

Genetische Anfälligkeit für Krebs Eine Analyse von Halb- und Vollgeschwisterdaten

Melanie Bevier
Deutsches Krebsforschungszentrum
Im Neuenheimer Feld 580
69120 Heidelberg
m.bevier@dkfz-heidelberg.de

Marianne Weires
69120 Heidelberg

Jan Sundquist
Lund University
Malmö (Schweden)
jan.sundquist@med.lu.se

Kari Hemminki
Deutsches Krebsforschungszentrum
Im Neuenheimer Feld 580
69120 Heidelberg
k.hemminki@dkfz-heidelberg.de

Zusammenfassung

Genetische Modelle zur Bestimmung genetischer Anfälligkeit oder Vererbbarkeit von Krankheiten (zum Beispiel Krebs) wie das ACE-Modell (A=additiv genetischer Effekt, C=geteilte Umwelteinflüsse, E=verbleibende Effekte) wurden bisher meist mit Hilfe von Zwillingsdaten analysiert. Jedoch sind Zwillingsdaten sehr selten und ergeben oft nur sehr kleine Datensätze. Unser Ziel war es, stattdessen Geschwister und Halbgeschwister zu analysieren, um mit einer größeren Datenmenge Vererbbarkeit von Krebs zu schätzen. Hierzu wurde ein vorhandenes SAS-Programm von Feng (ursprünglich zur Analyse von Zwillingsdaten geschrieben) angepasst und auf Geschwisterdaten aus der Schwedischen Krebsregisterdatenbank angewendet. Das SAS-Programm basiert auf der Prozedur Nlmixed, in der fixe (zum Beispiel Alter, sozioökonomischer Status) und zufällige Effekte (A, C) spezifiziert wurden.

Schlüsselwörter: PROC NLMIXED, NLMIXED-Prozedur, genetische Anfälligkeit für Krebs, ACE-Modell

1 Einleitung

Meist werden bei der Analyse der Vererbbarkeit einer Krankheit wie beispielsweise Krebs Daten aus Zwillingsstudien untersucht. Mit deren Hilfe lässt sich der genetische Anteil der Vererbbarkeit durch die Unterschiede in monozygoten und dizygoten Zwillingen berechnen. Es ist bekannt, dass monozygote Zwillinge 100% ihrer Gene teilen, wobei dizygote Zwillinge nur etwa 50% ihrer Gene teilen. Bei einer vererbaren Krankheit wird eine höhere Korrelation in monozygoten als in dizygoten Zwillingen erwartet. Durch diesen Unterschied der Korrelationen lässt sich der genetische Anteil der Vererbbarkeit ausdrücken. Bei Zwillingen geht man von einem gleichen Anteil an geteilten Umwelteinflüssen aus, da diese in derselben Umgebung aufwachsen.

Die Heritabilität, bzw. Vererbbarkeit einer Krankheit, lässt sich als Rate der Varianz der phänotypischen Merkmale gegenüber der totalen Varianz darstellen. Zur Berechnung der Vererbbarkeit werden Varianzkomponenten-Modelle genutzt, wobei hier die sichtbare phänotypische Varianz aufgeteilt wird in additiv genetische, dominant genetische Effekte, geteilte Umwelteinflüsse und verbleibende Effekte.

Da Zwillingsdaten selten sind beziehungsweise meist kleine Datensätze zur Analyse ergeben, war die Überlegung, diese Modelle auf andere Familienstrukturen anzuwenden. Der erste Ansatz ist nun der Vergleich von Geschwistern und mütterlichen Halbgeschwistern. Der Vorteil der Analyse von mütterlichen Halbgeschwistern ist, dass diese in der gleichen Umgebung (mit der Mutter) aufwachsen, weshalb man wie bei Zwillingen die gleichen geteilten Umwelteinflüsse annehmen kann.

2 Genetische Modelle und Zwillingsstudien

Ein genetisches Modell für quantitative Zielvariablen besteht aus fixen Effekten, additiv genetischen Effekten (a), dominant genetischen Effekten (d), geteilten Umwelteinflüssen (c) und verbleibenden Effekten (ε):

$$y_{ij} = X_{ij}\beta + a_{ij} + d_{ij} + c_i + \varepsilon_{ij},$$

wobei i für jedes beobachtete Paar (zum Beispiel: pro Zwillingspaar) steht und j den Index für das Individuum in einem Paar angibt ($j=1,2$).

Meist werden zur Analyse der Vererbbarkeit eines Merkmales Zwillingsstudien genutzt. Es ist bekannt, dass monozygote Zwillinge genetisch identisch sind, was bedeutet, sie teilen sich 100% ihrer Gene. Daher gilt: $a_{i1} = a_{i2}$ und $d_{i1} = d_{i2}$.

