



Roosebeckstr. 4-6  
30449 Hannover  
Fon 0511/4505-0  
Fax 0511/4505-140

## Statistische Erkennung räumlicher Cluster- Erfahrungen aus der Infektionsepidemiologie

J. Dreesman  
Niedersächsisches Landesgesundheitsamt

Johannes.Dreesman@NLGA.Niedersachsen.de

## Cluster - Häufungen

- Definition: Eine gegenüber der Erwartung erhöhte Anzahl von Fällen in einer Region und/oder einem Zeitintervall
- Gründe um nach Clustern zu suchen:
  - Um die Krankheit zu erforschen
  - Überprüfung einer Cluster-Meldung
  - Infektionssurveillance: Ausbruch identifizieren, Ursache ermitteln und abstellen

## Systematik von Clustertests

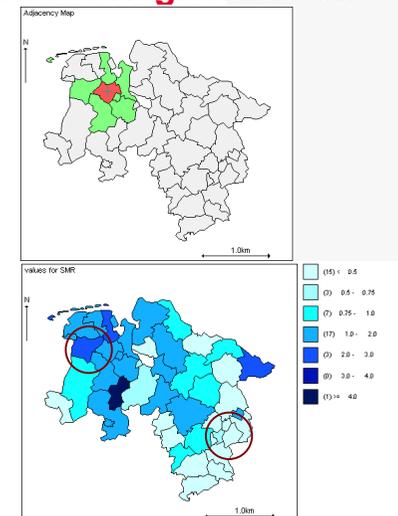
- Global
  - Gibt es eine generelle Tendenz zur Clusterung
- Lokal, nicht fokussiert
  - Gibt es irgendwo ein Cluster
- Lokal, fokussiert
  - Gibt es ein Cluster an einer bestimmten Stelle, z.B. an einer bestimmten Expositionsquelle

## Beispiel 1: Moran's I zur Messung der räumlichen Autokorrelation bei Regionaldaten

- „Auto-“ = die Daten korrelieren mit sich selber (d.h. den Werten in der Nachbarschaft)
- Messung durch Moran's I:

$$I = \frac{N \cdot \sum_{i=1}^N \sum_{j=1}^N w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\left( \sum_{i=1}^N (z_i - \bar{z})^2 \right) \cdot \left( \sum_{i=1}^N \sum_{j \neq i}^N w_{ij} \right)}$$

- Entspricht dem klassischen Korrelationskoeffizienten
- Wertebereich ist allerdings nicht auf (-1, 1) beschränkt
- Systematik: Global, Regionaldaten, Risikomaß

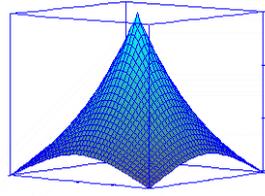


## Beispiel 2: Fokussierter Clustertest von Bithell für Regionaldaten (Idee des Tests)

- Zuerst wird eine Relative Risiko Funktion (RRF) für die Umgebung der Quelle geschätzt (siehe unten)
- Die Anpassung der Daten durch ein Modell mit RRF und ohne RRF für die Umgebung der Quelle wird durch Likelihood-Ratio-Test verglichen

$$f(d) = 1 + \beta \exp(-d / \varphi)$$

This model comes to a sharp point at the origin (focus): risk increases more rapidly the closer the subject is to the focus.



- Systematik: Lokal fokussiert, Regionaldaten, Risikomaß

## Public-Health-Surveillance

- Z.B. in der Infektionssurveillance: Wöchentliche Suche nach auffälligen Clustern, die auf Infektionsausbrüche hindeuten könnten
- Problem: Artefakte durch Laboreffekte, Zeitverzögerungen etc.
- Problem: Multiples Testen: Anzahl der Tests \* p-Wert = Anzahl der Fehlalarme  
=> Tests haben nur Hypothesen-generierenden Charakter

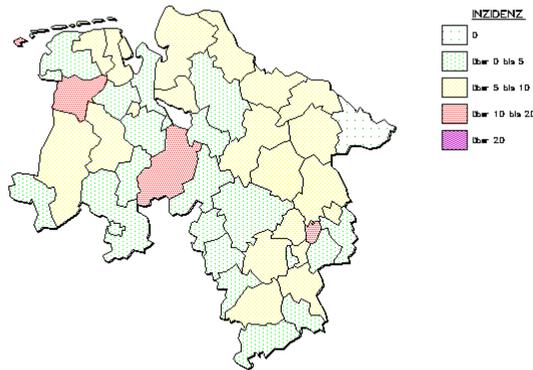
## Überregionale Infektionssurveillance

- Datenbasis: Meldedaten nach Infektionsschutzgesetz (IfSG)
- Gesundheitsämter übermitteln wöchentlich Daten an Landesstelle, Landesstelle übermittelt weiter an RKI
- Ziel u.a.: Identifikation von Ausbrüchen und Beseitigung der Ursache bzw. Unterbrechung der Übertragung
- Lokale Ausbrüche werden vom Gesundheitsamt i.d.R. ohne quantitative Methoden erkannt, aber nicht immer mitgeteilt
- Überregionale Ausbrüche werden oft lokal nicht wahrgenommen
- Zur routinemäßigen Analyse auf überregionaler Ebene sind quantitative/statistische Methoden ein wichtiges Hilfsmittel

