

Berechnung des relativen Überlebens auf Basis von Krebsregisterdaten mit SAS

Olaf Schoffer
Tumorepidemiologie,
Universitäts KrebsCentrum,
TU Dresden
Fetscherstraße 74
01307 Dresden
olaf.schoffer@uniklinikum-
dresden.de

Antje Niedostatek
Regionales Klinisches
Krebsregister Dresden

Fetscherstraße 74
01307 Dresden

antje.niedostatek@uniklinikum-
dresden.de

Stefanie J. Klug
Tumorepidemiologie,
Universitäts KrebsCentrum,
TU Dresden
Fetscherstraße 74
01307 Dresden
stefanie.klug@uniklinikum-
dresden.de

Zusammenfassung

Relative Überlebensraten eignen sich zum Vergleich von Überlebensangaben zwischen Patientengruppen mit unterschiedlicher Altersstruktur. Sie sind daher insbesondere bei der Auswertung von Krebsregisterdaten als gebräuchliche Methode etabliert. Nachstehend wird die Methodik zur Berechnung relativen Überlebens sowie zentrale Bestandteile der Umsetzung dieser Methodik mit SAS präsentiert. Abschließend werden beispielhafte Ergebnisse des vorliegenden Programms präsentiert und das Programm mit einer Auswahl anderer technischer Lösungen verglichen.

Schlüsselwörter: Relatives Überleben, Krebsregister, natürliche Mortalität, Kaplan-Meier-Verfahren, Nelson-Aalen-Schätzer, Ederer II, PROC LIFETEST

1 Einführung

Die Berechnung von Überlebendauerverteilungen ist eine essentielle Anforderung der epidemiologischen Forschung für Erkrankungen mit erhöhtem Sterberisiko. Neben der spezifischen Erkrankung beeinflusst die Altersstruktur der betrachteten Patientengruppe die Überlebendauerverteilung erheblich. Zur besseren Vergleichbarkeit wird der Einfluss der Altersstruktur insbesondere für Analysen auf Basis von Krebsregisterdaten häufig durch Verwendung relativer Überlebensraten eliminiert.

2 Methoden

Die bekannteste Methode zur Bestimmung von Überlebensdauerverteilungen ist das Kaplan-Meier-Verfahren. Hierbei wird die Überlebensdauerverteilung für Ereignisse an den Zeitpunkten $t_1 < t_2 < \dots < t_k$ bestimmt als

$$\hat{S}(t) = \begin{cases} 1, & t = 0 \\ \prod_{i=1}^j \left(1 - \frac{d_i}{n_i}\right), & t_j \leq t < t_{j+1} \leq t_k \\ \prod_{i=1}^k \left(1 - \frac{d_i}{n_i}\right), & t \geq t_k \end{cases}$$

wobei n_i die Anzahl der Individuen unter Risiko zum Zeitpunkt t_i und d_i die Anzahl der Ereignisse zu diesem Zeitpunkt angibt [2]. Ein Individuum steht zu einem bestimmten Zeitpunkt unter Risiko, wenn es bis dahin kein Ereignis (Versterben) aufweist oder zensiert ist (Ausscheiden aus der beobachteten Population ohne Ereignis). Üblicherweise wird die Überlebensdauerverteilung nur bis zum Zeitpunkt des letzten Ereignisses dargestellt bzw., wenn nach dem letzten Ereignis noch Zensierungen auftreten, bis zum Zeitpunkt der letzten Zensierung.

Die Auswahl der Population erfolgt zumeist, indem ein bestimmter Zeitraum vorgegeben wird, in welchem ein Individuum in die Population eintreten kann (beispielsweise Diagnosedatum in den Jahren 2000 bis 2009). Dieses Vorgehen wird als Kohortenansatz bezeichnet. Beginn und Ende dieses Zeitraums sind dabei passend zur interessierenden Fragestellung zu wählen. Aufgrund des medizinischen Fortschritts sollte für aktuelle Überlebensschätzungen der Beginn des Zeitraums nicht zu weit in die Vergangenheit verlegt werden. Zugleich ist das Ende des Zeitraums so zu wählen, dass für die betrachteten Individuen ein ausreichend langer Beobachtungszeitraum sichergestellt ist (zur Berechnung einer validen 5-Jahres-Überlebensrate sollten genügend Individuen mindestens 5 Jahre ab Eintritt in die Population überlebt haben).

Die ebenfalls gebräuchliche Aktuarialschätzung [7] unterscheidet sich von der Kaplan-Meier-Schätzung dadurch, dass einerseits die Zeitpunkte t_i nicht die Ereigniszeitpunkte sondern Grenzen (äquidistanter) fest gewählter Intervalle darstellen und andererseits in obiger Gleichung n_i durch $n_i - (c_i/2)$ ersetzt wird, wobei c_i die Anzahl der Zensierungen im Intervall $[t_{i-1}; t_i)$ angibt. Da die Intervalle zumeist breiter als die minimal möglichen Zeitabstände gewählt werden (z. B. jahres- vs. tagesgenau), muss hier eine entsprechende Ungenauigkeit in Kauf genommen werden.

