

Methodische Aspekte einer ökologischen Analyse von Daten des Krebsregisters Rheinland-Pfalz zum Zusammenhang zwischen Weinbaufläche und Krebsinzidenz

Jochem König, Gaël P. Hammer und Andreas
Seidler

*Institut für Medizinische Biometrie,
Epidemiologie und Informatik,
Johannes Gutenberg-Universität Mainz*

1

Ziel

- Gibt es Hinweise auf Zusammenhang zwischen Weinbaufläche und Krebsinzidenzen?

2

Studiendesign

- Ökologische Studie
- Datenbestand 2000-2003
- Weinbau-Gemeinden (> 0% Rebfläche, <100.000 Ew):
Σ 660.000 Ew.
- Risikofaktor (Pestizide)
 - Anteil Rebfläche an der Gemeindefläche
(<0-5% (Ref.), 5-20%, >= 20%)
- Stratifikation: Geschlecht, Diagnose (nach ICD-10)
- Counfounder
 - Alter („stetig“)
 - Stadt/Land (binär, > 5000 Einwohner)
 - Obstanbau (binär, >5% der Gemeindefläche)
- Auswertung: Poisson-Regression

3

Die Daten

- 498 Gemeinden mit Weinbau werden in drei Klassen geteilt:
 - Rebflächenanteil 0-5%, 5%-20%, >20%
- Von jeder Gemeinde in 2x18 Geschlechts- und Altersklassen
 - Bevölkerung und
 - inzidente Fälle zu 85 Krebsarten aus 2000-2003

4

Krebsregister Rheinland-Pfalz
Erfassungsgrad (2002)

Männer

Frauen

[image removed]

5

Krebsregister Rheinland-Pfalz
Erfassungsgrad (2002)

[image removed]

6

Krebsregister Rheinland-Pfalz Erfassungsgrad (2002)

- Erfassungsgrad von $\geq 90\%$ für folgende Krebslokalisationen:
 - Brustkrebs bei Frauen (C50)
 - Prostatakrebs (C61)
 - Malignes Melanom (C43)
 - Nicht-melanotische Hauttumoren (C44)
 - Hodenkarzinome (C62)
 - Harnblasenkarzinome bei Männern (C67)

7

Weinbaugebiete Rheinland-Pfalz

[image removed]

8

Weinbaugebiete Demographische Kennzahlen

	Rhineland- Palatinate*	Area under wine (% of community area)		
		Total	> 0%, ≤ 5%	>5%, ≤ 20%
Communities	2,305	162	171	170
Total area (ha)	1,984,688	222,736	200,709	129,444
Area under wine (ha)	69,043	2,996	25,101	40,683
% area under wine	3.5%	1.3%	12.5%	31.4%
Inhabitants (per ha)	4,000,567 (2.02)	564,210 (2.53)	526,486 (2.62)	301,193 (2.33)
Inhabitants per community (median, min-max)	566 (6-184,752)	1,188 (72-99,750)	1,193 (95-80,535)	984 (84-40,110)

9

Besonderheiten

- Nur 80% durchschnittliche geschätzte Vollzähligkeit.
- Durch den Ausschluss von Gemeinden ohne Weinbau werden nur benachbarte Gemeinden einbezogen.
- Zufällige Fehler werden mit Poissonregression wg. Overdispersion unterschätzt.

10

Limitationen

- Die ecologic fallacy
- Artefakte wegen differentieller Melderaten
- werden heute nicht weiter diskutiert.

11

Optionen

- Poisson-Regression
- Poisson-Regression mit Overdispersion-Parameter via
 - Pearson Residuen oder
 - Deviance-Residuen
- Negativ-Binomial Modell
- Zufallseffekt-Modell
 - Poisson-Gamma oder
 - Poisson-Normal

12

Modelle

$$O_i \sim \text{Poisson}(\exp(X_i' \beta)), \quad V(O_i) = E_i$$

$$O_i \sim \text{Poisson}(\exp(X_i' \beta)), \quad V(O_i) = \phi E_i$$

$$O_i \sim \text{NegBin}(\exp(X_i' \beta), k), \quad V(O_i) = E_i + kE_i^2$$

$$O_i \sim \text{Poisson}(\exp(X_i' \beta + \sigma Z_i)), Z_i \sim N(0,1) \quad V(O_i) = E_i + kE_i^2$$

$$O_i \sim Z_i \text{Poisson}(\exp(X_i' \beta)), Z_i \sim \Gamma(k^{-1}, k^{-1}) \quad V(O_i) = E_i + kE_i^2$$

(3) und (5) haben dieselbe Marginalverteilung.