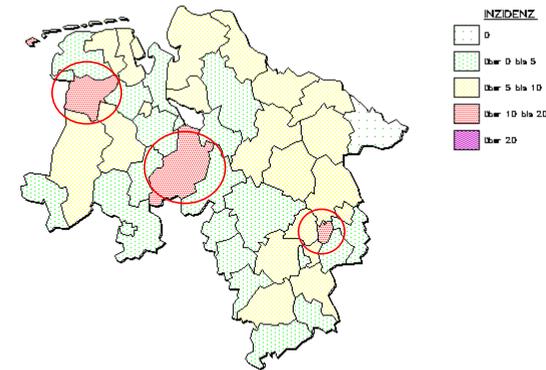
## Routineauswertungen in Niedersachsen

- Wöchentliche systematische Suche nach Cluster-Alarmen
  - MMWR-Methode für 16 häufige Krankheiten (4 Wochen Daten)
  - Farrington-Methode für 16 häufige Krankheiten mit R-package "Surveillance" (Höhle et al.)
  - MMWR-Methode für 2000 Salmonella-Serotypen (4 Wochen Daten)
  - Spatial Scan (SaTScan) für 16 häufige Krankheiten (4 Wochen Daten)
- Keine systematische Bestätigungs-Prozedur für Cluster-Alarme
- Typische Prozedur
  - Visuelle Untersuchung (Zeitreihe, Karte)
  - Vergleich interessanter Alarme mit elektronischen Ausbruchsberichten
  - Wenn keine Ausbruchsbericht vorliegt => Ggf. Kontakt GA
  - Keine Ressourcen zur systematischen Validierung der Alarme

## Inzidenzkarten, z.B. Salmonellosen 19.-22. KW (Fälle pro 100 000 E. je Landkreis / kreisfreie Stadt)



## Inzidenzkarten, z.B. Salmonellosen 19.-22. KW (Fälle pro 100 000 E. je Landkreis / kreisfreie Stadt)

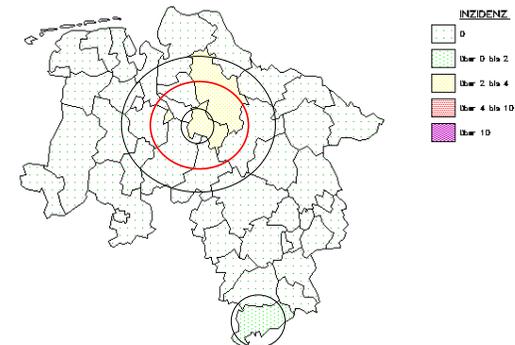


## Räumlicher Clustertest mit Scan-Statistik und SaTScan-Software

Region	Beobachtet	Erwartet	P-Wert	Bemerkung
BS	41	14	<0.001	Ausbruch durch kontaminiertes Speiseeis
DH	33	11	<0.001	Vermutlich ebenfalls durch kontaminiertes Speiseeis
LER	24	9	0.003	Keine Gemeinsamkeit erkennbar

## Identifikation überregionaler Cluster (Z.B. Masern 19.-22. KW)

Region	Beobachtet	Erwartet	P-Wert
VER, ROW, DEL	12	1	<0.001



## Scan-Statistik

- Test basiert auf Kreisen, die bei wachsender Größe eine oder mehrere Regionen beinhalten
- Inzidenzrate innerhalb der Kreise wird mit Inzidenzrate außerhalb verglichen
- Ist das Modell mit verschiedenen Inzidenzraten innerhalb und außerhalb im Likelihood-Ratio-Test signifikant besser als das Modell mit homogener Rate, dann erfolgt Alarm
- Es können auch mehrere Alarme für eine Karte erfolgen
- Systematik: Lokal (nicht fokussiert), Regionaldaten, Fallzahl und Bevölkerung

