

Polytome logistische Regression: Auswertung einer Fall-Kontroll-Studie

Silvia Sander
Institut für Epidemiologie und Medizinische Biometrie
Universität Ulm
Schwabstr. 13
89075 Ulm
silvia.sander@uni-ulm.de

Peter Kaskel

Asterweg 39
85591 Vaterstetten
kaskel@derma.de

Martina Kron
Institut für Epidemiologie und Medizinische
Biometrie, Universität Ulm
Schwabstr. 13
89075 Ulm
martina.kron@uni-ulm.de

Zusammenfassung

Es gibt nicht viele Studien, die die Risikofaktoren für das Auftreten eines Basalioms oder Melanoms untersuchen. Die vorliegende Fall-Kontroll-Studie vergleicht die Risikofaktoren, die klinische und verhaltensorientierte Aspekte abbilden für Basaliome, Melanome und Kontrollen. Am Beispiel der polytomen Zielgröße wird das Vorgehen beschrieben. Beim Vergleich der beiden Tumorarten mit den gesunden Kontrollpersonen lassen sich diverse Risikofaktoren identifizieren, die unter anderem aus dem Kontext des Sonnenexpositionsverhaltens stammen. Dies sind Faktoren der Freizeitaktivitäten in der Kindheit und 20 Jahre vor Diagnosestellung sowie des Berufs. Des Weiteren spielen aber auch phänotypische und klinische Charakteristika eine wichtige Rolle, die mit der Entstehung von Basaliom und Melanom assoziiert sind. Die Programme zur Durchführung der polytomen logistischen Regression werden vorgestellt.

Schlüsselwörter: polytome logistische Regression, Fall-Kontroll-Studie, Basaliom, Melanom

1 Hintergrund

In Europa besteht eine deutliche Zunahme der Neuerkrankungen an Basaliomen und Melanomen [1]. Nur wenige Studien untersuchen die vielfältigen Risikofaktoren, die klinische und verhaltensorientierte Aspekte abbilden, sowohl für das Basaliom als auch für das Melanom. Insbesondere für das Basaliom ist unklar, welche Faktoren als Risikofaktoren zu werten sind. Die vorliegende Fall-Kontroll-Studie untersucht die Sonnenexposition in Freizeit und Beruf, die Sonnenempfindlichkeit und deren Bedeutung im Hinblick auf die Entstehung von Basaliom und Melanom unter Berücksichtigung klini-

scher und soziodemographischer Charakteristika [2]. Die polytome Zielgröße der Studie mit den 3 Ausprägungen Basaliom, Melanom und Kontrolle wird mit einer logistischen Regression ausgewertet.

Anhand dieses Beispiels wird das zur Umsetzung der statistischen Auswertung notwendige Vorgehen in SAS illustriert.

2 Material und Methoden

Die Studie umfasst 503 Patienten (Basaliom (B) $n=212$, Melanom (M) $n=291$) und 329 Kontrollpersonen (K) aus Ulm und Dresden. Mit Hilfe standardisierter Fragebögen wurden soziodemographische sowie konstitutionelle Merkmale und Art und Ausmaß der Sonnenexposition in der Kindheit und 20 Jahre vor Diagnosestellung im Freizeitverhalten sowie im Beruf erhoben.

244 (49%) der Patienten und 165 (50%) der Kontrollen sind männlich. Das mediane Alter der Basaliompatienten beträgt 69.5 Jahre (27-92 Jahre), der Melanompatienten 55 Jahre (19-84 Jahre) und der Kontrollpersonen 57 Jahre (19-90 Jahre).

