



**DIE ANWENDUNG RÄUMLICH-STATISTISCHER ANALYSEMETHODEN
MIT GIS IN DER PRAXIS DES ÖFFENTLICHEN GESUNDHEITSWESENS**

**- dargestellt am Beispiel einer Untersuchung zur Verbreitung von
Legionelleninfektionen im Einzugsgebiet eines Brandenburger Klinikums -**

Master Thesis

**vorgelegt von
Dipl. Geogr. Ina Stalleicken**

Bonn, Februar 2005

Titel der Arbeit:

**Die Anwendung räumlich-statistischer Analysemethoden
mit GIS in der Praxis des Öffentlichen Gesundheitswesens**

**- dargestellt am Beispiel einer Untersuchung zur Verbreitung von
Legionelleninfektionen im Einzugsgebiet eines Brandenburger Klinikums –**

vorgelegt von:

Dipl. Geogr. Ina Stalleicken (U 1001)

Prof. Dr. Josef Strobl

Leiter des Institutes für Geographie und angewandte Geoinformatik
der Universität Salzburg
Heilbrunnerstraße 34
5020 Salzburg

PD Dr. Thomas Kistemann

Institut für Hygiene und Öffentliche Gesundheit
der Universität Bonn
Leiter des WHO Kollaborationszentrum für Wassermanagement und
Risikokommunikation zur Förderung der Gesundheit
Forschungsgruppenleiter Medizinische Geographie und Public Health
Sigmund-Freud-Str. 25
53105 Bonn

I. Inhalt

1.	Einleitung.....	9
1.1	Problemstellung.....	10
1.2	Zielsetzung.....	12
2.	GIS und räumlich-statistische Analysemethoden.....	14
2.1	Beschreibung der Datentypen.....	16
2.2	Darstellung wichtiger Techniken.....	17
2.2.1	Nächste Nachbar Statistik und K-Funktion.....	19
2.2.2	Punktdichteschätzungen.....	22
2.2.3	Räumliche Autokorrelation.....	24
3.	Das Anwendungsbeispiel: Legionellen und Legionellose.....	25
3.1	Legionellen.....	25
3.2	Legionellose in Deutschland.....	27
4.	Methodik.....	31
4.1	Datenerhebung.....	31
4.1.1	Serologische Tests.....	31
4.1.2	Der Fragebogen.....	32
4.2	Datenerfassung.....	32
4.3	Datenanalyse.....	33
4.3.1	Deskriptive statistische Analyse.....	34
4.3.2	Analytisch-epidemiologische Kohortenstudie.....	34
4.3.3	Räumlich-statistische Analyse.....	36
4.3.3.1	Flächenhafte Analysemethoden.....	37
4.3.3.2	Point Pattern Analysis Methoden.....	38
5.	Ergebnisse der Datenauswertung.....	45
5.1	Deskriptive Datenanalyse.....	45
5.1.1	Deskriptive Analyse der Patientendaten im Untersuchungszeitraum.....	45
5.1.2	Deskriptive Analyse des Serostatus der Patienten.....	46
5.2	Analytisch-epidemiologische Untersuchung.....	51
5.2.1	Risikoberechnung anhand der Ergebnisse der Befragung.....	51
5.3	Räumlich-statistische Analyse der Patientendaten.....	54
5.3.1	Flächenhafte räumliche Analyse.....	54
5.3.2	Punktbasierte räumliche Analyse der Patientendaten im Untersuchungsgebiet.....	58
5.3.2.1	Die Verteilung der serologisch positiver Patienten im Untersuchungsgebiet.....	60
5.3.2.2	Die Dichte serologisch positiver Patienten im Untersuchungsgebiet.....	64
5.3.2.3	Identifikation von Clustern serologisch positiver Patienten.....	66
6.	Diskussion.....	91
6.1	Erste Hinweise der deskriptiven Datenauswertung.....	91
6.2	Risikoanalyse auf Grundlage des Fragebogens.....	93
6.3	Interpretation der räumlich-statistischen Analyse.....	93

6.3.1	Flächenhafte räumliche Analyse	94
6.3.2	Geocodierung der Patientenadressen.....	95
6.3.3	Punktbasierte räumliche Analyse	96
7.	Ausblick	100
8.	Zusammenfassung	102
9.	Literatur.....	104
10.	Anhang	111

