

Raum-Zeit-Analyse der Kinderleukämie in Deutschland mit Kovariablen

Sven SCHMIEDEL

1 Einleitung

Leukämie ist die häufigste Krebserkrankung bei Kindern unter 15 Jahren. Das Risiko für ein Kind an Leukämie zu erkranken ist eins zu 2.000 (KAATSCH & SPIX 2008). Noch vor einigen Jahrzehnten war diese Erkrankung tödlich, heute liegt die 5-Jahres-Überlebenswahrscheinlichkeit in den USA und Europa bei rund 85 %. Jedoch ist diese sehr unterschiedlich je nach Subtyp der Erkrankung (GREAVES 2006).

Die Ätiologie der Leukämie ist weitestgehend unbekannt. Als einziger gesicherter Risikofaktor gilt die ionisierende Strahlung (LIGHTFOOT & ROMAN 2004). Eine Möglichkeit eine allgemeine Tendenz der Erkrankung zur Clusterbildung aufzudecken ist die Verwendung eines großen landesweiten Registers. Daten des Deutschen Kinderkrebsregisters stehen hierfür zur Verfügung.

Viele epidemiologische Studien, die auf räumlichen Daten beruhen, analysieren Regionen um Punktquellen, zum Beispiel Sendeantennen von Radio und Fernsehen (MERZENICH et al. 2008). Meist werden diese Studien initiiert aufgrund der Beobachtung erhöhter Inzidenzrate einer bestimmten Krankheit in einer begrenzten Region, hier kindliche Leukämie (MASKARINEC et al. 1994). Eine Beobachtung ist somit der Auslöser der Studie und nicht eine Hypothese. Diese Studien berichten häufig statistisch signifikante Ergebnisse. Diese können jedoch wegen des Studienansatzes dem sogenannten "Texas sharpshooter" Trugschluss unterliegen. Ein Beispiel hierfür sind die Studien von Dolk zum Zusammenhang zwischen hochfrequenter elektromagnetischer Strahlung und kindlicher Leukämie (DOLK et al. [1], DOLK et al. [2] 1997). Zunächst wurden um einen einzelnen Sender statistisch signifikante Ergebnisse gefunden, nach Einschluss aller Hochleistungssender Großbritanniens war dieselbe Assoziation nicht mehr signifikant. Damit wurde die gefundene Assoziation an einem einzelnen Sender anhand der Erweiterung der Studienregion nicht bestätigt. Beide Artikel wurden gleichzeitig im selben Journal veröffentlicht. Es ist wichtig zu untersuchen, ob eine allgemeine Tendenz bei kindlicher Leukämie besteht zu clustern. Eine Untersuchung von Daten des Deutschen Kinderkrebsregister (WESTERMEIER & MICHAELIS 1995) ergab keinen Hinweis auf eine Tendenz zum Clustern auf Gemeindeebene (Variabilität der Inzidenzraten waren im Rahmen der Annahmen des Poissonmodells). Dies steht im Widerspruch zu den Ergebnissen des EUROCLUS Projekts in dem Regionen aus 17 Ländern gemeinsam untersucht wurden und Hinweise zum Clustern gefunden wurden (ALEXANDER et al. 1997).

Da immer wieder Leukämiecluster gefunden werden (bekannt ist insbesondere das Leukämiecluster in der Elbmarsch in der Nähe des Leistungsreaktors Krümmel), besteht die Vermutung, dass relevante Risikofaktoren in ihrer räumlichen Verteilung eine Rolle spielen oder auch, dass eine allgemeine Tendenz der Erkrankung zur Clusterbildung besteht. Hierbei muss die Entstehung einer Leukämie näher erläutert werden. Eine Hypothese nach Knudsons besagt, dass zwei genetische Veränderungen hinreichend für die Entstehung für Krebs sind (KNUDSON 1971). Diese Hypothese wurde von Greaves auf die kindliche Leu-