13

Beispiel C71 Hirn

- $N=109+70+45=224$ (Weinbau <5, >5, >20%)
- Alle Modelle mit den Faktoren
 - Stadt (>5000 Einwohner)
 - Obstbau (>5% der Fläche)
 - Alter (linear)
 - Weinbau (3 Stufen)

14

Obstanbau, C71 Hirn

Modell	RR	95%-KI
Poisson (Bayes)	1.78394 1.776	1.31374 - 2.42242 1.3 - 2.384
Poisson GEE(Independent)	1.78394	1.17054 - 2.71877
Poisson GEE(exchangeable)	1.78660	1.17352 - 2.71997
Neg.Binomial (k=0.12)	1.69667	1.21031 - 2.37848
Neg.Binomial GEE (Indep.)	1.69667	1.08283 - 2.65850
GLMM Poisson-Gauss Bayes	1.22	0.6884 - 2.048
GLMM Poisson-Gamma Bayes prior k~Gamma(.001,.001)	1.233	0.7215 - 2.011
GLMM Poisson-Gamma Bayes prior k~unif(0,100)	1.196	0.6871 - 1.988
GLMM Poisson-Gamma Bayes prior k=0.12 fest	1.383	0.9108 - 2.067
GLMM Poisson-Gauss-CAR	1.188	0.6577 - 1.996

15

Weinbaufläche >5% vs <5%

Modell	RR	95%-KI
Poisson (Bayes)	0.98950 0.9842	0.72722 - 1.34638 0.7197 - 1.338
Poisson GEE(Independent)	0.98950	0.68308 - 1.43337
Poisson GEE(exchangeable)	0.98847	0.68286 - 1.43084
Neg.Binomial (k=0.12)	1.02226	0.74015 - 1.41190
Neg.Binomial GEE (Indep.)	1.02226	0.70014 - 1.49259
GLMM Poisson-Gauss Bayes	1.394	0.9133 - 2.196
GLMM Poisson-Gamma Bayes prior k~Gamma(.001,.001)	1.429	0.9214 - 2.295
GLMM Poisson-Gamma Bayes prior k~unif(0,100)	1.49	0.9563 - 2.364
GLMM Poisson-Gamma Bayes prior k=0.12 fest	1.228	0.847 - 1.764
GLMM Poisson-Gauss-CAR	1.1	0.7446 - 1.66

16

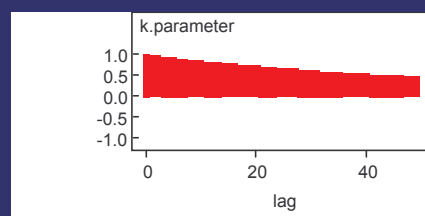
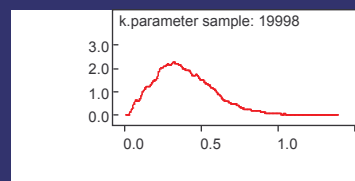
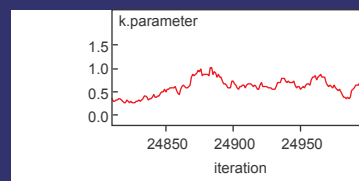
Weinbaufläche >20% vs <5%

Modell	RR	95%-KI
Poisson (Bayes)	1.04047 1.032	0.71040-1.52389 0.6983 - 1.515
Poisson GEE(Independent)	1.04047	0.66681-1.62351
Poisson GEE(exchangeable)	1.04215	0.66782-1.62631
Neg.Binomial (k=0.12)	1.06487	0.72162-1.57138
Neg.Binomial GEE (Indep.)	1.06487	0.68278-1.66076
GLMM Poisson-Gauss Bayes	1.365	0.8526 - 2.229
GLMM Poisson-Gamma Bayes prior k~Gamma(.001,.001)	1.399	0.8667 - 2.296
GLMM Poisson-Gamma Bayes prior k~unif(0,100)	1.448	0.9005 -2.377
GLMM Poisson-Gamma Bayes prior k=0.12 fest	1.224	0.798 -1.852
GLMM Poisson-Gauss-CAR	1.202	0.7396 - 1.995

