

## Die Berücksichtigung von „Competing risks“ zur Schätzung therapieassoziierter Mortalität bei allogener Stammzelltransplantation

Pfirschmann M<sup>1,2</sup>, Steigerwald N<sup>1,2</sup>, Maywald O<sup>3</sup>, Reiter A<sup>3</sup>, for the German CML Study Group

<sup>1</sup>Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie - IBE, Ludwig-Maximilians-Universität München, Deutschland

<sup>2</sup>Gesellschaft für Informationsverarbeitung und Statistik in der Medizin e.V., München, Deutschland

<sup>3</sup>III. Medizinische Universitätsklinik, Klinikum Mannheim, Universität Heidelberg, Deutschland

pfj@ibe.med.uni-muenchen.de

**Einleitung und Fragestellung** Bei Patienten mit chronischer myeloischer Leukämie (CML) sollten die Sterbewahrscheinlichkeiten nach allogener Stammzelltransplantation (SZT) betrachtet werden. Beim Vergleich dreier Patientengruppen, die sich zum Zeitpunkt der SZT in unterschiedlichen Krankheitszuständen befanden, interessierten insbesondere die Sterbewahrscheinlichkeiten, welche mit dem Verfahren der SZT an sich assoziiert werden mussten. Eine mögliche Beobachtung des Ereignisses „transplantationsassoziierter Tod“ wurde bei den zuletzt lebenden Patienten durch das Ende der Beobachtungszeit eingeschränkt (Zensierung) und bei anderen Patienten durch das Auftreten eines konkurrierenden Ereignisses („Competing risk“) verhindert. Zwischen beiden möglichen Ursachen für die Nichtbeobachtung eines transplantationsassoziierten Todes besteht ein erheblicher qualitativer Unterschied. Während eine Zensierung die Beobachtung des interessierenden Ereignisses theoretisch zu einem späteren Zeitpunkt erlaubt, schließt das Auftreten eines konkurrierenden Ereignisses die spätere Beobachtung eines therapieassoziierten Todes für alle Zeiten aus. Eine Beschreibung der therapieassozierten Mortalität über die Kaplan-Meier-Methode [1] und der Vergleich geschätzter Sterbewahrscheinlichkeiten der drei Gruppen mittels Logrank-Test [2], indem man das Ende der Beobachtungszeit wie den Zeitpunkt des Auftretens eines konkurrierenden Ereignisses gleichermaßen zensiert, ignoriert den beschriebenen qualitativen Unterschied. Als Alternative zum Kaplan-Meier-Schätzer, haben Kalbfleisch und Prentice [3, 4] den „Cumulative incidence“-Schätzer vorgestellt. Bei Berechnung der hier mit „kumulativen Inzidenzen“ bezeichneten, transplantationsbedingten Sterbewahrscheinlichkeiten werden durch konkurrierende Ereignisse reduzierte Beobachtungswahrscheinlichkeiten des interessierenden Ereignisses sowie Unterschiede zu echten Zensierungen berücksichtigt. Gray [5] entwickelte einen Test, um kumulative Inzidenzen verschiedener Stichproben miteinander vergleichen zu können. Anhand der Daten zum Überleben nach SZT, werden die unterschiedlichen methodischen Ansätze im Umgang mit konkurrierenden Ereignissen einander gegenübergestellt.