Als eine weitere Methode bietet sich die Konstruktion über den Nelson-Aalen-Schätzer für die kumulative Hazard-Funktion an. Dieser ist definiert als

$$\tilde{H}(t) = \begin{cases} 1, & t = 0 \\ \sum_{i=1}^j \frac{d_i}{n_i}, & t_j \leq t < t_{j+1} \leq t_k \\ \sum_{i=1}^k \frac{d_i}{n_i}, & t \geq t_k \end{cases}$$

wobei t_i , d_i und n_i wie für den Kaplan-Meier-Schätzer definiert sind. Daraus ergibt sich der Breslow-Schätzer für die Überlebensrate

$$\tilde{S}(t) = e^{-\tilde{H}(t)}$$

mit dem geschätzten Standard-Fehler (Greenwood-Formel, [12])

$$\hat{\sigma}(\tilde{S}(t_j)) = \tilde{S}(t_j) \sqrt{\sum_{i=1}^j \frac{d_i}{n_i(n_i - d_i)}}.$$

Der Breslow-Schätzer ist zum Kaplan-Meier-Schätzer asymptotisch äquivalent und modelltheoretisch gut begründet (d. h. der Nelson-Aalen-Schätzer lässt sich mittels Martingaltheorie über Zählprozesse herleiten [1]). Die Näherung von Kaplan-Meier- und Breslow-Schätzer wird nach einem Umformungsschritt deutlich. Wird im Breslow-Schätzer die Summe aus dem Nelson-Aalen-Schätzer als Produkt aus der Exponentialfunktion herausgezogen, verbleibt im Produkt der Term $\exp(-d_i/n_i)$. Die Differenz zum für den Kaplan-Meier-Schätzer an derselben Stelle zu findenden Term $(1-d_i/n_i)$ ist umso kleiner, je kleiner das Verhältnis d_i/n_i ist. Die Approximation wird somit umso besser, je größer die Population unter Risiko im Bezug auf die Zahl der Ereignisse pro Zeitpunkt ist.

Die Vor- und Nachteile beider Methoden werden in [6] diskutiert, jedoch lässt sich kein eindeutiger Vorzug für eine Methode ableiten. Die deutlichsten Unterschiede werden dabei wie erwartet im Small-Sample-Bereich festgestellt.

Der Standardfehler $\hat{\sigma}$ für beide Methoden wird in SAS unter Ausnutzung der asymptotischen Normalität von $\tilde{S}(t)$ bzw. $\hat{S}(t)$ geschätzt. Es ist dabei nicht auszuschließen, dass die Konfidenzgrenzen von $S(t)$ bei üblicher Konstruktion außerhalb des zulässigen Bereiches $[0;1]$ liegen. Um dies zu vermeiden, wird das $(1-\alpha)$ -Konfidenzintervall alternativ für eine geeignete Transformation g von $S(t)$ konstruiert (unter Normalverteilungsannahme für $g(S(t))$) [3]

$$KI_{1-\alpha}(S(t)) = \left[g^{-1} \left\{ g(\tilde{S}(t)) \pm u_{\frac{\alpha}{2}} \cdot g'(\tilde{S}(t)) \cdot \hat{\sigma}(\tilde{S}(t)) \right\} \right]$$

wobei g' die erste Ableitung und g^{-1} die Inverse von g ist sowie u_α das α -Quantil der Standardnormalverteilung. Eine geeignete Transformation stellt die LOGLOG-Transformation $g(x) = \log(-\log(x))$ [12] dar, welche nachstehend auch in der Umsetzung mit SAS angewendet wird.

Für etliche Anwendungsbereiche, nicht zuletzt die Auswertung von Krebsregisterdaten, ist diese Form der Überlebenszeitschätzung jedoch nicht ausreichend. Gesucht ist vielmehr eine Angabe zur Mortalität, die unabhängig von Alters- und Geschlechtsstruktur der betrachteten Population ist, um beispielsweise internationale Vergleiche der Mortalität spezifischer Erkrankungen zu ermöglichen. Hier kommt das Konzept des relativen Überlebens zur Anwendung. Dessen Grundidee besteht darin, die in den oben beschriebenen Verfahren betrachtete Gesamtmortalität aufzuteilen in die natürliche Mortalität und die Exzessmortalität (das diesem Vorgehen zugrundeliegende additive Hazardmodell wird beispielsweise in [13] beschrieben). Die natürliche Mortalität zum Zeitpunkt t beschreibt das anteilige Versterben einer zur betrachteten Population unter Risiko geschlechts- und altersidentischen Population aus der Standardbevölkerung. Die Mortalität der Standardbevölkerung lässt sich aus geschlechts- und altersspezifischen Mortalitätstabellen der amtlichen Statistik, beispielsweise den Periodensterbetafeln des Statistischen Bundesamtes [17], ablesen. Die nach Elimination der natürlichen Mortalität aus der Gesamtmortalität verbleibende Exzessmortalität beschreibt dann die zusätzliche Sterblichkeit aufgrund der spezifischen Erkrankung der betrachteten Population. Das zur Berechnung der natürlichen Mortalität verwendete Ederer II-Verfahren [9] sowie die daraus folgende Ableitung des relativen Überlebens werden nachstehend kurz heuristisch beschrieben.

1. Replikation der Angaben zur Population für jeden Ereigniszeitpunkt (d. h. kartesisches Produkt von n_0 Individuen und k Ereigniszeitpunkten).
2. Einschränkung der erhaltenen Datenbasis für jeden Ereigniszeitpunkt jeweils auf die zugehörige Population unter Risiko¹.
3. Bestimmung des Alters aller Individuen zu jedem Ereigniszeitpunkt.
4. Zuordnung der natürlichen Mortalität (jahres-, alters- und geschlechtsspezifisch²) für jedes Individuum zu jedem Ereigniszeitpunkt.
5. Pro Ereigniszeitpunkt: Mittelwertbestimmung über die individuellen natürlichen Mortalitäten aller Individuen unter Risiko zum jeweiligen Ereigniszeitpunkt³.
6. Gewichtung der mittleren Mortalität zum Zeitpunkt t_i mit der seit dem vorangegangenen Ereigniszeitpunkt vergangenen Zeit ($t_i - t_{i-1}$).
7. Kumulation der gewichteten natürlichen Mortalität über alle Zeitpunkte t_i zu: $H^*(t)$ ($t_i \leq t < t_{i+1}$).

