daher nicht berücksichtigt werden.

Zur einfachen Darstellung der *RAP*-Herleitung aus Regressionsmodellen erfolgt zunächst die Beschränkung auf den Fall eines dichotomen Expositionsfaktors E , dessen beide Zustände 0/1-kodiert sind ($E = 1$: „exponiert“, $E = 0$: „nicht exponiert“). Darüber hinaus wird zunächst angenommen, dass sich unter den Kovariablen C_i , $i = 1, \dots, m$, keine die Expositions- und Altersvariable beinhaltenden Interaktionsterme befinden. In dieser Situation gilt dann:

Modell 1 (keine Interaktionsterme)

$$\begin{aligned} f(R(E = 1, A = x, C = c)) &= f(R(E = 0, A = y, C = c)), \\ \beta_1 + \beta_2 \cdot x &= \beta_2 \cdot y, \\ y - x &= \frac{\beta_1}{\beta_2}, \\ RAP = g(\beta) &= \frac{\beta_1}{\beta_2}. \end{aligned} \quad (2)$$

Innerhalb dieses Modells lässt sich somit die expositionsbedingte *RAP*-Zeitspanne $y - x$, um welche die Gruppe die Exponierten dasselbe Erkrankungsrisiko (bzw. dieselbe Erkrankungsrate) früher als die Gruppe der Nichtexponierten bei ansonsten identischem Kovariablenmuster erreicht, durch den Quotienten der beiden Modellparameter β_1 und β_2 ausdrücken. Für protektive dichotome

Expositionsfaktoren, d.h. Einflussgrößen, deren Anwesenheit zu einer Reduktion des Erkrankungsrisikos (bzw. der Erkrankungsrate) führt, wird die obige Differenz automatisch negativ (da $\beta_1 < 0$), und der Absolutbetrag der Differenz ist als expositionsbedingte Zeitspanne des *verzögerten* Erreichens desselben Erkrankungsrisikos (bzw. derselben Erkrankungsrate) aufzufassen.

Die obige Herleitung zeigt, dass die *RAP* sich in einfacher Weise durch die Modellparameter eines Regressionsmodells darstellen lässt. In der Situation (2) entsteht eine alters- und kovariablenunabhängige *RAP*, die als konstante Größe den interessierenden Expositionseffekt widerspiegelt. In komplexeren Modellen, in denen sich unter den Kovariablen C_i , $i = 1, \dots, m$, auch die Expositions- und Altersvariable umfassende Interaktionsterme befinden, ist die *RAP* keine konstante Größe mehr. Im folgenden wird das Vorgehen in diesen Modellsituationen für dichotome Expositionsvariablen kurz zusammengestellt:

Modell 2 (Interaktionsterm zwischen E und A)

$$\begin{aligned} f(R(E = e, A = a, C = (e \cdot a, c_2, c_3, \dots, c_m))) &= \beta_0 + \beta_1 \cdot e + \beta_2 \cdot a \\ &\quad + \beta_3 \cdot e \cdot a + \sum_{i=2}^m \beta_{i+2} \cdot c_i \\ RAP_A = g(\beta) &= \frac{\beta_1 + \beta_3 \cdot A}{\beta_2} \end{aligned}$$

Modell 3 (Interaktionsterm zwischen E und einer spezifischen Kovariablen F)

$$\begin{aligned} f(R(E = e, A = a, C = (e \cdot f, c_2, c_3, \dots, c_m))) &= \beta_0 + \beta_1 \cdot e + \beta_2 \cdot a \\ &\quad + \beta_3 \cdot e \cdot f + \sum_{i=2}^m \beta_{i+2} \cdot c_i \\ RAP_F = g(\beta) &= \frac{\beta_1 + \beta_3 \cdot F}{\beta_2} \end{aligned}$$