Die Auswertung der polytomen Zielgröße mittels logistischer Regression erfolgt in SAS 9.3 mit der Prozedur PROC CATMOD. Um wichtige Faktoren zu identifizieren und ihren Einfluss auf die Entstehung von Basaliom und Melanom zu schätzen, werden für alle potentiellen Einflussfaktoren zunächst rohe Odds Ratio (OR) mit 95% Konfidenzintervall (KI) und p-Wert berechnet und zur Präselektion genutzt. Danach werden die präselektierten Faktoren mit einer multiplen polytomen logistischen Regression mit generalisierten Logits ausgewertet [3]. Backward Elimination (mit $p<0.05$) wird zur Variablenselektion verwendet. Der p-Wert des Globaltests, der alle Paarvergleiche (B vs. K, M vs. K, M vs. B) für einen Faktor einschließt, wird zur Selektion herangezogen. Dargestellt werden für das finale Modell die Odds Ratio mit 95% Konfidenzintervall und p-Wert.

3 Umsetzung in SAS

3.1 Programmcode

Im Rahmen einer explorativen Datenanalyse werden für die 66 potentiellen Risikofaktoren in einem ersten Auswertungsschritt rohe Odds Ratio mit 95% Konfidenzintervall berechnet. Dies erfolgt für die Vergleiche Basaliom versus Kontrolle, Melanom versus Kontrolle und Melanom versus Basaliom. Die Präselektion der Faktoren auf Basis der rohen Odds Ratio erfolgt mittels folgender Kriterien: (1.) 95% Konfidenzintervall schließt 1 aus oder (2.) Faktor ist bekannt, in Literatur unter Diskussion oder assoziiert mit Sonnenexpositionsverhalten.

Für die nachfolgende Backward Elimination stehen 37 präselektierte Einflussfaktoren zur Verfügung. Nach abgeschlossener Backward Elimination wird im Endmodell der multiplen polytomen logistischen Regression bei drei Ausprägungen in der Zielgröße

und 16 selektierten Einflussfaktoren die Berechnung der Effektschätzer wie folgt umgesetzt:

```

PROC CATMOD DATA=hautkrebs ORDER=internal;
  DIRECT alter;
  MODEL kollektiv = eKberg eKbrand e20garten e20brand landwirt
                  solarium hauttyp haarfarbe atopie elastose
                  verrucae cheilitis keratose lentsolares
                  wohnort alter / ml clparm;

  CONTRAST 'B vs K: Kindheit Bergaktivität' @1 eKberg -2 / EST=exp;
  CONTRAST 'M vs K: Kindheit Bergaktivität' @2 eKberg -2 / EST=exp;
  CONTRAST 'M vs B: Kindheit Bergaktivität' @1 eKberg 2
                                              @2 eKberg -2 / EST=exp;
  .
  .
  .
  CONTRAST 'B vs K: Landwirtschaft (hauptberuflich vs nein)'
          @1 landwirt -2 -1 / EST=exp;
  CONTRAST 'B vs K: Landwirtschaft (nebenberuflich vs nein)'
          @1 landwirt -1 1 / EST=exp;
  CONTRAST 'M vs K: Landwirtschaft (hauptberuflich vs nein)'
          @2 landwirt -2 -1 / EST=exp;
  CONTRAST 'M vs K: Landwirtschaft (nebenberuflich vs nein)'
          @2 landwirt -1 1 / EST=exp;
  CONTRAST 'M vs B: Landwirtschaft (hauptberuflich vs nein)'
          @1 landwirt 2 1
          @2 landwirt -2 -1 / EST=exp;
  CONTRAST 'M vs B: Landwirtschaft (nebenberuflich vs nein)'
          @1 landwirt 1 -1
          @2 landwirt -1 1 / EST=exp;
  .
  .
  .
  CONTRAST 'B vs K: Alter (J)' @1 alter 1 / EST=exp;
  CONTRAST 'M vs K: Alter (J)' @2 alter 1 / EST=exp;
  CONTRAST 'M vs B: Alter (J)' @1 alter -1
                              @2 alter 1 / EST=exp;
run;

```

Im Folgenden werden einige wichtige Statements und Optionen des vorgestellten SAS-Programms erläutert:

- In PROC CATMOD sind generalisierte Logits als default voreingestellt.
- **DIRECT:** dieses Statement definiert eine quantitative Einflussgröße
- **clparm:** die OPTION im MODEL Statement gibt die Konfidenzbereiche für die Schätzer aus
- **CONTRAST:** Definition der Schätzer, d.h. von Kontrasten, die sich auf Faktoren des MODEL Statements beziehen und für die OR und Konfidenzintervall ausgegeben werden

- **@n:** Ansprechen des jeweiligen Response Profiles auf Basis der generalisierten Logit Funktionen:
 Logit 1 mit $\ln p_1/p_3$ für den Vergleich B zu K,
 Logit 2 mit $\ln p_2/p_3$ für den Vergleich M zu K
- **EST=exp:** ESTIMATE=*keyword*, diese OPTION gibt die geschätzten ORs zum Zeichnen der Forest Plots [4] aus, wobei das *keyword* „exp“ (exponentiated contrast) verwendet wird

3.2 Parametrisierung

Die folgende Tabelle zeigt, wie binäre oder polytome Faktoren parametrisiert werden. Stetige Faktoren werden mit ihrer jeweiligen Ausprägung berücksichtigt.

Tabelle 1: Parametrisierung

Einflussgröße	Merkmal z.B.	Level	Codierung Design-Matrix		
binär	Bergaktivität	nein	1	1	
		ja	2	-1	
polytom	Landwirtschaft	nein	1	1	0
		nebenberuflich	2	0	1
		hauptberuflich	3	-1	-1

3.3 Erläuterung

Als Link-Funktion sind generalisierte Logits voreingestellt:

logit p1 = $\ln p_1/p_3$ für den Vergleich Basaliom zu Kontrolle

logit p2 = $\ln p_2/p_3$ für den Vergleich Melanom zu Kontrolle

Die Modellgleichung für das Response Profile 1 lautet für die Parameter β_1 und β_2 , die sich auf den Einflussfaktor Landwirtschaft beziehen:

$$\text{logit } p_1 = \dots + \beta_1 x_1 + \beta_2 x_2 + \dots$$

Die Berechnung des Kontrasts für den Effekt „hauptberuflich versus nicht in der Landwirtschaft tätig“ erfolgt damit als:

$$\begin{aligned} & \beta_1 * (-1) + \beta_2 * (-1) - (\beta_1 * 1 + \beta_2 * 0) \\ & = \beta_1 * (-2) + \beta_2 * (-1) \end{aligned}$$

Die Berechnung des Kontrasts für den Effekt „nebenberuflich versus nicht in der Landwirtschaft tätig“ erfolgt damit als:

$$\begin{aligned} & \beta_1 * 0 + \beta_2 * 1 - (\beta_1 * 1 + \beta_2 * 0) \\ & = \beta_1 * (-1) + \beta_2 * 1 \end{aligned}$$

4 Ergebnis

Nach Berechnung der rohen Odds Ratio und Präselektion der Risikofaktoren wird für 37 Faktoren eine multiple polytome logistische Regression mit Backward Elimination durchgeführt, um wichtige Faktoren zu identifizieren und ihren Einfluss zu schätzen.

Dargestellt werden für das ermittelte Endmodell, mit dem oben gezeigten Programm, die Odds Ratio mit 95% Konfidenzintervall und der p-Wert des Globaltest, der alle Paarvergleiche (Basaliom versus Kontrolle, Melanom versus Kontrolle, Melanom versus Basaliom über alle Ausprägungen des Einflussfaktors) einschließt.