Es wird angenommen, dass a, d, c, ε normalverteilt mit Mittelwert 0 und entsprechender Varianz $\sigma_a^2, \sigma_d^2, \sigma_c^2, \sigma_\varepsilon^2$ sind.

Falls das i -te Zwillingspaar monozygot ist, ist die Kovarianz von a und d zwischen den beiden Zwillingen: $\text{cov}(a_1, a_2) = \sigma_a^2$ und $\text{cov}(d_1, d_2) = \sigma_d^2$. Bei einem dizygoten Zwillingspaar ist folgende Kovarianzstruktur zwischen beiden Zwillingen gegeben: $\text{cov}(a_1, a_2) = 1/2\sigma_a^2$ und $\text{cov}(d_1, d_2) = 1/4\sigma_d^2$.

Zur Analyse der Erblichkeit von binären Merkmalen (zum Beispiel einer Krankheit: erkrankt, nicht erkrankt) lässt sich ein probit-Modell verwenden:

$$\text{probit}(\text{Pr}(y_{ij} = 1)) = X_{ij}\beta + a_{ij} + d_{ij} + c_i + \varepsilon_{ij},$$

wobei die $\text{probit}(\cdot)$ die Inverse der kumulierten Standardnormalverteilung ist.

3 Heritabilität

Die Heritabilität eines Merkmals lässt sich mit Hilfe von Varianzkomponenten-Modelle berechnen. Hierbei wird sie als Rate der phänotypischen Varianz gegenüber der totalen Varianz dargestellt.

Die phänotypische Varianz ist die Varianz der additiv- und dominant-genetischen Effekte. $\text{Var}(P) = \text{Var}(A) + \text{Var}(D)$

Die totale Varianz setzt sich zusammen aus der phänotypischen Varianz und der Varianz der geteilten und nicht-geteilten Umwelteinflüsse.

$$\begin{aligned}\text{Var (total)} &= \text{Var (P)} && + \text{Var (C)} + \text{Var (E)} \\ &= \text{Var (A)} + \text{Var (D)} && + \text{Var (C)} + \text{Var (E)}\end{aligned}$$

Heritabilität:
$$h^2 = \frac{\text{Var}(A) + \text{Var}(D)}{\text{Var}(A) + \text{Var}(D) + \text{Var}(C) + \text{Var}(E)}$$

Da es schwierig ist, die einzelnen Komponenten zu unterscheiden, ist es besser, einfachere Modelle zu betrachten, beispielsweise:

ein ADE-Modell:
$$h^2 = \frac{\text{Var}(A) + \text{Var}(D)}{\text{Var}(A) + \text{Var}(D) + \text{Var}(E)}$$

oder ein ACE-Modell:
$$h^2 = \frac{\text{Var}(A)}{\text{Var}(A) + \text{Var}(C) + \text{Var}(E)}$$

4 Daten

Die zu analysierenden Daten stammen aus der Schwedischen Familien-Krebsregister-Datenbank, in der Personen, die seit 1932 geboren wurden, mit ihren biologischen Eltern registriert sind. Für die über 11 Millionen registrierten Personen sind Informationen über Krebserkrankungen zwischen 1961 und 2006 verfügbar. Neben den klinischen Parametern zur Krebserkrankung sowie Todesursache sind auch Informationen über die Schulbildung, den sozioökonomischen Status, die Region, in der die Person lebt, sowie eventuelle Migrationsinformation vorhanden.

Durch die Verknüpfung der Person mit ihren Eltern ist es möglich, diverse Familienstrukturen zu identifizieren. Geschwister, bzw. Halbgeschwister können durch die Eltern-Personen-Tripel aufgezeigt werden.

5 Analysis of Twin Data Using SAS - Rui Feng et al.

Rui Feng et al. zeigen in ihrem Artikel "Analysis of Twin Data Using SAS" [1], wie es in SAS möglich ist, Zwillingsdaten zu analysieren. Für die Analyse der Heritabilität gibt es spezielle Software, die Zwillingsstudien auswerten können (zum Beispiel: LISREL, Mx Mplus). Da aber die weitverbreitetste Software zur statistischen Analyse SAS ist, haben sie ein auf der Prozedur Nlmixed basierendes SAS-Programm geschrieben, mit dem es auch möglich ist, diese Daten auszuwerten. Sie haben durch Simulationen und die Anwendung auf reale Daten ihre Ergebnisse mit denen von Mx verglichen und so ihr Programm validiert. Rui Feng et al. haben sowohl ein Programm für ein ADE- als auch für ein ACE-Modell angegeben. Zur Analyse der genetischen Anfälligkeit für RDS ("Respiratory Distress Syndrome") wurde später beispielsweise ein ACE-Modell verwendet [2].