Modell 4 (Interaktionsterm zwischen A und einer spezifischen Kovariablen F)

$$\begin{aligned} f(R(E = e, A = a, C = (a \cdot f, c_2, c_3, \dots, c_m))) &= \beta_0 + \beta_1 \cdot e + \beta_2 \cdot a \\ &\quad + \beta_3 \cdot a \cdot f + \sum_{i=2}^m \beta_{i+2} \cdot c_i \\ RAP_F = g(\beta) &= \frac{\beta_1}{\beta_2 + \beta_3 \cdot F} \end{aligned}$$

Die Modellkomplexität ließe sich durch die Betrachtung weiterer Interaktionsterme noch beliebig steigern. Die *RAP*-Herleitung folgt jedoch stets denselben schon im einfachsten Fall (2) dargelegten Prinzipien. Die aus den jeweiligen

Modellen gewonnenen Darstellungen der zugehörigen *RAP*-Parameter sind stets Quotienten von Linearkombinationen der Modellparameter. Dies bleibt auch in komplexeren Modellen gültig, jedoch steigt die Zahl der Summanden der jeweiligen Zähler- bzw. Nennerlinearkombinationen.

Im Fall der Erweiterung des obigen Ansatzes auf die Situation stetig erhobener Expositionsvariablen reflektieren die in den unterschiedlichen Modellen hergeleiteten *RAP*-Ausdrücke die mit einer Veränderung um eine Einheit der Expositionsvariablen einhergehende *RAP*. Bei ordinalen Expositionsvariablen spielt die Frage der Kodierung eine wesentliche Rolle. In dem bei epidemiologischen Anwendungen häufig anzutreffenden Fall einer *Dummy-Kodierung* eines ordinalskalierten Expositionsmerkmals (d.h. Repräsentanz des ordinalen Merkmals durch mehrere dichotome *Dummy-Variablen* für die einzelnen Expositions-kategorien im Modell) können die oben hergeleiteten *RAP*-Darstellungen ebenfalls direkt verwendet werden. Sie sind in diesem Fall als die mit der jeweiligen Expositions-kategorie assoziierte *RAP* im Vergleich zu dem bei der *Dummy-Kodierung* des ordinalen Expositionsmerkmals als Referenzkategorie gewählten Expositionslevel aufzufassen.

3 Punkt- und Intervallschätzung der *RAP*

Zuvor wurde gezeigt, wie der *RAP*-Parameter als Funktion der Modellparameter spezifischer Regressionsmodelle zur Charakterisierung des Erkrankungsrisikos in Abhängig von Expositions-, Alters-, Confounder- und eventuell Interaktionseffekten herzuleiten ist. Der Maximum-Likelihood-Schätzer für die *RAP* ist dann definiert durch das Einsetzen der Maximum-Likelihood-Schätzer für die Modellparameter β_i in die den Zusammenhang zwischen *RAP* und den Modellparametern beschreibende Funktion.

Die asymptotischen Eigenschaften der *RAP*-Schätzer in den verschiedenen Modellen sind direkt an die Eigenschaften der Parameterschätzer $\hat{\beta}_i$ in den jeweiligen Modellen gekoppelt. Allgemein gilt daher: Sind die Parameterschätzer $\hat{\beta}_i$ konsistent und asymptotisch unverzerrt für β_i , so sind auch die jeweiligen *RAP*-Schätzer in den Modellen konsistent und asymptotisch unverzerrt für das jeweilige *RAP*. Konvergiert der Vektor $\hat{\beta}_i$ gegen eine multivariate Normalverteilung, so ist auch der jeweilige *RAP*-Schätzer asymptotisch normalverteilt.

Neben der Punktschätzung ist die Bestimmung von $(1 - \alpha)$ -Konfidenzintervallen für die *RAP* von zentraler Bedeutung in der praktischen Anwendung des Konzepts. Nur durch die Angabe geeigneter Intervallschätzer für die *RAP* wird die Basis für eine auch die Stichprobenvariabilität berücksichtigende epidemiologische Interpretation der Studienergebnisse geschaffen. Somit muss der Konstruktion von Intervallschätzern, die möglichst präzise das vorgegebene Konfidenzniveau in allen praktisch relevanten Modellsituationen einhalten sollen,

besonderes Augenmerk in der statistischen Bearbeitung des *RAP*-Konzepts geschenkt werden.