Eine wichtige Rolle für die Entstehung der beiden Erkrankungen spielen die folgenden 16 selektierten Faktoren:

Tabelle 2a: Finales Modell nach Backward Elimination

a) Basaliom vs. Kontrolle
OR mit 95% KI

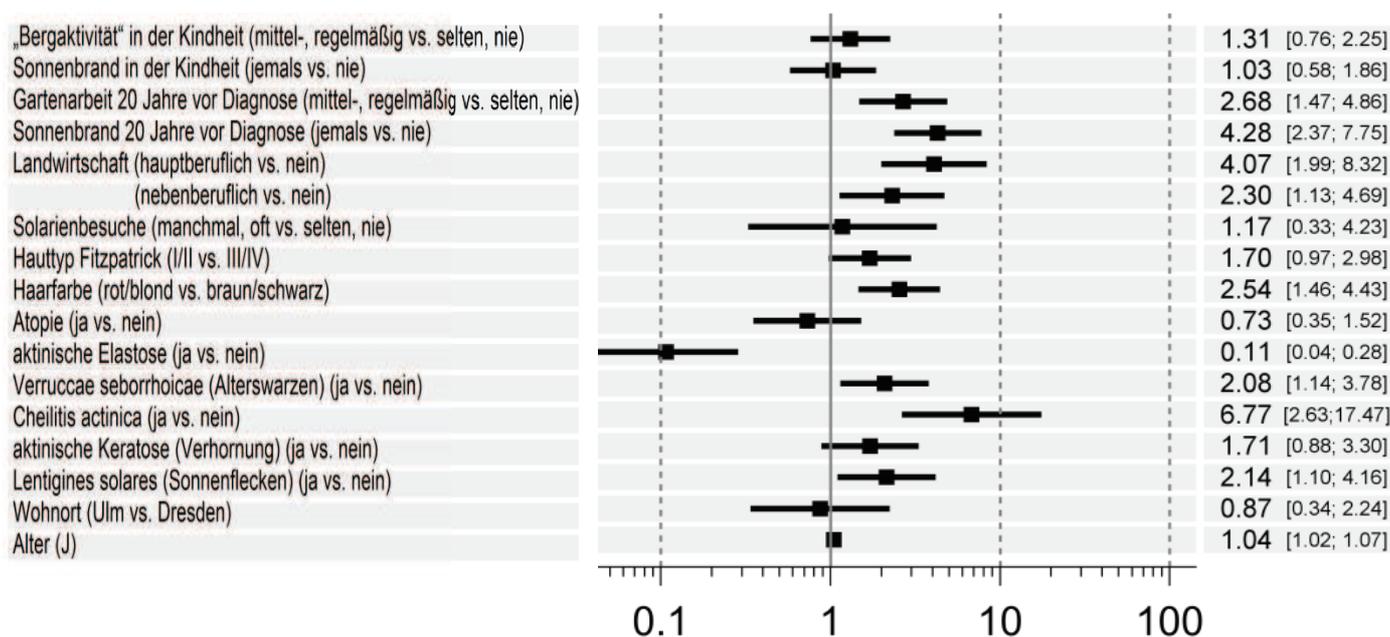


Tabelle 2b: Finales Modell nach Backward Elimination

b) Melanom vs. Kontrolle
OR mit 95% KI

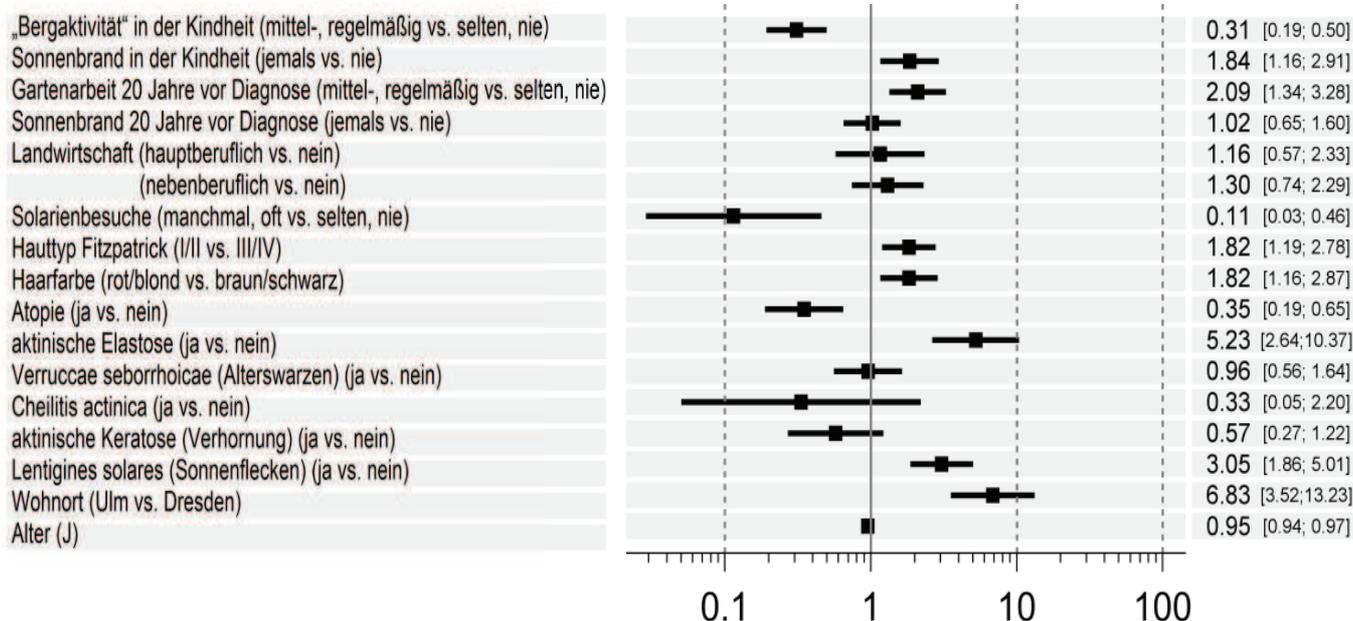


Tabelle 2c: Finales Modell nach Backward Elimination

c) Melanom vs. Basaliom
OR mit 95% KI

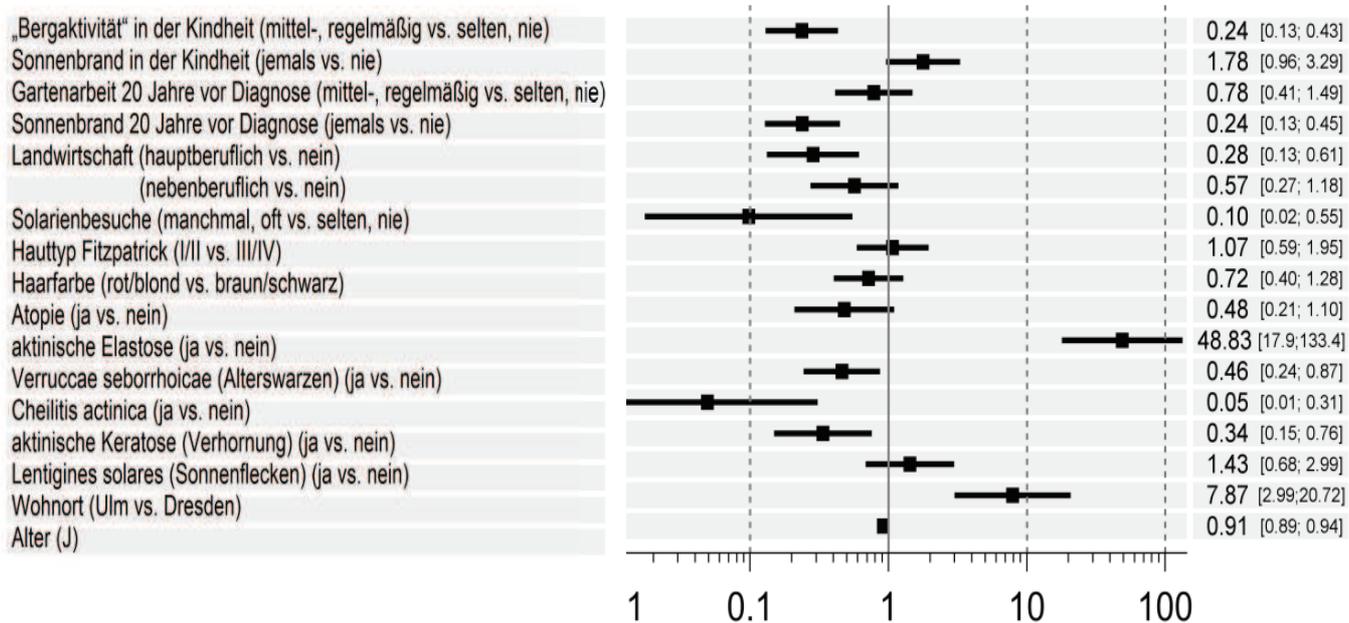


Tabelle 2d: Finales Modell nach Backward Elimination

d) p-Wert

„Bergaktivität“ in der Kindheit (mittel-, regelmäßig vs. selten, nie)	< 0.001
Sonnenbrand in der Kindheit (jemals vs. nie)	0.026
Gartenarbeit 20 Jahre vor Diagnose (mittel-, regelmäßig vs. selten, nie)	< 0.001
Sonnenbrand 20 Jahre vor Diagnose (jemals vs. nie)	< 0.001