17

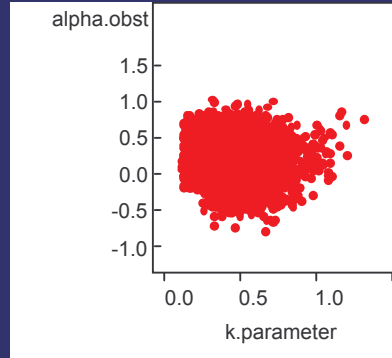
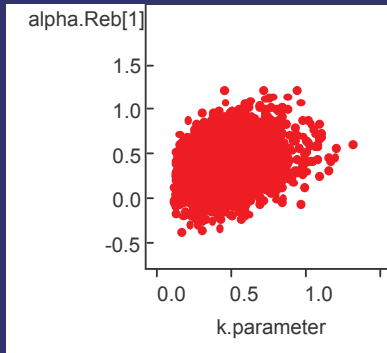
Problem mit Dispersionsparameter

k (in „ $\text{Var}=E+kE^2$ “)



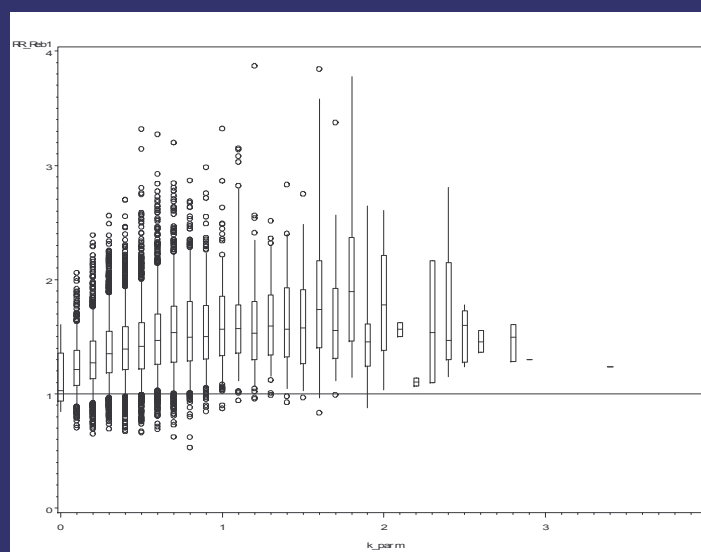
18

Dispersionsparameter k ($\text{Var} = E + kE^2$)



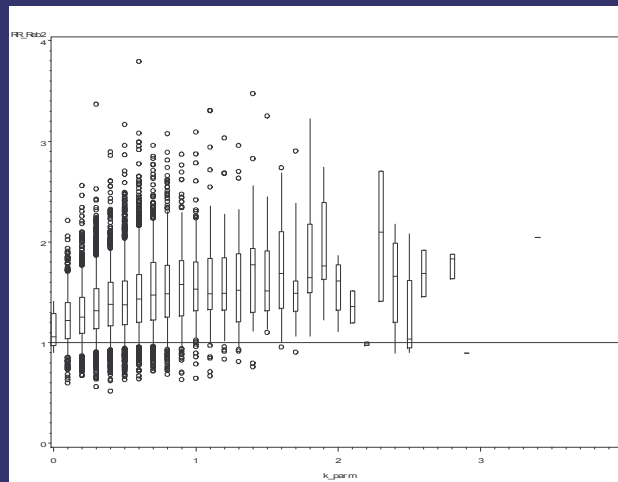
19

RR (Rebfläche 2:1) 98% Bereiche in Abhängigkeit vom Dispersionsparameter k ($\text{Var} = E + kE^2$)