II. Abbildungsverzeichnis

Abb. 1:	Zeitliches Auftreten der Erkrankungen im Klinikum des Landes Brandenburg	11
Abb. 2:	Schematischer Überblick über Transformationsoperationen in GIS	15
Abb. 3:	Verfahren der Punktdichteschätzung.....	22
Abb. 4:	Prinzip der Kernel density Schätzung	23
Abb. 5:	Legionellenausbreitungsmöglichkeiten in wassertechnischen Systemen	25
Abb. 6:	Absolute Anzahl der Legionellenerkrankungen im Jahr 2003	28
Abb. 7:	Legionelleninfektionen in Deutschland – Meldungen des Robert Koch Institutes -.....	28
Abb. 8:	Legionellenerkrankungen in Deutschland unter Berücksichtigung des bundesdeutschen Durchschnitts	29
Abb. 9:	Veränderung der Legionelleninzidenzen von 2001 bis 2003 in Deutschland.....	30
Abb. 10:	Eingabeformular der Datenbank „Legionellen FO“	32
Abb. 11:	Methodische Vorgehensweise	33
Abb. 12:	Vorgehensweise bei der deskriptiven Datenanalyse.....	34
Abb. 13:	Risikoanalyse in SPSS 12	35
Abb. 14:	Anwendung räumlicher Analysemethoden	36
Abb. 15:	Berechnung der räumlichen Autokorrelation mit Crimestat 2.0	37
Abb. 16:	Berechnung der räumlichen Punkteverteilung mit Crimestat 2.0	39
Abb. 17:	Beispiel der Kernel Density Berechnung in ArcGIS 8.....	41
Abb. 18:	Darstellung der hierarchischen Clustermethode.....	42
Abb. 19:	Beispiele der Nearest neighbour hierarchical clustering Methode	42
Abb. 20:	Beispiel der Risk-adjusted nearest neighbour hierarchical clustering Methode	44
Abb. 21:	Patientenaufnahme pro Monat des Untersuchungszeitraums.....	45
Abb. 22:	Patienten nach Geschlecht und Altersklassen	46
Abb. 23:	Auswertung der IgM und IgG Serologie.....	46
Abb. 24:	Auswertung des Serostatus nach dem Zeitpunkt der Immunantwort.....	47
Abb. 25:	Prävalenz pro Monat der IgM-positiven Patienten.....	48
Abb. 26:	Prävalenz pro Monat der IgG-positiven Patienten	48
Abb. 27:	Prävalenz der eingewiesenen Patienten des Klinikums	49
Abb. 28:	Häufigkeit der Krankenhausaufenthalte und Serostatus	50

Abb. 29: Auswertung der Befragung zum Besuch von Schwimmbädern/ Sportstätten	50
Abb. 30: Absolute Anzahl IgG positive Patienten je Postleitzahlbezirk.....	54
Abb. 31: Prävalenz de IgG positiven Patienten je Postleitzahlbezirk.....	55
Abb. 32: Berechnung der räumlichen Autokorrelation mit Crimestat 2.0	56
Abb. 33: Prävalenz und Lokationsquotient für die IgG-positiven Patienten	57
Abb. 34: Ergebnis der Geocodierung der Patientenadressen.....	58
Abb. 35: Patienten mit positiven Serostatus.....	59
Abb. 36: IgM und IgG-positive Patienten nach Serotypen klassifiziert	59
Abb. 37: Serologisch positive Patienten mit einem regelmäßigen Besuch von Sportanlagen und Sportstätten.....	60
Abb. 38: Ergebnis der Berechnung der Nächste-Nachbar Statistik für IgG positive Patienten	61
Abb. 39: Ergebnis der Berechnung der Nächste-Nachbar Statistik für IgM positive Patienten	62
Abb. 40: Ergebnis der K-Funktion für IgG positive Patienten.....	63
Abb. 41: Ergebnis der K-Funktion für IgM positive Patienten	63
Abb. 42: Kernel-density Interpolation für IgG positive Patienten (Radius in Metern).....	64
Abb. 43: Kernel-density-Interpolation für IgM positive Patienten (verschiedene Suchradien I).....	65
Abb. 44: Kernel-density-Interpolation IgM positiver Patienten (verschiedene Suchradien II).....	66
Abb. 45: Standardeinstellungen der NNH-Methode und deren Ergebnis	67
Abb. 46: NNH-Clustering, Minimum = 10 Punkte für IgG positive Patienten	67
Abb. 47: Vergleich der NNh-Methode mit der Kernel-density-Interpolation für IgG positive Patienten	68
Abb. 48: NNH-Clustering, Minimum = 5 Punkte für IgG positive Patienten	69
Abb. 49: Simulation der NNH-Methode (Mindestanzahl Patienten pro Cluster = 5)	70
Abb. 50: NNH-Clusterberechnung, Minimum = 15 Punkte für IgG positive Patienten.....	70
Abb. 51: NNH-Clusterberechnung, Minimum = 20 Punkte für IgG positive Patienten.....	71
Abb. 52: Variation des Wahrscheinlichkeitsgrades auf 99,9%.....	72
Abb. 53: NNH-Clusterberechnung, Minimum = 10 Punkte für IgM positive Patienten.....	72
Abb. 54: NNH-Clusterberechnung, Minimum = 5 Punkte für IgM positive Patienten.....	73
Abb. 55: NNH-Clusterberechnung, Minimum = 10 Punkte für IgG positive Patienten mit einem regelmäßigen Besuch von Schwimmbädern und sonstigen Sportstätten	74
Abb. 56: RNNH-Clusterberechnung unter Berücksichtigung der absoluten Patientenzahl für IgG mit Standardeinstellungen (min=10).....	75
Abb. 57: Simulation der RNNH-Methode (Mindestanzahl Patienten pro Cluster =10)	76
Abb. 58: RNNH-Clusterberechnung unter Berücksichtigung der absoluten Patientenzahl für IgG mit Standardeinstellungen (Min =15).....	77
Abb. 59: RNNH-Clusterberechnung – detaillierte Betrachtung der Straßenzüge –	78
Abb. 60: Simulation der RNNH-Methode (Mindestanzahl Patienten pro Cluster =15)	79
Abb. 61: RNNH-Clusterberechnung – detaillierte Betrachtung der Straßenzüge II –	79
Abb. 62: RNNH-Clusterberechnung für IgG positive Patienten (min=20 Patienten).....	80
Abb. 63: RNNH-Clusterberechnung für IgG positive Patienten (min=20) II	81
Abb. 64: RNNH-Clusterberechnung für IgG positive Patienten (min=25).....	81
Abb. 65: RNNH-Clusterberechnung mit einer Standardabweichung der Ellipse von 2	82