Angesichts der Rahmenbedingungen bei der epidemiologischen Anwendung beschränkt sich die Beschäftigung mit der Konstruktion von Intervallschätzern auf die Behandlung *asymptotischer Verfahren*, exakte $(1 - \alpha)$ -Konfidenzintervalle sind in der vorliegenden Situation nicht allgemein herzuleiten. Mit Hilfe der Delta-Methode lässt sich in den verschiedenen Modellsituationen direkt ein asymptotisches $(1 - \alpha)$ -Konfidenzintervall für *RAP* wie folgt angeben:

$$\left[\widehat{RAP} - u_{1-\alpha/2} \cdot \sqrt{\widehat{VarRAP}}, \widehat{RAP} + u_{1-\alpha/2} \cdot \sqrt{\widehat{VarRAP}} \right]. \quad (3)$$

Dieses asymptotische Standardkonfidenzintervall basiert auf der Verteilungskonvergenz des *RAP*-Schätzers gegen eine Normalverteilung und wurde bereits von Brenner et al. [3] für die dort behandelten Situationen der *RAP*-Schätzung explizit angegeben, dort jedoch als einzige vorgestellte Möglichkeit zur Intervallschätzung.

In Analogie zur Intervallschätzung für den Erwartungswertparameter aus einer Stichprobe normalverteilter Daten mit unbekannter Varianz, bei der sich die Notwendigkeit der zusätzlichen Schätzung dieser Varianz in der Verwendung des entsprechenden t-Verteilungsquantils anstelle des Normalverteilungsquantils niederschlägt, werden in den praktischen Anwendungen der Intervallschätzung für unterschiedlichste Parameter ebenfalls häufig t-Verteilungsquantile verwendet, um das $(1 - \alpha)$ -Konfidenzintervall bei kleinen Stichprobenumfängen bewusst etwas breiter zu machen (siehe z.B. [5]). Im vorliegenden Fall ergibt sich konkret

$$\left[\widehat{RAP} - t_{df;1-\alpha/2} \cdot \sqrt{\widehat{VarRAP}}, \widehat{RAP} + t_{df;1-\alpha/2} \cdot \sqrt{\widehat{VarRAP}} \right], \quad (4)$$

wobei df der Anzahl der Freiheitsgrade bei der Varianzschätzung von \widehat{RAP} entspricht. Es sei nochmals betont, dass obiges Konfidenzintervall (4) auf einer rein heuristischen Argumentation beruht. Eine finite Verteilungsaussage für \widehat{RAP} existiert nicht, so dass die Verwendung des t-Verteilungsquantils hier kein statistisch zu begründendes Vorgehen, sondern eher eine ad hoc Idee zur vermeintlichen Verbesserung der Eigenschaften des Konfidenzintervalls bei kleinen Stichprobenumfängen darstellt.

Die Anwendung des Fieller-Theorems [6] bietet für die Intervallschätzung der *RAP* ebenfalls eine Möglichkeit, in allen Modellsituationen $(1 - \alpha)$ -Konfidenzintervalle zu konstruieren. Für die *RAP*-Parameter, bei denen im Zähler bzw. Nenner des Quotienten jeweils eine Linearkombination von Zufallsvariablen auftaucht, kann das Theorem von Fieller unter Einschaltung eines einfachen Substitutionschrittes angewendet werden.