**Material und Methoden** Zu 84 Patienten, die vor einer allogenen SZT im Rahmen der deutschen CML-Studien I-III mindestens drei Monate mit Interferon-alpha behandelt worden waren, lagen zum Zeitpunkt der SZT aktuelle Daten zur zytogenetischen Remission bzw. zur Krankheitsphase vor [6]. Die transplantationsassozierte Mortalität dreier Gruppen sollte verglichen werden. Die erste Gruppe (ZR) bestand aus Patienten, die in zytogenetischer Remission transplantiert wurden, d.h. bei welchen der Anteil der CML-typischen Knochenmarkszellen mit abnormem Philadelphia-Chromosom zuvor mit Interferon-alpha-Therapie von 100% auf 0-95% hatte reduziert werden können. Bei der zweiten Gruppe (CP) handelte es sich um Patienten, die sich nicht in Remission, aber – wie die erste Gruppe – immer noch in chronischer Phase befanden. Die Patienten der dritten Gruppe (kCP) konnten keine Remission aufweisen und hatten zusätzlich über den Verlust ihrer chronischen Phase ein fortgeschrittenes Krankheitsstadium erreicht. Ohne Unterscheidung wurde zunächst zum Zeitpunkt der letzten Beobachtung bzw. des Auftretens eines der konkurrierenden Ereignisse „Tod aus anderen Gründen“ oder „Rezidiv“ zensiert und die Sterbewahrscheinlichkeiten der drei Gruppen mit dem Kaplan-Meier-Verfahren geschätzt und per Logrank-Test miteinander verglichen. Beim „Cumulative incidence“-Schätzer werden bis zum Zeitpunkt  $t$  die Summanden aller bisherigen Auftrittszeitpunkte  $j$  des interessierenden Ereignisses  $e$ , mit  $j \leq t$ , aufaddiert. Jeder Summand besteht aus dem Quotienten der Anzahl der Ereignisse zum Zeitpunkt  $j$  dividiert durch die noch bis einschließlich  $j$  unter Risiko stehenden Patienten, multipliziert mit dem Kaplan-Meier-Schätzer zur Wahrscheinlichkeit, kein Ereignis, weder das interessierende noch ein konkurrierendes, bis zum maximalen Ereigniszeitpunkt  $k$ , so dass  $k < j$ , erfahren zu haben. Die so geschätzten kumulativen Inzidenzen der drei Gruppen wurden mit dem Gray-Test verglichen.

**Ergebnisse** Die transplantationsassozierten Sterbewahrscheinlichkeiten wurden zunächst über das Komplement der mit der Kaplan-Meier-Methode berechneten Überlebenswahrscheinlichkeiten geschätzt. Für die Gruppe ZR (25 Patienten) waren die Schätzergebnisse zu den Ereigniswahrscheinlichkeiten für beide Methoden über alle Zeitpunkte identisch. Dies ist genau solange der Fall, bis zum ersten Mal ein konkurrierendes vor dem interessierenden Ereignis beobachtet wurde. Alle fünf transplantationsassozierten Todesfälle traten innerhalb der ersten vier Monate auf. Da die übrigen 20 Patienten alle länger unter Beobachtung standen (1 SZT-unabhängiger Todesfall, 2 Rezidive) betrug die geschätzte transplantationsbedingte Sterbewahrscheinlichkeit von Monat 4 bis zum Jahr 9 0,20. In der Gruppe CP (29 Patienten) wurden alle 14 therapieassozierten Todesfälle in den ersten 15 Monaten registriert. Zwischen dem 10. und 11. Ereignis lag im 6. Monat ein Rezidiv, so dass sich die Kaplan-Meier-Sterbewahrscheinlichkeiten (nach Jahr 1: 0,42) und die kumulativen Inzidenzen (nach Jahr 1: 0,41) ab diesem Zeitpunkt unterschieden. Die 14 übrigen Patienten wurden wieder länger als 15 Monate beobachtet (1 SZT-unabhängiger Todesfall, 3 Rezidive). Die deutlichsten Differenzen traten in der Gruppe kCP (30 Patienten) auf. Drei Rezidive im ersten Jahr führten nach Jahr 1 zur Kaplan-Meier-Sterbewahrscheinlichkeit 0,50 und zur kumulativen Inzidenz 0,54. Insgesamt wurden 17 transplantationsbedingte Todesfälle und fünf Rezidive beobachtet; die acht zensierten Patienten standen wenigstens zwei Jahre unter Beobachtung. Beim Vergleich der Ereigniswahrscheinlichkeiten wiesen der Logrank-Test den p-Wert 0,021 und der Gray-Test den p-Wert 0,063 auf.