kämie übertragen (GREAVES 2002). Hier liegt die erste Veränderung zum präleukämischen Klon schon in der Gebärmutter vor und die zweite Veränderung, welche die Leukämie auslöst, wird in der frühen Kindheit durch bestimmte Expositionen hervorgerufen. Das Auslösen der Erkrankung in der Gebärmutter konnte für die akute lymphoblastische Leukämie bestätigt werden (HJALGRIM et al. 2002). Greaves vermutet, dass Leukämie eine Erkrankung des Immunsystems ist und fehlendes immunologisches Training in der frühen Kindheit dazu beiträgt. Kovariablen, die direkt eine Infektionsrate beschreiben in Deutschland nicht vorhanden, jedoch hat eine Studie in Großbritannien gezeigt, dass Bevölkerungsdichte ein Indikator für die Infektionsrate darstellt (TAYLOR et al. 2008).

Anhand der Kinderleukämiedaten des Deutschen Kinderkrebsregisters soll überprüft werden, ob

- die räumliche Verteilung einen Hinweis auf Cluster gibt (Überdispersion in der Poissonverteilung: Varianz > Erwartungswert),
- lokale, zeitliche Trends zu beobachten sind,
- Kovariablen die Anpassung des Modells verbessern.

2 Material und Methoden

Krebsdaten

Das Deutsche Kinderkrebsregister am Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI) der Johannes Gutenberg-Universität Mainz registriert seit 1980 alle Krebserkrankungen innerhalb Deutschlands. Die Vollständigkeit der Registrierung liegt insbesondere für Leukämien bei über 95 %. Alle Leukämieerkrankungen der Jahre 1987-2007 auf Gemeindeebene werden analysiert. Insgesamt beläuft sich die Anzahl auf 11.946 Fälle in 12.263 Gemeinden bzw. 426 Landkreisen getrennt nach Geschlecht und den Altersjahren null bis 14.

Leukämien sind kodiert nach der dritten Auflage der International Classification of Childhood Cancer (ICCC-3) (STELIAROVA-FOUCHER et al. 2005) (siehe Tabelle 1).

Tabelle 1: Gemeldete Fälle von Leukämie bei Kindern unter 15 Jahren in Deutschland (1987-2007) klassifiziert nach der International Classification of Childhood Cancer, dritte Version (ICCC-3)

International Classification of Childhood Cancer, dritte Version	n	%
I Leukämien	11.946	100,0
I.a Lymphatische Leukämie	9.638	80,7
I.b Akute myeloische Leukämie	1.642	13,7
I.c Chronische myeloproliferative Erkrankung	145	1,2
I.d Myelodysplastisches Syndrom und andere myeloproliferative Erkrankung	417	3,5
I.e Unspezifizierte und andere spezifizierte Leukämien	104	0,9

Methoden

Grundlage für die Modellierung von Erkrankungsraten ist das Binomialmodell. Sei o_{ij} die Anzahl Fälle und n_{ij} die Anzahl der Individuen in der Region $i=1, \dots, I$ und Altersgruppe $j=1, \dots, J$. Damit ist

$$o_{ij} \sim \text{Binomial}(n_{ij}, \pi_{ij}),$$

wobei π_{ij} das Erkrankungsrisiko ist.

Bei kleinem Erkrankungsrisiko und großer Anzahl von Individuen kann die Binomialverteilung durch die Poissonverteilung angenähert werden.

$$o_{ij} \sim \text{Binomial}(n_{ij}, \pi_{ij}) \approx \text{Poisson}(n_{ij}\pi_{ij})$$

Wenn Proportionalität unterstellt wird, kann π_{ij} folgendermaßen dargestellt werden

$$\pi_{ij} = p_j \theta_i,$$

wobei p_j die Referenzwahrscheinlichkeit in Altersgruppe j und θ_i das relative Risiko in Region i ist. Durch Einsetzen und Summation von $\sum_j n_{ij} p_j = e_i$ über alle Altersgruppen hinweg, kann die erwartete Anzahl e_i Fälle in Region i berechnet werden.

Die direkte Schätzung von θ_i durch das Maximum Likelihood Verfahren führt zum standardisierten Inzidenz Ratio (SIR). Dieses Vorgehen ist allerdings problematisch, da durch das geringe Erkrankungsrisiko die Varianz häufig größer als der Erwartungswert ist, damit den Annahmen des Poissonmodells widerspricht und zu der sogenannten Überdispersion führt. Weiterhin wird die räumliche Struktur der Daten nicht genutzt.