Abb. 66: RNNH-Clusterberechnung für IgG positive Patienten mit einem angepassten Suchradius von 50 Punkten	84
Abb. 67: RNNH-Clusterberechnung mit festgesetztem Radius der Kernel density Berechnungen.....	85
Abb. 68: RNNH-Clusterberechnung für IgM positive Patienten (min=5) mit einem angepassten Suchradius von 100	86
Abb. 69: RNNH-Clusterberechnung für IgM positive Patienten (min=5) und angepasster Suchradius von 50 Punkten	87
Abb. 70: RNNH-Clusterberechnung für IgM positive Patienten (min=5) unter Anwendung eines festgesetzten Suchradius von 5000 m.....	88
Abb. 71: RNNH-Clusterberechnung für IgG positive Patienten (min=5) mit regelmäßigen Besuch von Schwimmbädern und sonstigen Sportstätten	89
Abb. 72: RNNH-Clusterberechnung für IgG positive Patienten (min=10) mit regelmäßigen Besuch von Schwimmbädern und sonstigen Sportstätten	90
Abb. 73: Probleme der Gebietsreform im Land Brandenburg.....	95
Abb.74: Anzahl Cluster bei festgelegter Mindestanzahl Patienten pro Cluster (NNH Methode)	97
Abb. 75: Anzahl Cluster bei einer festgelegten Mindestanzahl Patienten pro Cluster.....	98

III. Tabellenverzeichnis

Tab. 1: Auswahl räumlich-statistischer Analysemethoden (in Anlehnung an Bailey 1998)	18
Tab. 2: Gemeldete Legionelleninfektionen aus 25 europäischen Ländern von 1993 bis 2002	26
Tab. 3: Berechnung des relativen Risikos für die Ergebnisse der IgM Serologie	51
Tab. 4: Berechnung des relativen Risikos für die IgG Serologie.....	52
Tab. 5: Berechnung des relativen Risikos für beide Antikörperklassen.....	52
Tab. 6: Ergebnisse der Risikoberechnung zu den Expositionsfaktoren Geschlecht und Alter	53
Tab. 7: Ergebnis der Clusterberechnung IgG positiver Patienten (min=10)	68
Tab. 8: Ergebnis der Clusterberechnung IgG positiver Patienten (min=5)	69
Tab. 9: Ergebnis der Clusterberechnung IgG positiver Patienten (min=15)	71
Tab. 10: Standardeinstellungen zur RNNH-Clustering Methode	75
Tab. 11: Ergebnis der Rnnh-Clusterberechnung IgG positiver Patienten (min=10)	76
Tab. 12: Ergebnis der Rnnh-Clusterberechnung IgG positiver Patienten (min=15)	78
Tab. 13: Veränderung der Standardabweichung der Ellipsen für IgG positive Patienten.....	82
Tab. 14: Ergebnis der RNNH-Clusterberechnung IgG positiver Patienten (min=15) mit veränderter Standardabweichung	83
Tab. 15: Veränderung des Suchradius bei der RNNH-Methode IgG positiver Patienten	83
Tab. 16: Ergebnis der RNNH-Methode mit einem variablen Suchradius (n=50)	83
Tab. 17: Modifikation der Standardeinstellungen für IgG positive Patienten: Festgesetzter Radius	84
Tab. 18: Ergebnisse der RNNH-Berechnungen mit einem festgesetzten Suchradius von 5000m.....	85
Tab. 19: Modifikation der Standardeinstellungen für IgM positive Patienten: Variabler Radius	86