6 Anwendung des SAS-Programms auf Halb- bzw. Vollgeschwisterdaten

Da Zwillingsdaten selten sind und meist nur kleine Datensätze ergeben, war die Überlegung, die genetische Anfälligkeit einer Krankheit (hier am Beispiel von Brustkrebs) mit Hilfe von Voll- bzw. Halbgeschwisterdaten zu berechnen.

Der vorliegende Datensatz wurde auf die Frauen reduziert, die entweder eine Halbschwester mütterlicherseits (und keine anderen Geschwister) oder eine Schwester (und keine anderen Halbgeschwister) hatten. Außerdem wurde die Gruppe der Schwestern zur Gruppe der Halbschwestern bezüglich der Altersstruktur gematched. Die Analyse wurde für den sozioökonomischen Status der Personen adjustiert.

Zur Berechnung der Heritabilität wurde das ACE-Modell verwendet. Da mütterliche Halbgeschwister und Schwestern analysiert wurden, kann man von gleichen geteilten Umwelteinflüssen innerhalb der Paare ausgehen (mütterliche Halbschwestern leben meist weiterhin mit der Mutter zusammen).

7 SAS-Code

```
/* Datensatz breast2 enthält alle relevanten Informationen: Person,
Halb- bzw. Schwester, Alter, sozioökonomischer Status,
Krankheitsstatus, Familien-ID */
/* Dieser erste Teil dient zur Bildung der Kovarianz-Strukturen */

proc iml;
covA1=sqrsym({1,0.5,1});
gA1=t(root(covA1));          /* Schwestern */
covA2=sqrsym({1,0.25,1});
gA2=t(root(covA2));          /* Halbschwestern*/
use breast2;
  read all var {fam hsm} into fam4;
close breast2;
nsub=nrow(fam4);
g=j(nsub,2,.);
do i=1 to nsub;
  ind=2-mod(i,2);
  if(fam4[i,2]=1) then g[i,]=gA1[ind,];    /* Schwestern */
  if(fam4[i,2]=0) then g[i,]=gA2[ind,];    /* Halbschwestern*/
end;
cname={"aone" "atwo" };
create rmatrix from g[colname=cname];
append from g;
close rmatrix;
quit;

data rmatrix;
set rmatrix;
id=_n_;
run;
```

```
data two;
merge breast2 rmatrix;
by id;
run;

/* Programm basierend auf der Prozedur Nlmixed */

title 'ACE model';
proc nlmixed data=two tech=congra;
parms beta1=-0.0360 beta2=-0.4501 beta3=-0.4601 beta4=-0.7199
beta5=1.3575;
fixed1 = beta0 + beta1*age + beta2*sei3 + beta3*sei4 + beta4*sei5 +
beta5*sei6;
random1 = a1*aone + a2*atwo + c;
eta = fixed1 + random1;
p = 1 - probnorm(eta);
model code ~ binary(p);
random a1 a2 c ~ normal([0,0,0],[s1,0,s1,0,0,s2]) subject=fam;
estimate 'icc' s1/(1+s1+s2);
run;
```

8 Ergebnisse

Mit Hilfe eines ACE-Modells konnte die genetische Vererbbarkeit von Brustkrebs geschätzt werden. Danach sind 31,88% der Varianz der Anfälligkeit für Brustkrebs allein genetischen Faktoren zuzuschreiben. In einem Artikel von Locatelli et al. [3] wird die Vererbbarkeit von Brustkrebs durch Analyse von Zwillingsdaten auf ca. 30 % geschätzt, was unser Ergebnis bestätigt.

9 Schlussfolgerung

Mit Hilfe des vorgestellten SAS-Programms lässt sich die genetische Anfälligkeit einer Krankheit nicht nur durch die Analyse von Zwillingsdaten berechnen, sondern auch anhand anderer Familienstrukturen. Außerdem kann man je nach Ziel der Analyse zwischen einem ADE- und einem ACE-Modell wählen.

Literatur

- [1] Rui Feng, Gongfu Zhou, Meizhuo Zhang, and Heping Zhang. Analysis of twin data using SAS. *Biometrics*, 65(2):584–589, Jun 2009.
- [2] Orly Levit, Yuan Jiang, Matthew J Bizzarro, Naveed Hussain, Catalin S Buhim-schi, Jeffrey R Gruen, Heping Zhang, and Vineet Bhandari. The genetic susceptibility to respiratory distress syndrome. *Pediatr Res*, 66(6):693–697, Dec 2009.
- [3] Isabella Locatelli, Paul Lichtenstein, and Anatoli I Yashin. The heritability of breast cancer: a Bayesian correlated frailty model applied to Swedish twins data. *Twin Res*, 7(2):182–191, Apr 2004.