¹ Da die aus dem kartesischen Produkt hervorgehende Datenbasis schnell sehr groß wird, ist hier eine frühzeitige Einschränkung sinnvoll. Andernfalls entstehen aufgrund der SAS-spezifischen Datenhaltung unnötig lange Rechenzeiten. Zudem empfiehlt es sich, die resultierende Datenbasis komprimiert abzulegen. Eine weitere Reduktion bietet sich nach dem dritten Schritt an, indem die für den vierten Schritt relevanten Merkmalskombinationen nicht mehrfach mitgeführt, sondern die Zahl der Wiederholungen in einer Häufigkeitsvariablen aufgefangen und weiterverarbeitet wird. Eine Reduktion der aus dem kartesischen Produkt hervorgehenden Datenbasis durch Vergrößerung der Zeitangaben ist ebenfalls denkbar, führt allerdings zu Genauigkeitseinbußen.

² Sofern geeignet strukturierte Mortalitätsdaten vorliegen, kann die Zuordnung problemlos weiter stratifiziert werden (z. B. regional, ethnisch und soziodemografisch).

³ Damit erhält man zur jeweiligen Population unter Risiko den bezüglich der alters- und geschlechtsidentischen Referenzpopulation erwarteten Anteil von Verstorbenen. In der amtlichen Statistik wird die Mortalität zumeist jahresbezogen angegeben, daher bezieht sich der genannte Anteil auf ebendiesen Zeitraum und muss im sechsten Schritt noch mit dem Abstand zwischen den zugehörigen Ereigniszeitpunkten gewichtet werden.

Analog zur Herleitung des Breslow-Schätzers aus dem Nelson-Aalen-Schätzer errechnet sich die erwartete Überlebensrate aus der natürlichen Mortalität gemäß

$$S^*(t) = e^{-H^*(t)}.$$

Damit ergibt sich die relative Überlebensrate als

$$S_{rel}(t) = \frac{\tilde{S}(t)}{S^*(t)}.$$

Da die natürliche Mortalität aus der gesamten Standardbevölkerung bestimmt wird und somit im Prinzip keinen Schätzfehler hat, gilt aufgrund des additivem Hazardmodells für den geschätzten Standard-Fehler

$$\hat{\sigma}(S_{rel}(t)) = \hat{\sigma}(\tilde{S}(t)).$$

Ein relatives Überleben von x% beschreibt den Anteil der Überlebenden, nicht bezogen auf die Ausgangspopulation sondern bezogen auf den Teil, welcher auch in der alters- und geschlechtsidentischen Standardbevölkerung überlebt hätte. Wurde beispielsweise mittels Breslow-Schätzer ein 5-Jahres-Gesamtüberleben von 60% berechnet, während in diesem Zeitraum die Referenzpopulation ein erwartetes Überleben von 90% aufweist, beträgt das relative Überleben gerade 66,7% (=60% / 0,9). In Einzelfällen kann das relative Überleben auch mehr als 100% betragen (wenn das Gesamtüberleben der betrachteten Population größer ist als das erwartete Überleben der Standard-Referenzpopulation).

3 Umsetzung in SAS

Die nachfolgend vorgestellte Umsetzung in SAS stellt eine Weiterentwicklung des Programms RSURV [13] dar. Allerdings wird nicht das komplette Programm sondern nur einzelne wesentliche Programmschritte vorgestellt. Auf Nachfrage wird das Programm aber gern vom Autor zur weiteren Verwendung und Weiterentwicklung zur Verfügung gestellt.

Die vorgestellten Programmschritte stellen selbst keine wesentlichen Neuerungen in der Anwendung von SAS-Syntax dar, wohingegen diese Kombination bekannter Programmier-elemente mit dem Ziel der Schätzung relativer Überlebensraten bisher noch nicht andernorts beschrieben wurde.

Nachstehend wird die Umsetzung einer stratifizierten Schätzung beschrieben. Die notwendigen Vereinfachungen für den unstratifizierten Fall dürften daraus aber leicht abzuleiten sein.

3.1 Makro-Funktionalität

Die vorgestellte Schätzung wurde in ein Makro eingebettet, um einerseits die Flexibilität in der Programmausführung nutzen zu können, etwa wiederholte (%DO <...>) und bedingte (%IF <...>) Ausführung von Programmcode, und andererseits einen nutzer-

freundlichen Aufruf der Schätzung zu ermöglichen. Das Makro ist bisher mittels Positionsparametern definiert: `%macro rsurv_s(ds, strat, nzus, tit2);`.

Nachstehend werden hiervon die Makrovariablen für die Namen des basierenden Datensatzes⁴ (`&ds`) sowie der Stratifizierungsvariable (`&strat`) benötigt. Als weitere Parameter können gegebenenfalls ein Text für Untertitel im Output (`&tit2`) sowie ein Dateinamenzusatz (`&nzus`) angegeben werden.

Neben diesen werden innerhalb des Makros globale, vorab definierte Makrovariablen genutzt. Diese betreffen das Ausgabeformat (`&kana1`), z. B. PDF oder RTF, die maximal darzustellende Beobachtungszeit (`&maxtime`) sowie die Pfadangabe für die Dateiausgabe (`&path`).

Eine weitere Erläuterung der verwendeten Makrofunktionalität erübrigt sich an dieser Stelle, hierfür sei stattdessen auf die umfangreiche Literatur zur SAS-Makroprogrammierung hingewiesen, beispielsweise [14] und [15].