Tab. 20: Modifikation der Standardeinstellungen für IgM positive Patienten: Angepasster Radius (n=50).....	87
Tab. 21: Modifikation der Standardeinstellungen für IgM positive Patienten: Festgesetzter Radius....	88
Tab. 22: Modifikation der Standardeinstellungen für IgG positive Patienten mit einem regelmäßigen Besuch von Schwimmbädern und sonstigen Sportstätten (variabler Radius)	89

Abkürzungsverzeichnis:

AG: Antigen

IgG: Immunglobulin G

IgM: Immunglobulin M

NNH: Nearest neighbour hierarchical cluster analysis

PPA: Point Pattern Analysis

RNNH: Risk-adjusted nearest neighbour hierarchical cluster analysis

RKI: Robert Koch Institut

Begriffe:

ANTIGEN: als Antigen wird jede Substanz bzw. jedes Merkmal auf Oberflächen bezeichnet, das eine Bildung von Antikörpern hervorruft

ANTIKÖRPER: Eiweißstoffe, die vom Immunsystem gebildet werden. Vereinfacht ausgedrückt, sind sie in der Lage, an die Fremdkörper anzudocken und diese zu zerstören

GRAM-NEGATIV: Färbung von Bakterien im luftgetrockneten, hitzefixierten Ausstrichpräparat nach dem Gramschen Färbeverfahren ins Rote (im Gegensatz zu gram-positiven Bakterien, die sich blau färben)

ICD 10 CODE: internationale Klassifikation von Krankheiten (Diagnose). ICD10 SGBV wird in Deutschland als Schlüssel zur Angabe von Diagnosen, vor allem zur Abrechnung mit Krankenkassen verwendet

INZIDENZ: Neuerkrankungsrate in einem bestimmten Zeitraum (meist Anzahl pro 100.000 Einwohner)

IMMUNGLOBULINE: Antikörper werden in der Fachsprache als Immunglobuline bezeichnet. Es handelt sich hierbei um Eiweißstoffe, die Menschen zur Abwehr von Antigenen (Infektionserregern oder körperfremdem biologischen Material) produzieren

LETALITÄTSRATE: Ist das Verhältnis der Anzahl der an einer bestimmten Krankheit Verstorbenen zur Anzahl neuer Fälle

NOSOKOMIAL: im Krankenhaus erworben

PNEUMONIE: akute oder chronische Entzündung des Lungengewebes, meist verursacht durch Bakterien, Viren, Pilze, Protozoen, etc.

PRÄVALENZ: Anteil der Bevölkerung, der an einer bestimmten Krankheit zu einem bestimmten Zeitpunkt oder in einem bestimmten Zeitraum erkrankt ist

SURVEILLANCE: Beobachtung, Registrierung und Auswertung des aktuellen Infektionsgeschehens

1. Einleitung

In der Epidemiologie werden schon seit vielen Jahren räumliche Analysemethoden mit der Hilfe Geographischer Informationssysteme angewendet. So fällt Geoinformationssystemen in allen Bereichen der Gesundheitsforschung eine zentrale Rolle zu, unabhängig davon, ob es um die Beschreibung und Erklärung raumbezogener Variation von Krankheit oder um die Planung von Gesundheitseinrichtungen geht (Gatrell und Senior 1999). In der Krankheitsökologie (Geographische Epidemiologie) finden GIS heute Anwendung bei der Erkennung von Gesundheitsproblemen, dem Nachweis signifikanter räumlicher Muster von Krankheitsfällen mit Hilfe räumlich-statistischer Verfahren (z.B. Clusteridentifikation) sowie zur eigentlichen geographisch-ökologischen Analyse, bei der Gesundheitsdaten und Risikofaktoren der physischen, chemischen, biologischen oder sozialen Umwelt räumlich miteinander verknüpft werden. In der Gesundheitsforschung werden raumbezogene, zeitliche und organisatorische Strukturen und Prozesse, die mit der Planung von Gesundheitssystemen verbunden sind, untersucht, analysiert und bewertet (Schweikart & Kistemann 2004).