Landwirtschaft (hauptberuflich vs. nein)	0.001
(nebenberuflich vs. nein)	
Solarienbesuche (manchmal, oft vs. selten, nie)	0.007
Hauttyp Fitzpatrick (I/II vs. III/IV)	0.012
Haarfarbe (rot/blond vs. braun/schwarz)	0.001
Atopie (ja vs. nein)	0.004
aktinische Elastose (ja vs. nein)	< 0.001
Verrucae seborrhoicae (Alterswarzen) (ja vs. nein)	0.029
Cheilitis actinica (ja vs. nein)	< 0.001
aktinische Keratose (Verhornung) (ja vs. nein)	0.029
Lentigines solares (Sonnenflecken) (ja vs. nein)	< 0.001
Wohnort (Ulm vs. Dresden)	< 0.001
Alter (J)	< 0.001

5 Schlussfolgerung

Eine Fall-Kontroll-Studie mit polytomer Zielgröße kann in SAS 9.3 mit der Prozedur CATMOD ausgewertet werden. Die Backward Elimination muss schrittweise von Hand durchgeführt werden. Der dafür entwickelte SAS-Code für PROC CATMOD ist leicht zu adaptieren.

Das frei verfügbare Macro zur graphischen Darstellung der Forest Plots [4] kann auch für komplexe Anwendungssituationen modifiziert werden.

Die dargestellte Vorgehensweise ist ebenfalls anwendbar auf andere Studiendesigns mit kategorialer Zielgröße.

Literatur

- [1] Karim-Kos H, de Vries E, Soerjomataram I, Lemmens V, Siesling S, Coebergh JW: Recent trends of cancer in Europe. A combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s. *Eur J Cancer* 2008; 44: 1345–1389.
- [2] Kaskel P, Lange U, Sander S, Huber MA, Utikal J, Leiter U, Krähn G, Meurer M, Kron M: Ultraviolet exposure and risk of melanoma and basal cell carcinoma in Ulm and Dresden, Germany. *J Eur Acad Dermatol Venereol*. 2014 Mar 31. doi: 10.1111/jdv.12488. [Epub ahead of print]
- [3] Hosmer DW, Lemeshow S: *Applied Logistic Regression*. Wiley, New York, 1989: 25-36
- [4] http://blogs.sas.com/content/graphicallspeaking/files/2012/09/ForestPlot_931.txt