3.2 PROC LIFETEST mit ODS

Zentraler Bestandteil der Schätzung relativer Überlebensraten ist zunächst die Schätzung des Gesamtüberlebens (mittels Kaplan-Meier- oder Breslow-Schätzer). Dies ist in SAS ab Version 9.22 mit der Prozedur LIFETEST [16] umgesetzt, sodass keine Eigenentwicklungen erforderlich sind. Soll das Programm mit älteren SAS-Versionen genutzt werden, kann es gegebenenfalls durch Weglassen der dort noch nicht definierten Optionen (und entsprechend reduzierter Funktionalität) lauffähig gemacht werden.

Nachstehend wird zunächst der im Programm verwendete Prozeduraufruf dargestellt und anschließend die verwendeten Anweisungen und Optionen kurz beschrieben. In allen nachfolgenden Programmteilen wird jeweils das Vorgehen inklusive Stratifizierung beschrieben, die leicht auf eine nichtstratifizierte Variante reduziert werden kann.

```
PROC LIFETEST DATA=&ds METHOD=BRESLOW NELSON ATRISK
              INTERVALS=(0 TO &maxtime BY 1)
              CONFTYPE=LOGLOG OUTSURV=&ds._cl;
  TIME time*status(0);
  STRATA &strat;
RUN;
```

⁴ Erwartet wird ein Datensatz mit folgenden Variablen: `patid` (eindeutige Patientenkenntung), `diagnosedatum` (taggenaues Datum der Diagnose), `geburtsdatum` (taggenaues Geburtsdatum), `time` (Zahl der Tage zwischen Datum der Diagnose und Ereignis- oder Zensorisierungsdatum), `status` (`0` $\hat{=}$ zensorisierte Beobachtung, `1` $\hat{=}$ Beobachtung mit Ereignis), `geschlecht` (`1` $\hat{=}$ männlich, `2` $\hat{=}$ weiblich).

Anweisungen:

- `STRATA` – Stratifizierungsvariable(n); in Klammern können sogar Bereiche zur Kategorisierung angegeben werden
- `TIME` – Angabe der Variablen für Zeitbezug und Zensorisierung; in Klammern wird hinter der Zensorisierungsvariable der Wert festgelegt, welcher eine Zensorisierung anzeigt

Optionen⁵:

- `ATRISK` – [NEU] Ausgabe der Anzahl an Individuen unter Risiko im Textoutput
- `CONFTYPE=ASINSQRT|LOGLOG|LINEAR|LOG|LOGIT` – [NEU] Angabe der Transformation zur Berechnung der Konfidenzgrenzen
- `INTERVALS=` – Ausgabe der Überlebenszeitschätzung im Textoutput nur für bestimmte Zeitpunkte
- `METHOD=BRESLOW|FH` – [NEU] Auswahl des Breslow-Schätzers zur Berechnung des Gesamtüberlebens (die Methode `FH` – Fleming-Harrington ist eine auf dem Breslow-Schätzer basierende Modifikation zur Auflösung verbundener Werte) zusätzlich zu bereits früher verfügbaren Methoden `KM` (Kaplan-Meier) und `LT` (Aktuarialschätzung)
- `NELSON` – [NEU] Ausgabe des Nelson-Aalen-Schätzers im Textoutput
- `OUTSURV=` – Ausgabedatensatz der Überlebenszeitschätzung und der zugehörigen Konfidenzgrenzen

Die Ergebnisausgabe wird im Programm mittels ODS gesteuert. Die Angabe des Ausgabekanals erfolgt dabei mittels der zum Programmbeginn definierten Makrovariable `&kanal` (im Programm wird angenommen dass der Name des Ausgabekanals identisch zur Dateieindung der erzeugten Ausgabe ist, andernfalls ist hier eine entsprechende Anpassung erforderlich). Die Angabe, welche die Ausgabe in das gewünschte Dateiformat umleitet, umschließt alle Prozeduren, welche Text- und Grafikoutput erzeugen und hat folgende Form (die `ODS LISTING`-Angabe verhindert, dass parallel Ergebnisse im SAS-Output-Fenster erzeugt werden).

```
ODS LISTING CLOSE;
ODS &kanal FILE="&path.\rsurv.&kanal";
  <Prozeduren mit Text- und Grafikoutput>
ODS &kanal CLOSE;
ODS LISTING;
```

Zusätzlich zur Text- und Grafikausgabe wird ODS auch genutzt, um die Schätzergebnisse und einige Angaben als SAS-Datensatz zu erhalten, die nicht in dem mittels der `OUTSURV=`-Option von `PROC LIFETEST` erzeugten Datensatz enthalten sind. Dieser Datensatz dient als Grundlage für die anschließende Berechnung des relativen Überlebens und wird später mit dem Datensatz verknüpft, der die Konfidenzschätzung für das Gesamtüberleben enthält. Hierzu wird `PROC LIFETEST` von folgender `ODS OUTPUT`-Angabe umschlossen.

⁵ In Version 9.2 bzw. 9.22 neu eingeführte oder erweiterte Optionen sind mit [NEU] gekennzeichnet.

```
ODS OUTPUT BRESLOWESTIMATES=&ds._breslow;
PROC LIFETEST <...>
ODS OUTPUT CLOSE;
```

Mittels der ODS GRAPHICS-Funktionalität können verschiedene hochauflösende Grafiken zur Überlebenszeitschätzung erzeugt werden (Steuerung über die PLOTS=-Option von PROC LIFETEST). Allerdings lässt sich damit nur das Gesamtüberleben, aber nicht das relative Überleben darstellen (letzteres wird außerhalb von PROC LIFETEST bestimmt). Daher wird statt dieser Art der Darstellung später eine Grafikausgabe mit PROC GPLOT erzeugt. Der Vollständigkeit halber sei dennoch hier die zugehörige Syntax kurz skizziert.

```
ODS GRAPHICS ON;
PROC LIFETEST PLOTS=SURVIVAL(CL) <...>
ODS GRAPHICS OFF;
```