Bereits im 18. Jahrhundert wurde die Kartierung von Krankheiten zu einer zentralen Methode. Ein historisches Beispiel ist die Weltkarte der Krankheiten von Finke aus dem Jahr 1792 (Kistemann et al. 2002, Barret 2000) und eine Kartierung von Gelbfieberanfällen im Jahr 1798 im Hafen von New York (Stevenson 1965). Im Jahr 1852 kartierte August Petermann die großen Cholera-Epidemien von 1831, 1832 und 1833, wobei neben der reinen Visualisierung die Karte zunehmend als exploratives Werkzeug zur Datenanalyse diente (Schweikart und Kistemann 2004, Kistemann 2002). Die wohl bekannteste Karte zur Aufklärung von Krankheitsverbreitungen ist die des Arztes John Snow (1813-1858). Anhand der Kartierung der örtlichen Wasserversorgungsverhältnisse konnte im Jahre 1849 die Ursache der Cholera in London, in diesem Fall die Kontamination einer öffentlichen Pumpe in der Broad Street, enthüllt werden. Die Pumpe wurde daraufhin abgestellt, die Erkrankung eingedämmt. Im Zusammenhang mit der Kartierung des Cholera-Ausbruchs im Jahr 1854 (Snow 1855) wurde John Snow – zwar zu Unrecht (Brody et al. 2000) – die Einführung der kartengestützten raumzeitlichen Analyse in die induktive Krankheitsursachenforschung zugeschrieben (Schweikart und Kistemann 2004). Im 20. Jahrhundert wurden verschiedene medizinisch-geographische Atlanten, die sich mit der Verbreitung von Krankheiten auseinandersetzten, veröffentlicht, darunter „The world atlas of infectious diseases“ (Rodenwaldt und Jusatz 1952-1961), „The National Atlas of Disease Mortality in the United Kingdom“ (Howe 1963), „The Atlas of Avoidable Death“ (European Community 1991), „The London International Atlas of AIDS“ (Smallman-Raynor 1992) oder „The Atlas of Cancer Diseases of the Federal Republic of Germany“ (Becker und Wahrendorf 1998) (Kistemann et al. 2002). Walter und Birnie (1991) erfassten in einer Erhebung 49 internationale, nationale und regionale Atlanten, die die Verbreitung von Krankheiten zum Inhalt hatten (Elliot und Wartenberg 2004)

Heute werden zunehmend Geographische Informationssysteme in der Krankheitsökologie und der Gesundheitssystemforschung angewendet. Sie verfügen prinzipiell über die automatisierte Durchführung von räumlich-statistischen Analysemethoden, die ein solches System zu mehr als nur einem Computersystem machen, welches in der Lage ist, raumbezogene Daten zu verarbeiten. GIS ist daher „mehr als Denkweise zu begreifen: eine besondere Art,

Probleme zu bearbeiten.“ (Schweikart 2004: 33). Eine Vielzahl von Studien haben die Anwendung Geographischer Informationssysteme im Gesundheitswesen dokumentiert. Einen Überblick geben Cromely und Mc Lafferty (2002), Gatrell und Loytönen (1998), Gatrell und Senior (1999), Higgs und Gould (2001), de Lepper et al. (1994) und Maheswaran und Craglia (2004). Im Jahr 2003 wurde die Bezeichnung „Geographical Information System“ zum Thesaurus der US National Library of Medicine hinzugefügt, was die große Bedeutung von GIS im Gesundheitssektor reflektierte (Boulous 2004).

1.1 Problemstellung

Der Einsatz von GIS in der Epidemiologie erfolgte anfänglich überwiegend in der Bekämpfung und Kontrolle parasitärer Erkrankungen, weil die komplexen Wechselwirkungen zwischen Umweltfaktoren und Krankheitserregern, Vektoren, den verschiedenen Wirtstieren und Erregerreservoirien in einem System erfasst werden können (Mott et al. 1995). Die eigentliche Stärke von GIS besteht dabei in dessen Funktionalität, raumbezogene Daten zu analysieren.