Die Veröffentlichung von WESTERMEIER & MICHAELIS hat folgende Teststatistik für Überdispersion verwendet:

$$Q = \frac{1}{n-1} \frac{\sum_{i=1}^n (o_i - \bar{o})^2}{\bar{o}}.$$

Da diese Teststatistik auf den beobachteten Fällen beruht und diese von der Einwohnerzahl in einer Gemeinde abhängt, wurden die beobachteten Fälle nach Gemeindegröße sortiert und 100 Gruppen gebildet. Innerhalb dieser Gruppen wurde Q dann berechnet. Bis auf die Gruppe mit den größten Gemeinden, gab es keine auffälligen Ergebnisse.

Eine weitere Methode, die bei Untersuchungen möglicher Leukämiecluster verwendet wurde, ist die Potthoff-Whittinghill Teststatistik (POTTHOFF & WHITTINGHILL [1, 2] 1966). Diese beruht zusätzlich auf den erwarteten Fällen, so dass die Unterteilung in Gemeindegruppen hinfällig wird. Eine Erweiterung der Methode ermöglicht die Einbeziehung der hierarchischen Struktur von Kreisen und Gemeinden (MUIRHEAD 2006). Damit wird die Analyse einzelner Komponenten der Varianz (beispielsweise hervorgerufen durch Gemeinden innerhalb von Kreisen) ermöglicht.

Ein Nachteil der oben beschriebenen Methoden ist, dass die Einbeziehung der räumlichen Struktur nicht möglich ist. Erst durch Formulierung eines Modells wird dies machbar.

Die Anpassung des sogenannten Besag (BESAG et al. 1991) Modells ist vorgesehen. Dieses beruht auf einem Bayes'schen hierarchischen Modell. Die logarithmierten SIR werden folgendermaßen modelliert:

$$\log(\theta_i) = \mu + u_i + v_i,$$

wobei μ das allgemeine Niveau ist, u_i die räumliche Struktur berücksichtigende Komponente und v_i eine zufällige Komponente ist. u_i kann durch ein „gaussian random field“ dargestellt werden. Letzteres ermöglicht die Einbeziehung der Nachbarschaftsstruktur innerhalb der Daten. Weiterhin kann das Modell um eine zeitliche Komponente erweitert werden.

Ein weiterer Vorteil dieser Methode ist, dass das Modell leicht durch Kovariablen ergänzt werden, die die eventuelle Variation in den Kinderleukämiedaten erklären und damit die Anpassung des Modells verbessern. Zum Modellvergleich wird das Devianz Informations Kriterium (DIC) herangezogen (SPIEGELHALTER et al. 2002).

Basis der kartographischen Darstellung sind die Verwaltungsgrenzen vom Stand 31.12.2007 (Bundesamt für Kartographie und Geodäsie).

3 Ausblick

Zur Zeit der Erstellung des Abstracts lagen noch keine Ergebnisse vor, werden aber im Rahmen des Vortrags vorgestellt.

Als Kovariablen werden zurzeit geprüft:

- Ost-West-Indikator,
- Bundesländer,
- Bevölkerungsdichte (Indikator für Infektionsrate) und
- Arbeitslosenquote auf Kreisebene.

Literatur

- ALEXANDER, F. E., BOYLE, P., CARLI, P.M., COEBERGH, J. W., DRAPER, G. J., EKBOM, A., LEVI, F., MCKINNEY, P. A., MCWHIRTER, W., MICHAELIS, J., PERIS-BONET, R., PETRIDOU, E., POMPE-KIRN, V., PLISKO, I., PUKKALA, E., RAHU, M., STORM, H., TERRACINI, B., VATTEN, L & WRAY N. (1998): Spatial clustering of childhood leukaemia: summary results from the EUROCLUS project. In: *British Journal of Cancer*, 77, S. 818-824.
- BESAG, J., YORK, J. & MOLLIE, A. (1991): Bayesian image restoration with two applications in spatial statistics. In: *Annals of the Institute of Statistical Mathematics*, 43, S. 1-59.
- DOLK, H., SHADDICK, G., WALLS, P., GRUNDY, C., THAKRAR, B., KLEINSCHMIDT, I. & ELLIOTT, P. (1997 [1]), Cancer Incidence Near Radio and Television Transmitters in Great Britain, I. Sutton Coldfield Transmitter. In: *American Journal of Epidemiology*, 145, S. 1-9.