**Diskussion** Wegen des im Vergleich zu den konkurrierenden Ereignissen sehr frühen Auftretens der therapieassozierten Todesfälle, bestehen für zwei der drei Gruppen nur relativ geringe Unterschiede zwischen den Kaplan-Meier-Sterbewahrscheinlichkeiten und den kumulativen Inzidenzen. Insofern zeigt das bei unseren Daten auftretende Beispiel keinen Extremfall für methodenabhängige Schätzunterschiede. Dennoch können hier die prinzipiellen Differenzen in den geschätzten Wahrscheinlichkeiten demonstriert werden, bis hin zu verschiedenen Schlussfolgerungen aus den sich unterscheidenden p-Werten zu den beiden Teststatistiken. Sobald vor einem interessierenden Ereignis ein konkurrierendes Ereignis beobachtet wurde, beginnen sich die geschätzten Wahrscheinlichkeiten zum interessierenden Ereignis für die beiden Methoden zu unterscheiden. Die über die Kaplan-Meier-Methode errechneten Wahrscheinlichkeiten liegen nun immer höher und führen dementsprechend zu einer Überschätzung der interessierenden Ereigniswahrscheinlichkeiten: Stehen von  $n$  Patienten zum Zeitpunkt  $j$  noch  $n-m$  unter Beobachtung, so verteilt die Kaplan-Meier-Methode als potentiellen Beitrag eines Patienten zur geschätzten Ereigniswahrscheinlichkeit  $1/(n-m)$  [4]. Dagegen werden bei der kumulativen Inzidenzschätzung über die Multiplikation mit dem Kaplan-Meier-Schätzer alle Ereignisse, auch die konkurrierenden, als beobachtete Endpunkte berücksichtigt. Die durch konkurrierende Ereignisse bedingte Reduktion des noch aufzuteilenden Ereignisbeitrages wird damit beachtet, statt 1 nur  $1-x$  (mit  $0 < x < 1$ ) potentiell unter den  $j$  verbliebenen  $n-m$  Patienten verteilt. Die Zahl  $x$  steht für die Wahrscheinlichkeit, bis  $j$  eines der konkurrierenden Ereignisse beobachtet zu haben (und so das interessierende Ereignis definitiv nie beobachten zu können). In einer „Competing risk“-Situation, in welcher unabhängig vom interessierenden Ereignis konkurrierende Ereignisse auftreten können, die das interessierende Ereignis ausschließen, betrachten wir die Schätzung der Ereigniswahrscheinlichkeiten durch kumulative Inzidenzen und die Verwendung des Gray-Tests beim Gruppenvergleich der so geschätzten Wahrscheinlichkeiten als angemessenere Methoden denn der Rückgriff auf den Kaplan-Meier-Schätzer und den Logrank-Test.

### Literatur

- [1] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. JASA 1958; 53: 457-481.
- [2] Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. JNCI 1959; 22: 719-748.
- [3] Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: John Wiley & Sons; 1980.
- [4] Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Stat Med. 1999; 18: 695-706.
- [5] Gray R.J. A class of k-sample tests for comparing the cumulative incidence of a competing risk. Annals of statistics. 1988; 16: 1141-1154.
- [6] Maywald O, Pfirschmann M, Berger U, Breitschdel L, Gratwohl A, Kolb HJ, Beelen DW, Metzgeroth G, Gnad SU, Hochhaus A, Hasford J, Hehlmann R, Reiter A for the German CML Study Group and the SAKK. Cytogenetic response to prior treatment with interferon-alpha is predictive for survival after allogeneic

hematopoietic stem cell transplantation in chronic myeloid leukemia. *Leukemia* 2006 [Epub ahead of print].