Quelle: Kulldorf and Nagrawalla, Statistics in Medicine, 1995, S. 799-810

## Die räumliche Scan-Statistik (Kulldorf and Nagrawalla 1995)

- $Z = \{\text{Kreise } Z \text{ mit unterschiedlicher Lokation und Größe}\}$
- Testproblem  $H_0 : \lambda_Z = \lambda_{\bar{Z}} \quad H_1 : \lambda_Z > \lambda_{\bar{Z}}$
- Scan-Statistik ist der Maximum Likelihood Quotient über alle Kreise  $Z$ :

$$S_Z = \frac{\sup_{Z \in \mathbf{Z}, \lambda_Z > \lambda_{\bar{Z}}} L(Z, \lambda_Z, \lambda_{\bar{Z}})}{\sup_{Z \in \mathbf{Z}, \lambda_Z = \lambda_{\bar{Z}}} L(Z, \lambda_Z, \lambda_{\bar{Z}})}$$

- Die Likelihood basiert auf der Poisson-Verteilung

## Die räumliche Scan-Statistik und die Realisierung in SaTScan

- Verfügbare Freeware: SaTScan  
<http://srab.cancer.gov/satscan/>
- Simulation von 999 Konstellationen unter  $H_0$   
(Gleichmäßige Verteilung der  $n$  Fälle über die Regionen gemäß der Bevölkerung)
- Beobachteter Wert der Statistik wird eingeordnet: Monte Carlo p-Wert
- Durch Simulation wird multiples Testproblem umgangen
- Auch sekundäre, nicht überlappende Cluster werden ausgewiesen (konservativ, da stärkeres Cluster in  $\bar{Z}$ )
- Auch möglich für Zeit bzw. Raum und Zeit

(Quelle: Kulldorf and Nagrawalla, Statistics in Medicine, 1995, S. 799-810)

## Einstellungen von SaTScan für die niedersächsische Surveillance

- Der Landkreis wird durch den geografischen Schwerpunkt der Region repräsentiert
- Bei zweigeteilten Regionen / Inseln wird nur der größere Teil verwendet
- Der Scan erfolgt immer über die Fallzahlen der letzten 4 Wochen
- $P = 0,05$
- Anzahl Simulationen = 999
- Maximale Clustergröße 20%

(Dreesman und Scharlach: Gesundheitswesen 2004; 66:783-9)

## SaTScan: Ergebnis-Datei

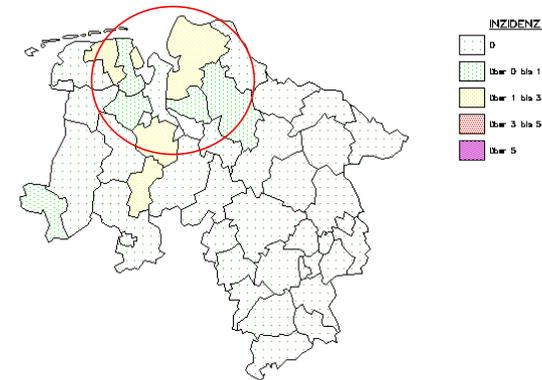
### SUMMARY OF DATA

Study period .....: 2002/1/1 - 2002/12/31  
 Number of census areas: 46  
 Total population .....: 7851907  
 Total cases .....: 13  
 Annual cases / 100000.: 0.2

### MOST LIKELY CLUSTER

1. Census areas included.: VER, ROW, DEL  
 Coordinates / radius..: (2339,6352) / 352.93  
 Population.....: 365273  
 Number of cases.....: 12 (0.60 expected)  
 Annual cases / 100000.: 3.3  
 Overall relative risk.: 19.842  
 Log likelihood ratio..: 33.336576  
 Monte Carlo rank.....: 1/1000  
 P-value.....: 0.001

## Inzidenzkarte Kryptosporidiose 42.-45. KW 2003 (gemeldete Fälle pro 100 000 E. je Landkreis / kreisfreie Stadt)



- Kryptosporidien sind ein Darmparasit, der Durchfallerkrankungen auslöst
- Zu den natürlichen Wirten gehören Rinder
- Fall Kontroll-Studie zum Laboreffekt mit 146 Kryptosporidiose-Fällen und 146 gematchte Salmonella and Campylobacter-Fällen als Kontrollen

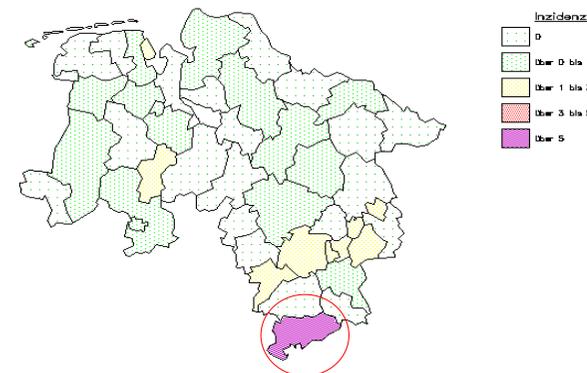
## Ergebnisse der Fall-Kontroll-Studie zum Labor