3.3 Kartesisches Produkt mit SQL

In Abschnitt 2 ist beschrieben, dass zur Bestimmung der alters- und geschlechtsidentischen Referenzpopulation zur Population unter Risiko für alle Ereigniszeitpunkte das kartesische Produkt aller Individuen (hier repräsentiert durch die Patientennummern: patid) und Ereigniszeitpunkte zu bilden sowie die erzeugte Datenmenge auf die Population unter Risiko zum jeweiligen Zeitpunkt einzuschränken ist. Im nachstehenden PROC SQL-Programmschritt werden beide Anforderungen umgesetzt (die Einschränkung erfolgt in der WHERE-Bedingung, die übrige SQL-Syntax dient der Erzeugung des kartesischen Produktes). Voraussetzung ist neben dem Ausgangsdatensatz ein weiterer Datensatz, der ausschließlich die Ereigniszeitpunkte je Stratum ohne Wiederholungen enthält. Dies wird mit dem vorgeschalteten PROC FREQ erreicht.

```
PROC FREQ DATA=&ds;
  TABLES time*&strat /NOPRINT OUT=&ds._ereig(KEEP=time &strat
    RENAME=(time=globtime &strat=globstrat));
  WHERE status=1;
RUN;

PROC SQL NOPRINT;
  CREATE TABLE kart AS
  SELECT a.patid, a.time, a.&strat,
         b.globtime, b.globstrat
  FROM &ds a, &ds._ereig b
  WHERE (a.time>=b.globtime AND
        a.&strat=b.globstrat)
  ORDER BY a.patid, b.globtime;
QUIT;
```

3.4 DATA-Step mit LAST

Die Überlebenskurve wird grafisch als Treppenfunktion dargestellt (PROC GPLOT mit SYMBOL1 I=STEPJS). Da der von SAS erzeugte Datensatz nur zu Ereigniszeitpunkten

Werte für die Überlebensrate aufweist, endet dieser Treppenfunktion am letzten Ereigniszeitpunkt. Bei einer Zensierung nach dem letzten Ereigniszeitpunkt wird aber die Überlebenskurve zumeist bis zum Zeitpunkt der letzten Zensierung fortgeschrieben (der Wert der Überlebenskurve ändert sich ab dem letzten Ereigniszeitpunkt jedoch nicht mehr). Um das in SAS darstellen zu können, muss die letzte betrachtete Beobachtung eines Stratums den zuletzt gültigen Wert der Überlebenskurve (`survival`) erhalten. Dies lässt sich mit dem nachstehenden DATA-Step umsetzen. Dabei wird der nach `stratum` sortierte Datensatz mittels der `BY`-Anweisung strukturiert und die letzte Beobachtung jedes Stratums mit `LAST.stratum` identifiziert, sodass die zuvor initialisierte temporäre Variable `_t` an dieser Stelle ausgelesen und der Wert der Überlebenskurve zugewiesen werden kann.

```
DATA &ds._breslow;
  SET &ds._breslow;
  BY stratum;
  RETAIN _t 0;
  IF survival NE . THEN _t=survival;
  IF LAST.stratum THEN survival=_t;
  DROP _t;
RUN;
```

3.5 Ergebnisdarstellung

In dem vorliegend beschriebenen Programm werden neben dem Output von `PROC LIFETEST` (nur Angaben zum Gesamtüberleben) auch Grafiken und tabellarische Übersichten zum erwarteten und relativen Überleben erzeugt. Diese werden nachstehend Auszugsweise am Beispiel von Daten Klinischer Krebsregister (KKR) vorgestellt. Im Rahmen einer Initiative der Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT) lieferten 22 bevölkerungsbezogene und 4 einrichtungsbezogene KKR für eine Auswertung zum malignen Melanom (ICD-10: C43 und D03) mehr als 45.000 Fälle im Zeitraum 2000 bis 2009 gemeldeter Primärtumoren. Davon gehen mehr als 30.000 Fälle in die nachstehend beschriebene Überlebenszeitanalyse ein⁶. Aufgrund der Größe dieser Datenmenge verbunden mit der oben erwähnten asymptotischen Äquivalenz ist eine starke Annäherung des Breslow-Schätzers an den Kaplan-Meier-Schätzer gegeben.

In Abbildung 1 wird das Gesamtüberleben (A) sowie das erwartete (B) und relative Überleben der eingeschlossenen Fälle innerhalb der ersten 5 Jahre nach Diagnose dargestellt. Für das relative Überleben sind zudem die 95%-Konfidenzgrenzen abgebildet. Hier ist gut zu erkennen, wie das relative Überleben durch Gesamtüberleben und erwartetes Überleben bestimmt wird.

⁶ In die Überlebenszeitanalyse wurden nur bevölkerungsbezogene Register mit hinreichend guten Angaben zum Vitalstatus der Krebspatienten eingeschlossen. Fälle, die auf mehrfache Tumormeldungen beim selben Patienten zurückzuführen waren, wurden (bis auf eine Primärmeldung) ausgeschlossen.