Um Unterschiede und Ähnlichkeiten zwischen den aus der Anwendung der Delta-Methode und des Fieller-Theorems resultierenden Intervallschätzern zu verdeutlichen, wurden unterschiedlichste algebraisch äquivalente Umformungen untersucht. In der folgenden Darstellung erkennt man die strukturellen Unterschiede am besten:

Delta-Methode:

$$\widehat{RAP} \mp q \cdot \widehat{RAP} \cdot \sqrt{\widehat{V}},$$

Fieller-Theorem:

$$\left(\widehat{RAP} - q^2 \cdot \widehat{U} \mp q \cdot \widehat{RAP} \cdot \sqrt{\widehat{V} - \widehat{W}} \right) \cdot \frac{1}{1 - \widehat{G}},$$

wobei q das jeweilige $(1 - \alpha/2)$ -Quantil der Normal- bzw. t -Verteilung bezeichnet und \widehat{RAP} , \widehat{U} , \widehat{V} , \widehat{W} und \widehat{G} in Tabelle 1 auf Seite 231 für Standardsituationen explizit spezifiziert sind. Formal betrachtet würden für $\widehat{G} = \widehat{U} = \widehat{W} = 0$ die beiden Ansätze zu identischen Intervallschätzern führen. Dies ist jedoch unmöglich, da stets sowohl $\widehat{G} > 0$ als auch $\widehat{U} = 0 \Rightarrow \widehat{W} \neq 0$ gilt. Somit liefern die beiden Verfahren immer unterschiedliche Intervallschätzer für RAP , wobei die Frage zu klären bleibt, in welcher Größenordnung die Unterschiede liegen und wodurch sie maßgeblich beeinflusst werden.

Die auf der Anwendung von Resampling-Techniken fußenden Ansätze zur Intervallschätzung können in allen Modellsituationen direkt von ihrer allgemeinen Verfahrensbeschreibung auf den hier interessierenden Fall der Schätzungen von $(1 - \alpha)$ -Konfidenzintervallen für die RAP übertragen werden. Beim Jackknife-Verfahren [7] müssen die Pseudowerte der Stichprobe auf der Basis der RAP -Schätzer der Gesamtstichprobe und der um die i -te Beobachtung verringerten Stichprobe ermittelt werden. Die Bildung des arithmetischen Mittels dieser Pseudowerte führt zum Jackknife-Schätzer für RAP und die Berechnung der entsprechend standardisierten mittleren quadratischen Abweichung der Pseudowerte vom Jackknife-Schätzer zur Jackknife-Varianzschätzung für $Var\widehat{RAP}$. Das mittels t -Verteilungsquantilen gebildete $(1 - \alpha)$ -Konfidenzintervall nach Tukey [8] und das auf der Normalverteilungsapproximation basierende $(1 - \alpha)$ -Konfidenzintervall nach Miller [9] ist damit bei Verwendung der Jackknife-Schätzer für RAP und für $Var\widehat{RAP}$ entsprechend definiert.

Auch die Bootstrap-Methode lässt sich zur Schätzung der RAP und des zugehörigen $(1 - \alpha)$ -Konfidenzintervalls auf der Grundlage der allgemeinen Definition direkt verwenden. Zur Intervallschätzung mittels Bootstrap wird in allen einschlägigen Lehrbüchern [10] eine hohe Anzahl von Bootstrap-Replikationen empfohlen, da sonst die Ordnungsstatistik der aus den Bootstrap-Replikationen ermittelten Schätzer in den zur Konfidenzintervallschätzung heranzuziehenden Randbereichen zu große Sprünge aufweist und der Effekt der stichprobenbedingten Variabilität der Bootstrap-Replikationen auf die praktische

Realisation der Intervallschätzung einen zu großen Einfluss hat. Nur durch eine entsprechend große Zahl von Bootstrap-Replikationen – in der Regel wird die Ziehung von mindestens 1000 Bootstrap-Resamplingstichproben empfohlen – können zuverlässige Intervallschätzungen erhalten werden, was allerdings zu einem enormen Zeitaufwand für die Realisation der Bootstrap-Methoden führt.