- DOLK, H., ELLIOT, P., SHADDICK, G., WALLS, P. & THAKRAR, B. (1997 [2]): Cancer Incidence Near Radio and Television Transmitters in Great Britain, II. All High Power Transmitters. In: *American Journal of Epidemiology*, 145, S. 10-17.
- GREAVES, M. (2006): Childhood leukaemia. In: *British Medical Journal*, 324, S.283-287.
- GREAVES, M. (2006): Infection, immune responses and the aetiology of childhood leukaemia. In: *Nature Review Cancer*, 6, S.193-203.
- HJALGRIM, L. L., MADSEN, H. O., MELBYE, M., JØRGENSEN, P., CHRISTIANSEN, M., ANDERSEN, M. T., PALLISGAARD, N., HOKLAND, P., CLAUSEN, N., RYDER, L. P., SCHMIEGELOW, K. & HJALGRIM, H. (2002): Presence of clone-specific markers at birth in children with acute lymphoblastic leukaemia. In: *British Journal of Cancer*, 87, S. 994-999.
- KAATSCH, P. & SPIX, C. (2008): Jahresbericht 2006/07 des Deutschen Kinderkrebsregisters. Universität Mainz.
- KNUDSON, A. (1971): Mutation and cancer: statistical study of retinoblastoma. In: *Proceedings of the National Academy of Sciences*, 68. S. 820-823.
- LIGHTFOOT, T. J. & ROMAN, E. (2004): Causes of childhood leukaemia and lymphoma. In: *Toxicology and Applied Pharmacology*, 199. S. 104-117.
- MASKARINEC, G., COOPER, J. & SWYGERT, L. (1994), Investigation of Increased Incidence in Childhood Leukemia near Radio Towers in Hawaii: Preliminary Observations. In: *Journal of Environmental Pathology, Toxicology and Oncology*, 19, S. 33-37.
- MERZENICH, H., SCHMIEDEL, S., BENNACK, S., BRÜGGEMEYER, H., PHILIPP, J., BLETTNER, M. & SCHÜZ, J. (2008): Childhood leukemia in relation to radio frequency electromagnetic fields in the vicinity of TV and radio broadcast transmitters. In: *American Journal of Epidemiology*, 168, S. 1169-1178.
- POTTHOFF & WHITTINGHILL (1966 [1]): Testing for homogeneity. I. The binomial and multinomial distribution. In: *Biometrika*, 53. S. 167-182.
- POTTHOFF & WHITTINGHILL (1966 [2]): Testing for homogeneity. II. The Poisson distribution. In: *Biometrika*, 53. S. 183-190.
- SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. & VAN DER LINDE, A. (2002): Bayesian Measures of Model Complexity and Fit (with Discussion). In: *Journal of the Royal Statistical Society, Series B*, 64, S. 583-616.
- STELIAROVA-FOUCHER, E., STILLER, C., LACOUR, B. & KAATSCH, P. (2005): International Classification of Childhood Cancer, Third Edition. In: *Cancer*, 103, S. 1457-1467.
- TAYLOR, J. C., LAW, G. R., BOYLE, P. J., FENG, Z., GILTHORPE, M. S., PARSLow, R.C., RUDGE, G. & FELTBOWER, R. G. (2008): Does population mixing measure infectious exposure in children at the community level? In: *European Journal of Epidemiology*, 23, S. 593-600.
- WESTERMEIER, T. & MICHAELIS J. (1995): Applicability of the Poisson distribution to model the data. In: *Radiation and Environmental Biophysics*, 34, S. 7-11.