Die Möglichkeiten des Einsatzes von GIS im Öffentlichen Gesundheitswesen sind weltweit bekannt, insbesondere in Forscherkreisen. Die Anwendung räumlich-statistischer Analysemethoden mit GIS ist jedoch in der Praxis des Öffentlichen Gesundheitswesens noch nicht sehr verbreitet. *„GIS have been usually applied to time-limited, single, isolated aetiological research or surveillance issues processing mainly retrospective data rather than to ongoing, broad efforts and wide scale applications processing real-time or near real-time data for health planning, promotion and protection“* (Boulous 2004: 2). Gould (1992) beobachtete in einer Erhebung und Bewertung von Gesundheitseinrichtungen des Vereinigten Königreiches, dass das Bewusstsein über die Existenz von GIS durchaus groß ist, die Anwendung sich jedoch lediglich auf die einfache Visualisierung (mapping) beschränkte. In Veröffentlichungen des staatlichen Gesundheitsdienstes (National Health Service) wird die Notwendigkeit von Informationstechnologien betont, Geographische Informationssysteme werden jedoch nicht erwähnt (Cockings et al. 2004). Im Gegensatz dazu finden GIS in anderen Bereichen, wie dem Dienstleistungssektor (Longley und Clarke 1995), der Telekommunikation (Fry 1999), dem Transportwesen (Waters 1999) und der Stadtplanung (Yeh 1999) breite Anwendung. Die Ursachen für eine derzeit noch limitierte Nutzung von GIS in der Praxis des Gesundheitswesens sind u.a. die limitierte Verfügbarkeit von georeferenzierten Gesundheitsdaten, aber auch Wissensdefizite hinsichtlich der vielfältigen Anwendungs- und Einsatzmöglichkeiten (Cockings et al. 2004). Die Weltgesundheitsorganisation WHO sieht ebenso Anwendungshindernisse in den hohen Anschaffungskosten der Hard- und Software sowie in der großen Komplexität der Programme, die eine praktische Anwendung kosten- und zeitintensiv werden lassen (WHO 2004).

Auch in Deutschland gibt es verschiedene Ansätze, die sich mit der Anwendung von GIS im Rahmen der Krankheitsökologie oder Gesundheitssystemforschung beschäftigen (Schweikart & Kistemann 2004). So wurden beispielsweise GIS-gestützte geostatistische Interpolationsverfahren (Kriging) zur Überwachung der Influenza-Aktivität genutzt (Uphoff et al. 2004, Stalleicken et al. 2004). Ein Beispiel aus dem Bereich der Gesundheitssystemforschung ist

die Optimierung der Versorgung chronisch Kranker mittels GIS-gestützter Analyse der vorhandenen ärztlichen Infrastruktur (Krafft et al. 2004) oder die Untersuchung der räumlichen Verbreitung von Gesundheitseinrichtungen mit GIS am Beispiel der Stadt Hamburg (Traub 2004).

Die Anwendung von GIS in der Praxis des Öffentlichen Gesundheitswesens ist somit in Deutschland begrenzt. Vornehmlich werden GIS weiterhin als Visualisierungswerkzeuge verwendet, z.B. zur Visualisierung der Daten zu Infektionskrankheiten, die im Rahmen der Infektionssurveillance erhoben werden (Rissland et al. 2003). Räumlich-statistische Analysemethoden mit GIS werden kaum eingesetzt.

In der vorliegenden Studie zur Verbreitung von Legionelleninfektionen im Einzugsgebiet eines Brandenburger Klinikums sollen neben statistischen Methoden insbesondere räumlich-statistische Methoden mit GIS angewendet werden und deren Nutzen in der Praxis des Öffentlichen Gesundheitswesens beleuchtet werden.

Legionellen als fakultativ pathogene Keime haben die Eigenschaft, sich vor allem in vom Menschen geschaffenen wässrigen oder feuchten Umgebungen zu vermehren. Aus diesem Grund sind sie nicht nur für Ärzte, die sich mit Legionellose befassen müssen, sondern ebenso für Krankenhaushygieniker und Konstrukteure von besonderem Interesse (Pagano 1994).

In einem Klinikum im Land Brandenburg traten in den Jahren 2002 und 2003 mehrere Legionellen-Erkrankungen auf. Dabei handelte es sich um zwei Ausbrüche, bei denen einmal sieben und einmal fünf Patienten des Klinikums erkrankten. Eine direkte Zuordnung der Untersuchungsergebnisse für Legionellen aus dem Trinkwasser-Installationsnetz des Klinikums zu den erkrankten Patienten war nur selten möglich, weil Untersuchungsergebnisse nicht immer im zeitlichen und örtlichen Zusammenhang zu den jeweils betroffenen Patienten vorlagen (RKI 2004) (Abb. 1)