4. Implementation in SAS

Die Realisation der Punkt- und Intervallschätzung der RAP aus verschiedenen verallgemeinerten linearen Modellen ist eingebettet in die SAS-Software. Eine flexible SAS-Macrosammlung bietet dem Anwender die Möglichkeit, die RAP-Schätzung in die statistische Analyse der epidemiologischen Fragestellung direkt zu integrieren und die komplette, oben vorgestellte Methodik ohne zusätzlichen eigenen Programmieraufwand zur Verfügung zu haben. Da je nach SAS/STAT-Prozedur, die zur statistischen Analyse benutzt wird, kleinere technische Adaptionen nötig sind, existieren die folgenden drei SAS-Macros:

```
SAS/STAT/PROC LOGISTIC - choose Macro RAP_LOGISTIC
SAS/STAT/PROC PHREG - choose Macro RAP_PHREG
SAS/STAT/PROC GENMOD - choose Macro RAP_GENMOD
```

Bei der Verwendung einer der ersten beiden RAP-Macrovarianten müssen zunächst folgende SAS-Hilfsmacros hereingeladen werden:

```
%include mod1.sas;
%include mod2.sas;
%include mod3.sas;
%include mod4.sas;
```

Falls RAP_GENMOD verwendet wird, sind hingegen die folgenden SAS-Hilfsmacros notwendig:

```
%include mod1_genmod.sas;
%include mod2_genmod.sas;
%include mod3_genmod.sas;
%include mod4_genmod.sas;
```

Die Syntax des Macroaufrufs für alle drei RAP-Macrovarianten unter Angabe aller größtenteils optionaler Möglichkeiten zur Anpassung des Macros an die Benutzerwünsche sieht wie folgt aus:

```
%Rap_...(< DATA=SAS-Dateiname,>
Y = Zielvariable,
< CENSOR = nur für RAP_PHREG: Zensierungsvariable(Ausprägung),>
CLASS = Liste der Variablen im CLASS-Statement der SAS/STAT-Prozedur,
```

RAPVAR = Liste der Expositionsvariablen für RAP-Schätzung,
 < RESTVAR = Liste von Kovariablen (default: keine Kovariablen),>
 AGE = Altersvariable
 < , INT_RAPVAR_AGE = Expositionsvariable, die mit dem Alter interagiert
 (default: keine Interaktion) >
 < , INT_RAPVAR_RESTVAR = Expositionsvariable, die mit einer Kovaria-
 blen interagiert (default: keine Interaktion)>
 < , INT_RESTVAR_RAPVAR = Kovariable, die mit der Expositionsvariable
 interagiert (default: keine Interaktion)>
 < , INT_RESTVAR_AGE = Kovariable, die mit dem Alter interagiert (default:
 keine Interaktion)>
 < , REF = spezifiziert das Referenzlevel für die CLASS-Variablen>
 < , MODELOPTIONS = weitere Modelloptionen der SAS/STAT-Prozedur>
 < , CILEVEL = Konfidenzintervallniveau (default: 0.95)>
 < , OUT = Output in SAS-Datei oder als Ausdruck (default: PRINT)>
**Folgende Optionen werden nur für die graphische Illustration der
 RAP benötigt:**
 < , PLOT = Anforderung einer Graphik (PLOT=1, default: 0)>
 < , MIN = untere Grenze der RAP-Berechnung (default: 0) >
 < , MAX = obere Grenze der RAP-Berechnung (default: 100) >
 < , INCR = Abstand für RAP-Berechnung (default: 5) >
 < , DEV = Name des Device (default: win)>
 < , DEVPLOT = Name des Targetdevice (default: psepsf)>
 < , HS = Horizontale Größe des Plots (default: 20cm) >
 < , VS = Vertikale Größe des Plots (default: 15cm)>
 < , TITLE_F = Schriftart für die Überschrift des Plots (default: swissl)>
 < , TEXT_F = Schriftart für den Text des Plots (default: swissl)>
 < , TITLE_H = Schriftgröße für den Titel des Plots (default: 1.5)>
 < , TEXT_H = Schriftgröße für den Text des Plots (default: 1)>
 < , C1 = Farbe der RAP-Kurve (default: black)>
 < , C2 = Farbe der Konfidenzintervallgrenzen (default: black)>);