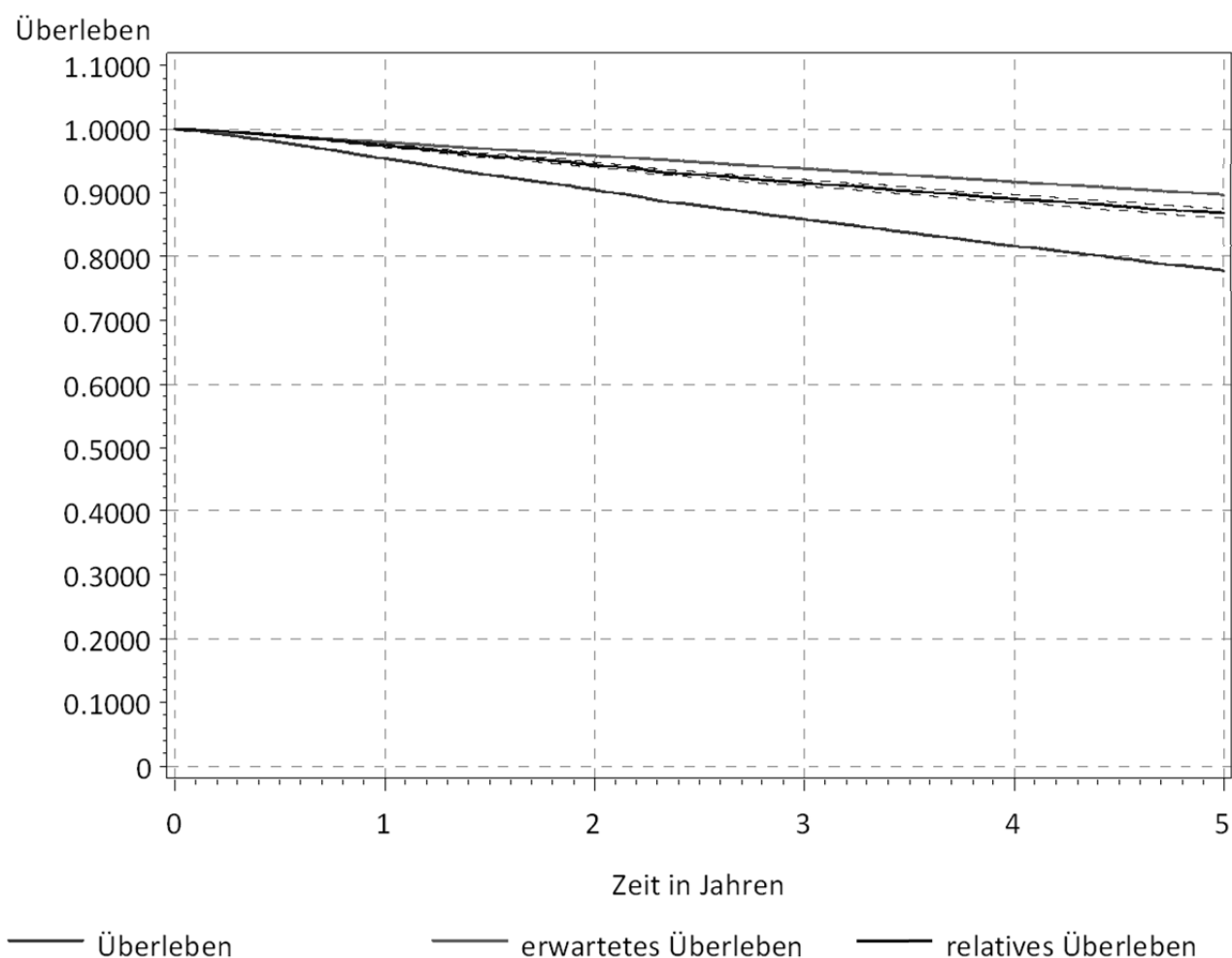


Abbildung 1: Überleben (Gesamt, erwartetes, relatives) ab Diagnose „malignes Melanom“

Abbildung 2 stellt ausschließlich das relative Überleben inklusive 95%-Konfidenzgrenzen dar, allerdings stratifiziert nach UICC-Stadien [18]. Für die Stadien 0 (in situ-Fälle) und 1 (kleine invasive Tumoren ohne Metastasen in umliegenden Lymphknoten oder anderen Organen) liegt das relative Überleben bei etwa 100%, zeitweise sogar darüber. Offensichtlich sind diese Stadien so gut therapierbar, dass gegenüber der Standardbevölkerung kein Mortalitätsnachteil erkennbar ist. Für höhere Stadien (größere Tumoren, Tumoren mit Metastasen in umliegenden Lymphknoten oder Tumoren mit Fernmetastasen) ist hingegen ein deutlicher Unterschied gegenüber der Standardbevölkerung zu erkennen. Die Schwere der Erkrankung hat demnach erhebliche Auswirkungen auf das Überleben der betroffenen Patienten – auch wenn der Einfluss der Altersstruktur der Patienten aus den Angaben zum Überleben herausgerechnet wurde.