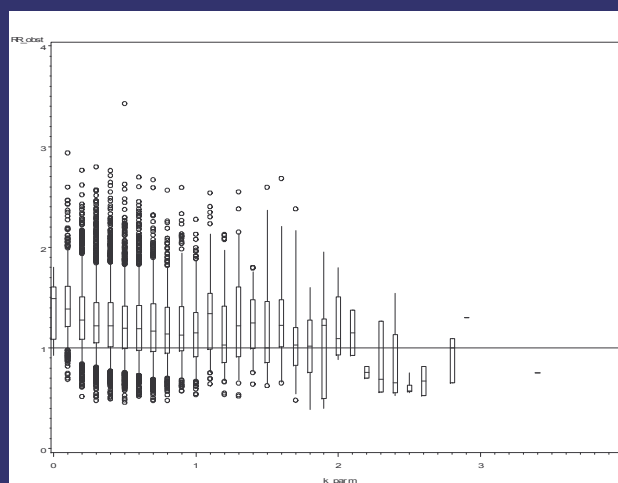
20

RR (Rebfläche 3:1) 98% Bereiche in Abhängigkeit vom Dispersionsparameter k ($\text{Var}=E+kE^2$)



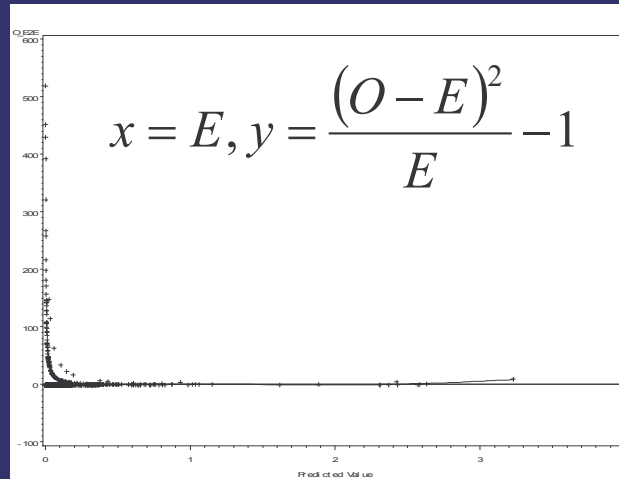
21

RR (Obstanbau) 98% Bereiche in Abhängigkeit vom Dispersionsparameter k ($\text{Var}=E+kE^2$)



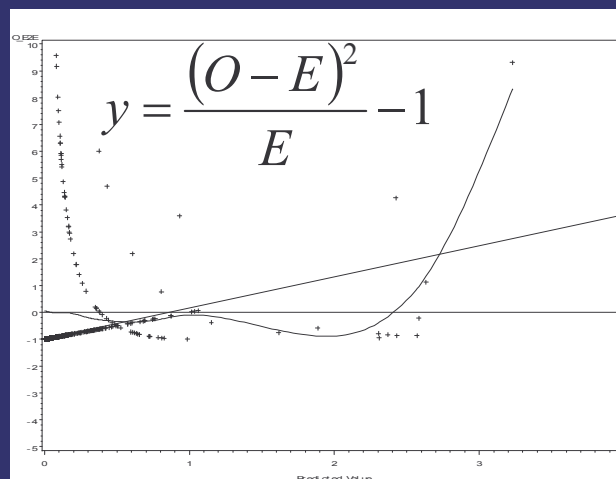
22

Plot der quadrierten Residuen des Poisson-Fits



23

Plot der quadrierten Residuen des Poisson-Fits (Ausschnitt)



24

Kleine E's machen unsichere Reste

- Regression von $\frac{(O-E)^2}{E} - 1$ auf E

	Parameter	Standard		
Variable	Estimate	Error	t Value	Pr > t
Intercept	0.029	0.148	0.19	0.847
E	0.114	1.381	0.08	0.934

- Die Overdispersion-Korrektur mit Pearson-Residuen wird extrem instabil.