Das Macro ist — für nichtkommerzielle Anwendungen kostenlos — mit ei-
 ner ausführlichen Beschreibung unter folgender Web-Adresse herunterzuladen:
<http://www.imbe.med.uni-erlangen.de/issan/issan.htm> .
 Diese Web-Seite wird kontinuierlich gepflegt, so dass dort auch Updates des
 Macros zu finden sein werden.

Literatur

- [1] Dempsey M (1947): Decline in tuberculosis; the death rate fails to tell the entire story. Am Rev Tuberculosis 86, 157.

- [2] Robins JM, Greenland S (1991): Estimability and estimation of expected years of life lost due to hazardous exposure. *Stat Med* 10, 79–93.
- [3] Brenner H, Gefeller O, Greenland S (1993): Risk and rate advancement periods as measures of exposure impact on the occurrence of chronic disease. *Epidemiology* 4, 229–236.
- [4] McCullagh P, Nelder JA: Generalized linear models. 2nd ed. Chapman & Hall, New York 1989.
- [5] Herson J (1975): Fieller’s theorem vs. delta method for significance intervals for ratios. *J Statist Comput Simul* 3, 265–274.
- [6] Fieller EC (1944): A fundamental formula in the statistics of biological assays and some applications. *Q J Pharm Pharmacol* 17, 117–123.
- [7] Quenouille M (1949): Approximation tests of correlation in time series. *J Royal Statist Soc B* 11, 18–44.
- [8] Tukey JW (1958): Bias and confidence in not quite large samples. (abstract) *Ann Math Statist* 29, 614.
- [9] Miller RG (1964): A trustworthy jackknife. *Ann Math Statist* 39, 1594–1605.
- [10] Efron B, Tibshirani RJ: An introduction to the bootstrap. Chapman & Hall, New York 1993.