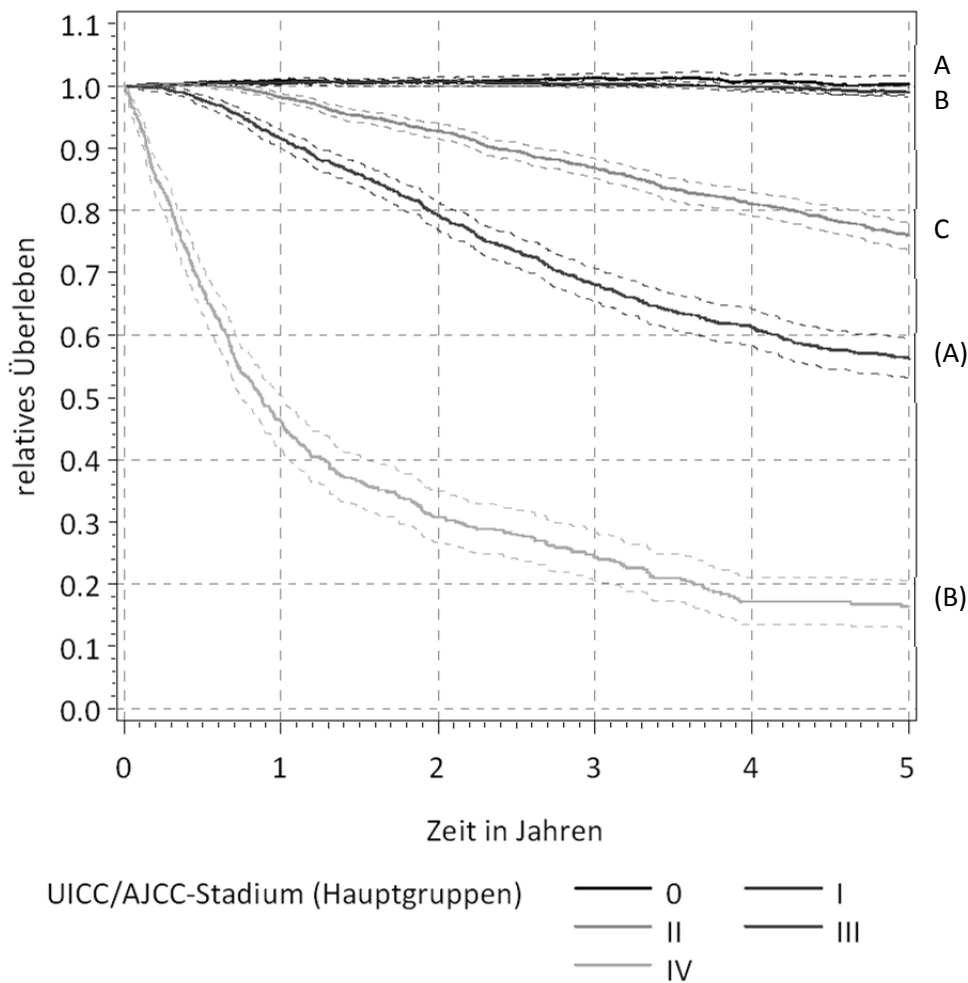


Abbildung 2: Relatives Überleben ab Diagnose „malignes Melanom“ stratifiziert nach UICC-Stadium

Die in Tabelle 1 ersichtliche Übersicht quantifiziert den bereits in Abbildung 1 veranschaulichten Zusammenhang zwischen relativem, erwartetem und Gesamtüberleben. Vergleicht man das Gesamtüberleben zwischen beiden Geschlechtern 5 Jahre nach Diagnose, haben Frauen mit einer Überlebensrate von 81,5% eine deutlich bessere Perspektive als Männer mit einer Überlebensrate von 74,2%. Allerdings hätte auch eine bezüglich des Alters jeweils identisch zur betrachteten Population zusammengesetzte Gruppe aus der Standardbevölkerung für Frauen deutlich größere Überlebensaussichten (erwartete Überlebensrate von 91,4%) als für Männer (erwartete Überlebensrate von 87,9%). Es stellt sich also die Frage, ob das bessere Überleben von an malignem Melanom erkrankten Frauen gegenüber erkrankten Männern allein auf die ohnehin geringere Lebenserwartung der Männer zurückzuführen ist oder ob nach Herausrechnen dieses Effektes ein Unterschied zwischen beiden Geschlechtern verbleibt. Diese Frage beantwortet die Angabe zum relativen Überleben. Hier verbleibt auch nach Herausrechnen des erwarteten Überlebens ein Überlebensvorteil für Frauen.

Tabelle 1: Überleben (Gesamt, erwartetes, relatives) ab Diagnose „malignes Melanom“ stratifiziert nach Geschlecht⁷

Geschlecht	Jahr (nach Diagnose)	Überleben	erwartetes Überleben	relatives Überleben	relatives Überleben 95%-LCL	relatives Überleben 95%-UCL
männlich	1	94,6%	97,5%	97,0%	96,6%	97,4%
	2	88,8%	95,1%	93,4%	92,8%	94,0%
	3	83,3%	92,7%	89,9%	89,0%	90,6%
	4	78,3%	90,3%	86,8%	85,8%	87,7%
	5	74,2%	87,9%	84,4%	83,3%	85,4%
weiblich	1	96,1%	98,3%	97,8%	97,4%	98,1%
	2	92,1%	96,5%	95,4%	94,8%	95,9%
	3	88,4%	94,8%	93,2%	92,5%	93,8%
	4	85,0%	93,1%	91,3%	90,5%	92,1%
	5	81,5%	91,4%	89,2%	88,2%	90,1%

4 Vergleich mit anderen Lösungen

In Tabelle 2 wird die vorliegend beschriebene Umsetzung mit einer Auswahl anderer bestehender Lösungen hinsichtlich der Plattform und verfügbaren Methoden verglichen. In der ersten Zeile ist dabei die oben beschriebene Lösung zu finden.

Tabelle 2: Übersicht über verschiedene Lösungen zur Berechnung des relativen Überlebens (kein Anspruch auf Vollständigkeit)

	Plattform	Methode		
		Überleben ⁸	Erwartetes Überleben	Relatives Überleben
Erweiterung von RSURV [13]	SAS	Kaplan-Meier & Nelson-Aalen	Ederer II	Kohortenansatz
SURVSOFT [10]	Eigenentwicklung	Kaplan-Meier	Ederer II & Hakulinen ⁹	Kohorten- & Periodenansatz ¹⁰
Programm von P. Dickman [8]	SAS	Kaplan-Meier	Ederer II	Kohorten- & Periodenansatz
period(R/H) [4]	SAS / R	Kaplan-Meier	Ederer II & Hakulinen	Periodenansatz

⁷ Die Tabelle wurde zugunsten eines zum vorliegenden Text passenden Layouts aufbereitet. Die im Programm per ODS erzeugten Tabellen haben eine etwas andere Form.