- Mit Gewicht $\frac{E}{2E+1} = V^{-1} \left(\frac{(O-E)^2}{E} \right)$

	Parameter	Standard		
Variable	Estimate	Error	t Value	Pr> t
• Intercept	-0.072	0.088	-0.81	0.416
• E	0.333	0.187	1.78	0.076

25

Zusammenhang zu Poisson und Negativ-Binomialmodell

- Regression nur mit Intercept ergibt Dispersionsparameter (minus) im overdispersed Poissonmodell 0.031 ungewichtet, -0.08 gewichtet),
- Regression nur mit Slope ergibt Parameter k aus Negativ-Binomialmodell (0.18 ungewichtet, 0.27 gewichtet),

26

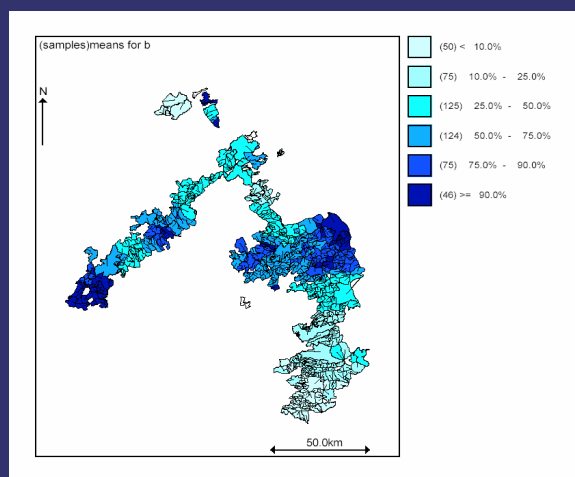
Beobachtungen

ohne Anspruch auf Allgemeinheit

- Overdispersion ist schwer zu schätzen, wenn die Counts klein sind.
- Marginale Modelle (GEE) unterschätzen den Dispersionsparameter
- Die Bayes-Analyse im Gemischten Modell bewahrt vor voreiligen Anschuldigungen und zeigt, dass die Datenlage zu dünn ist.

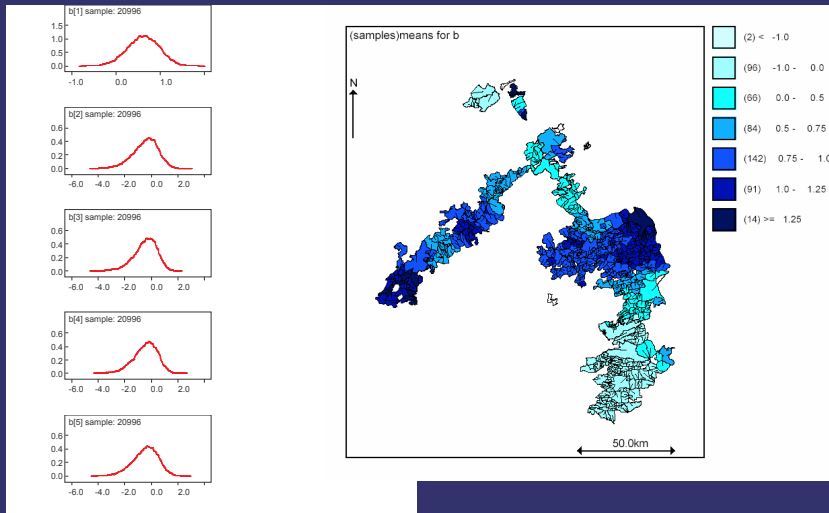
27

Räumlich korrelierter Zufallseffekt



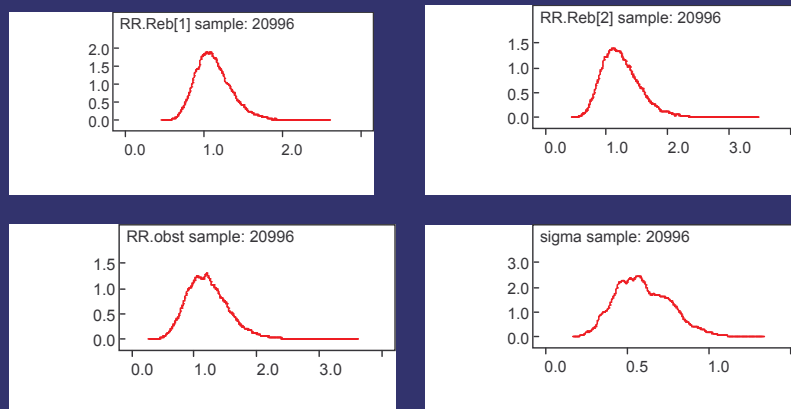
28

Räumlich korrelierter Zufallseffekt



29

CAR-Modell



30