Tabelle 1: Zusammenfassung der Konfidenzintervalle nach Fieller und Delta

\widehat{RAP}	\widehat{U}	\widehat{V}	\widehat{W}	\widehat{G}
$\frac{\hat{\beta}_1}{\hat{\beta}_2}$	$\frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}{\hat{\beta}_2^2}$	$\frac{\widehat{Var}\hat{\beta}_1}{\hat{\beta}_1^2} + \frac{\widehat{Var}\hat{\beta}_2}{\hat{\beta}_2^2} - 2 \cdot \frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}{\hat{\beta}_1 \cdot \hat{\beta}_2}$	$q^2 \left(\frac{\widehat{Var}\hat{\beta}_1}{\hat{\beta}_1^2} \cdot \frac{\widehat{Var}\hat{\beta}_2}{\hat{\beta}_2^2} - \left(\frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}{\hat{\beta}_1 \cdot \hat{\beta}_2} \right)^2 \right)$	$q^2 \cdot \frac{\widehat{Var}\hat{\beta}_2}{\hat{\beta}_2^2}$
$\frac{\hat{\beta}_1 + \hat{\beta}_3 \cdot A}{\hat{\beta}_2}$	$\frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) + A \cdot \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3)}{\hat{\beta}_2^2}$	$\frac{\widehat{Var}\hat{\beta}_1 + A^2 \cdot \widehat{Var}\hat{\beta}_3 + 2 \cdot A \cdot \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3)}{(\hat{\beta}_1 + \hat{\beta}_3 \cdot A)^2} + \frac{\widehat{Var}\hat{\beta}_2}{\hat{\beta}_2^2} - 2 \cdot \frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) + A \cdot \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3)}{(\hat{\beta}_1 + \hat{\beta}_3 \cdot A) \cdot \hat{\beta}_2}$	$q^2 \left(\frac{\widehat{Var}\hat{\beta}_1 + A^2 \cdot \widehat{Var}\hat{\beta}_3 + 2 \cdot A \cdot \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3)}{(\hat{\beta}_1 + \hat{\beta}_3 \cdot A)^2} \cdot \frac{\widehat{Var}\hat{\beta}_2}{\hat{\beta}_2^2} - \left(\frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) + A \cdot \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3)}{(\hat{\beta}_1 + \hat{\beta}_3 \cdot A) \cdot \hat{\beta}_2} \right)^2 \right)$	$q^2 \cdot \frac{\widehat{Var}\hat{\beta}_2}{\hat{\beta}_2^2}$
$\frac{\hat{\beta}_1}{\hat{\beta}_2 + \hat{\beta}_3 \cdot C}$	$\frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) + C \cdot \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3)}{(\hat{\beta}_2 + \hat{\beta}_3 \cdot C)^2}$	$\frac{\widehat{Var}\hat{\beta}_1}{\hat{\beta}_1^2} + \frac{\widehat{Var}\hat{\beta}_2 + C^2 \cdot \widehat{Var}\hat{\beta}_3 + 2 \cdot C \cdot \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3)}{(\hat{\beta}_2 + \hat{\beta}_3 \cdot C)^2} - 2 \cdot \frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) + C \cdot \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3)}{\hat{\beta}_1 \cdot (\hat{\beta}_2 + \hat{\beta}_3 \cdot C)}$	$q^2 \left(\frac{\widehat{Var}\hat{\beta}_1}{\hat{\beta}_1^2} \cdot \frac{\widehat{Var}\hat{\beta}_2 + C^2 \cdot \widehat{Var}\hat{\beta}_3 + 2 \cdot C \cdot \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3)}{(\hat{\beta}_2 + \hat{\beta}_3 \cdot C)^2} - \left(\frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) + C \cdot \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3)}{\hat{\beta}_1 \cdot (\hat{\beta}_2 + \hat{\beta}_3 \cdot C)} \right)^2 \right)$	$q^2 \cdot \frac{\widehat{Var}\hat{\beta}_2}{(\hat{\beta}_2 + \hat{\beta}_3 \cdot C)^2} + \frac{C^2 \cdot \widehat{Var}\hat{\beta}_3}{(\hat{\beta}_2 + \hat{\beta}_3 \cdot C)^2} + \frac{2 \cdot C \cdot \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3)}{(\hat{\beta}_2 + \hat{\beta}_3 \cdot C)^2}$
$\frac{\hat{\beta}_1 + \hat{\beta}_3 \cdot C}{\hat{\beta}_2}$	$\frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) + C \cdot \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3)}{\hat{\beta}_2^2}$	$\frac{\widehat{Var}\hat{\beta}_1 + C^2 \cdot \widehat{Var}\hat{\beta}_3 + 2 \cdot C \cdot \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3)}{(\hat{\beta}_1 + \hat{\beta}_3 \cdot C)^2} + \frac{\widehat{Var}\hat{\beta}_2}{\hat{\beta}_2^2} - 2 \cdot \frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) + C \cdot \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3)}{(\hat{\beta}_1 + \hat{\beta}_3 \cdot C) \cdot \hat{\beta}_2}$	$q^2 \left(\frac{\widehat{Var}\hat{\beta}_1 + C^2 \cdot \widehat{Var}\hat{\beta}_3 + 2 \cdot C \cdot \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3)}{(\hat{\beta}_1 + \hat{\beta}_3 \cdot C)^2} \cdot \frac{\widehat{Var}\hat{\beta}_2}{\hat{\beta}_2^2} - \left(\frac{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) + C \cdot \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3)}{(\hat{\beta}_1 + \hat{\beta}_3 \cdot C) \cdot \hat{\beta}_2} \right)^2 \right)$	$q^2 \cdot \frac{\widehat{Var}\hat{\beta}_2}{\hat{\beta}_2^2}$