⁸ Ohne Aktuarial- bzw. Sterbetafel-Methode

⁹ Hakulinen [11]: Alternative zu Ederer II zur Berechnung des erwarteten Überlebens

¹⁰ Periodenansatz [5]: Alternative zu Kohortenansatz (Auswahl der Population, sodass möglichst aktueller Stand des Überlebens abgebildet wird)

Eine zu Tabelle 1 analoge Auswertung mit dem Programm SURVSOFT [10] ergibt nahezu identische Ergebnisse für die Prozentangaben und Konfidenzgrenzen zum relativen Überleben (Abweichungen bezüglich 5-Jahres-Überleben erst in zweiter Nachkommastelle der Prozentwerte). Andere Lösungen ergeben hierzu etwas größere Abweichungen (aufgrund anderer Datumsgenauigkeit oder Methodik), die abzuleitenden Aussagen bleiben jedoch vergleichbar.

Neben den dargestellten Unterschieden und Gemeinsamkeiten der verschiedenen Lösungen bestehen teilweise erhebliche Unterschiede hinsichtlich der Ergebnisdarstellung und Bedienung, die an dieser Stelle allerdings nicht detailliert dargestellt werden können.

5 Zusammenfassung und Ausblick

Mit dem vorgestellten Programm steht eine bezüglich der Ergebnisse zu bestehenden Programmen vergleichbare Lösung zur Berechnung relativer Überlebensraten zur Verfügung. Kleine Unterschiede in den Ergebnissen entstehen beispielsweise durch die verschiedene Methodik der Lösungen, etwa bezüglich der Verfahren zur Berechnung des erwarteten Überlebens und des Gesamtüberlebens, der Auswahl der betrachteten Population, der Genauigkeit der Zeitachse sowie der verwendeten Sterbetafeln.

Das Programm kann aufgrund der reduzierten Komplexität hinsichtlich der Syntax sowie der Makrofunktionalität flexibel angepasst werden und bietet das gesamte Spektrum von Methoden, die in SAS 9.22 mit der Prozedur LIFETEST zur Verfügung stehen. Die erzeugten Ergebnisse können mittels ODS und SAS/GRAPH in verschiedenen Dateiformaten ausgegeben werden.

Bei der Programmierung der Syntax wurde auf einen optimierten Programmablauf geachtet. Insbesondere seien hier die Reduktion der Datenbasis bereits bei Erzeugung des kartesischen Produktes sowie die konsequente Ausnutzung der Neuerungen bis SAS 9.22 (LIFETEST, ODS, SQL) genannt.

Dennoch sind Erweiterungen und Verbesserungen des Programms vorgesehen. So sollen die Hakulinen-Methode sowie der Periodenansatz noch in das Programm aufgenommen sowie die Parametrisierung des Makros optimiert werden.

Literatur

- [1] Aalen, O.O.: *Nonparametric inference for a family of counting processes*, The Annals of Statistics, 1978 6:4, 701–726.
- [2] Allison, P.: *Survival analysis using SAS - a practical guide* (2nd ed.), SAS Publishing, Cary, NC. 2010.
- [3] Borgan, Ø., Liestøl, K.: *A Note on Confidence Interval and Bands for the Survival Curves Based on Transformations*, Scand J Statistics, 1990 18, 35–41.

- [4] Brenner, H., Gefeller, O., Hakulinen, T.: *Period analysis for up-to-date cancer survival data: theory, empirical evaluation, computational realization and applications*, Eur J Cancer, 2004 40:3, 326-335.
- [5] Brenner, H., Gefeller, O.: *An alternative approach to monitoring cancer patient survival*. Cancer, 1996 78:9, 2004–2010.
- [6] Colosimo, E.A., Ferreira, F.F., Oliveira, M.D., Sousa, C.B.: *Empirical comparisons between Kaplan-Meier and Nelson-Aalen survival function estimators*, J Stat Comput Simul, 2002 72:4, 299-308.
- [7] Cutler, S.J., Ederer, F.: *Maximum utilization of the life table method in analyzing survival*, J Chronic Dis. 1958 8, 699-712.
- [8] Dickman, P.: *Estimating and modelling relative survival using SAS*, manual: http://www.pauldickman.com/rsmodel/sas_colon/, 2004.
- [9] Ederer, F., Heise, H.: *Instructions to IBM 650 programmers in processing survival computations - methodological note No. 10 - end results section*, National Cancer Institute, Bethesda, MD, 1959.
- [10] Geiss, K., Meyer, M., Radespiel-Tröger, M., Gefeller, O.: *SURVSOF T—Software for nonparametric survival analysis*, Comput Methods Programs Biomed, 2009 96:1, 63-71.
- [11] Hakulinen, T.: *Cancer survival corrected for heterogeneity in patient withdrawal*, Biometrics, 1982 39, 933–942.
- [12] Kalbfleisch, J.D., Prentice, R.L.: *Statistical analysis of failure time data*, Wiley, New York, 1980.
- [13] Koch, R.: *Analyse von Überlebensdauerverteilungen unter Berücksichtigung der Altersstruktur der Patienten in Klinischen Krebsregistern*, in: Jahresbericht 2001 des Tumorzentrums Dresden e.V., 67-74.
- [14] Krämer, W., Schoffer, O., Tschiersch, L.: *Datenanalyse mit SAS (2. Aufl.)*, Springer, Heidelberg, 2008.
- [15] SAS Institute Inc.: *SAS 9.2 Macro Language - Reference*, SAS Publishing, Cary, NC, 2009.
- [16] SAS Institute Inc.: *SAS/STAT 9.22 User's Guide*, SAS Publishing, Cary, NC, 2010.
- [17] Statistisches Bundesamt: *Periodensterbetafeln für Deutschland 1871/1881 bis 2007/2009*, Statistisches Bundesamt, Wiesbaden, 2011.
- [18] Wittekind, C., Meyer, H.-J., Bootz, F. [Hrsg.]: *UICC - TNM Klassifikation maligner Tumoren (6. Aufl.)*, Springer, Heidelberg, 2002.