Labor		CUX		BRA		ROW		FRI		VEC		Gesamt	
		Fälle	Kon.	Fälle	Kon.								
A	n	42	21	39	21	16	1	24	2	10	0	130	45
	%	95,5	47,7	92,9	50,0	72,7	4,5	96,0	8,0	76,9	0,0	89,0	30,8
B	n	1	6	2	7	6	1	0	2	3	0	13	16
	%	2,3	13,6	4,8	16,7	27,3	4,5	4,0	8,0	23,1	0,0	8,9	11,0
Andere	n	1	17	1	14	0	20	1	21	0	13	3	85
	%	2,3	38,6	2,4	33,3	0,0	91,0	4,0	84,0	0,0	100	2,1	58,2
Gesamt		44	44	42	42	22	22	25	25	13	13	146	146

- 89% der K.-Fälle wurden in einem örtlichen Labor (A) nachgewiesen, aber nur 31% der Kontrollen. (mOR für Lab. A: 44 (95%-KI: 11 – 177; p<0,001).
- Labor A verwendete eine breitere Indikation für die Durchführung der K.-Diagnostik

(Dreesman et al: Gesundheitswesen 2007; 69:483-7)

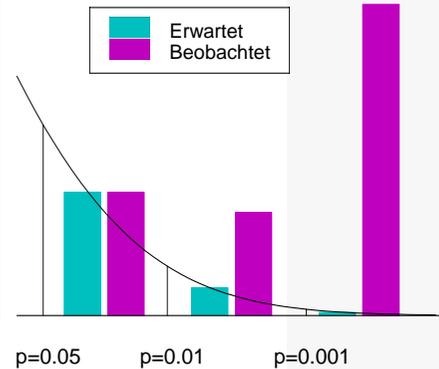
## Inzidenzkarte Tuberkulose 49.-52. KW 2004 (Fälle pro 100 000 E. je Landkreis)



- Im Landkreis Göttingen befindet sich das Grenzdurchgangslager Friedland für Spätaussiedler
- Bei der Ankunft werden die Spätaussiedler durch ein Röntgenbild der Lunge auf Tuberkulose gescreent (Feil, Dreesman: Eurosurveillance weekly 2004; 8(18))

## p-Werte der SaTScan-Alarme 2003 – 2005 (2652 Tests => 123 Alarme erwartet unter H0)

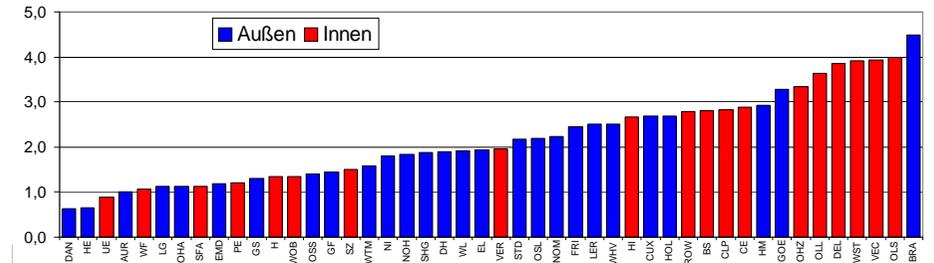
P-Wert	Anzahl Cluster	%
0,01 – <0,05	434	24
0,001 – <0,01	375	20
0 - <0,001	1031	56
Gesamt	1840	



## Verhalten am Rand



Beteiligung an Clustern in %



## EpiScanGIS als Beispiel für Routine-Verwendung und Verknüpfung mit Internet-Darstellung

- Nationales Referenzlabor für Meningokokken-Erkrankungen
- Aufgabe: Feintypisierung von Meningokokken-Stämmen
- Dynamische Karten im Internet
- Clustersuche mit SaTScan™
- Für Nutzer aus dem ÖGD werden Cluster auf der Internetseite dargestellt (Frühwarnsystem)
- [www.meningococcus.de](http://www.meningococcus.de)

## Eigenschaften von SaTScan

- Einfach anwendbar, geeignet für Anwendung im Routinebetrieb
- Identifikation überregionaler Cluster:
  - Als Ort der Erkrankungen wird der Repräsentant (Centroid) der Region angenommen
  - Als Nachbarschaftskriterium wird der euklidische Abstand der Regions-Repräsentanten verwendet
  - Verhalten am Rand des Untersuchungsgebietes
- Erweiterung für ellipsenförmige Cluster ist möglich
- Erweiterung auf raum-zeitliche Cluster ist möglich
- Für Einzelregionen ist der Test vergleichbar mit einer Perzentilenprüfung anhand der Poissonverteilung