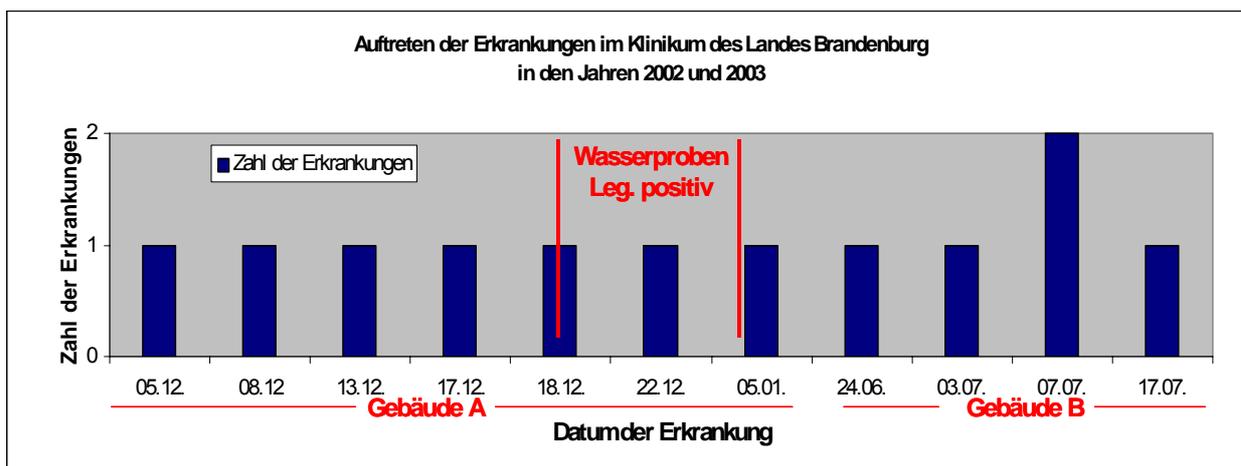


Abb. 1: Zeitliches Auftreten der Erkrankungen im Klinikum des Landes Brandenburg (in Anlehnung an RKI 2004)

Es liegt zwar nahe, dass beide Ausbrüche von Legionellen-Pneumonien im Zusammenhang mit Legionellen-kontaminiertem Trinkwasser stehen. Ein eindeutiger Infektionsweg für die Übertragung der Legionellen konnte jedoch nicht verifiziert werden. Auch die Analyse der Patientendaten ergab keine eindeutigen Hinweise auf bestimmte Faktoren, die mit einem erhöhten Erkrankungsrisiko vergesellschaftet waren (RKI 2004).

Vor dem Hintergrund der ungeklärten Ursache des Auftretens von Legionelosen wurde eine prospektive epidemiologische Studie durchgeführt, bei der der Legionella-Serostatus der Klinikumspatienten zum Zeitpunkt ihrer Aufnahme sowie mögliche Risikofaktoren für Legionella-Infektionen außerhalb des Klinikums erfasst und analysiert wurden. Diese Studie wurde im Zeitraum August 2003 – Juli 2004 durchgeführt. Zur Klärung des Infektionsweges und des Risikos der Patienten, die von beiden Ausbrüchen betroffen waren sowie von Patienten, die in Zukunft in das Krankenhaus eingewiesen werden, wurden die für die Erkrankung relevanten Informationen der Patienten per Fragebogen erfasst. Dabei handelt es sich um allgemeine Patientendaten (Adresse mit Angabe von Straße und Hausnummer), um die Infektionsserologie der Patienten sowie um Fragen zu möglichen relevanten Risikofaktoren vor der Einlieferung in das Krankenhaus. Dazu zählen Reisen, der Besuch von Schwimmbädern oder Sportstätten, Luftbefeuchter, Rückkühlwerke etc.

1.2 Zielsetzung

In dieser Studie sollen am Beispiel der Verbreitung von Legionelleninfektionen im Einzugsgebiet eines Brandenburger Klinikums räumlich-statistische Analysemethoden mit Hilfe von Geoinformationssystemen angewendet und gleichzeitig deren Möglichkeiten und Grenzen für die Praxis des Öffentlichen Gesundheitswesens aufgezeigt werden.

Aufgrund der Epidemiologie der Legionellen, insbesondere deren Vermehrung und Verbreitung in der Umwelt, sind räumliche Analysemethoden mit Hilfe von GIS hier von besonderer Bedeutung. Im Bereich der umwelt-assoziierten Erkrankungen, zu denen auch die Legionelosen zu rechnen sind, ist die räumlich-statistische Analyse von Risikofaktoren, die mit der Krankheit in Verbindung gebracht werden können, eine Möglichkeit, Fragen des Infektionsweges sowie des Infektionsrisikos zu beantworten. In der Regel wird der Wohnort der Probanden als räumlicher Bezugspunkt angenommen. Hierfür wurden zwei Wege ausgewählt:

1. die flächenhafte Analyse der räumlichen Verteilung serologisch positiver Patienten auf Basis der Postleitzahlenbezirke
2. die punktbasierte Analyse serologisch positiver Patienten auf der Basis der Geocodierung von Patientenadressen

Aus dem Bereich der GIS-gestützten räumlichen Analyse wurden vornehmlich Verfahren der Point Pattern Analysis (PPA-Methoden) angewendet, und zwar im einzelnen:

- zur Bestimmung der räumlichen Punktverteilung mit dem Schwerpunkt auf die Nächste Nachbar Statistik
- zur Bestimmung der räumlichen Punktdichte, u.a. nach positivem und negativem Serostatus (Erstellung von Punktdichtekarten mit Hilfe der Kernelmethode).
- zur Identifikation räumlicher Cluster, d.h. signifikant höherer Konzentrationen von Punkten in Teilbereichen des Untersuchungsgebietes

Voraussetzung einer solchen räumlich-statistischen Analyse ist die so genannte Geocodierung der Patientendaten, wobei die Wohnadresse eines jeden Patienten zugrunde gelegt wurde. Die geographischen Koordinaten können bestenfalls Hausnummer-genau zugewiesen werden.

Die räumliche und statistische Analyse der erhobenen Daten von etwa 14.500 Patienten soll Aufschlüsse über mögliche umwelt-assoziierte Risiken erbringen. Es soll ermittelt werden, ob die Patienten schon vor einer Einweisung mit Legionellen in Kontakt gekommen sind und wo, wenn serologisch nachgewiesen, möglicherweise eine ursächliche Infektion stattgefunden haben kann. Zudem kann festgestellt werden, ob die Patienten im Einzugsgebiet des Krankenhauses einem überdurchschnittlich hohen Infektionsrisiko ausgesetzt sind. Die geocodierten Patientenadressen sind die Ausgangsbasis der räumlich-statistischen Analyse. Sie ermöglicht die Analyse der Verteilung der Patienten im Untersuchungsgebiet, die Untersuchung der Patientendichte sowie die Identifikation von Patientenclustern mit positiver Legionellen-Serologie.

Die vielfältigen Möglichkeiten räumlicher Analysemethoden in dieser Studie bedingen gleichzeitig den Einsatz verschiedener Softwareprodukte (u.a. ArcInfo 8, Crimestat 2.0), da nicht eine einzige GIS-Software über die genannten Methoden verfügt. Es soll ermittelt werden, welche räumlich-statistischen Methoden mit der Unterstützung von GIS einen geeigneten Beitrag zur Ursachenforschung leisten kann und worin Vor- und Nachteile der eingesetzten Methoden in dieser Studie im Speziellen, aber auch in der Praxis des Öffentlichen Gesundheitswesens im Allgemeinen bestehen. Die Anwendung der Analysetechniken von GIS und GIS-verwandten Softwareprodukten in dieser Studie soll deren Nutzen für die Praxis des Öffentlichen Gesundheitswesens unter Beweis stellen.

2. GIS und räumlich-statistische Analysemethoden

In der Literatur finden sich viele Definitionen von GIS, die sich in Abhängigkeit von Erfahrungen und Wahrnehmungen des jeweiligen Anwendungsbereiches Geographischer Informationssysteme unterscheiden „*As such the term invokes different perceptions dependent on whether the viewpoint is that of the software developer, the system marketer, the data provider, the application specialist or the academic researcher, among others.*“ (Reader 1995:3). Die Definition von Strobl (1988) stellt dabei in diesem Zusammenhang eine umfassende Sicht von GIS dar. „*Ein Geographisches Informationssystem ist ein Computersystem zur Erfassung, Speicherung, Prüfung, Manipulation, Integration, Analyse und Darstellung von Daten, die sich auf räumliche Objekte beziehen.*“ (Strobl 1988)

Verschiedene Konzepte von GIS erschienen schon Anfang des 20. Jahrhunderts auf Grundlage des „Schichtenprinzips“, welches der deutsche Geograph Alfred Hettner (1859-1941) im Jahre 1927 in die Landschaftsforschung einführte. In den 1960er Jahren wurden erste Anwendungen der digitalen Bildverarbeitung eingesetzt, und es entstand parallel in dem „Harvard Laboratory for Computer Graphics and Spatial Analysis“ sowie an der ETH Zürich die Idee der unabhängigen Datenebenen. Durch eine Trennung von Geometriedaten einschließlich zugehöriger Attributdaten in thematische Ebenen, entstehen verschiedenen Modelle mit der Möglichkeit der Separation und Verknüpfung (Blaschke 1997). Dieses Prinzip wurde in viele Variationen (Layer, Ebenen, Themen) in der GIS-Software umgesetzt. Der Begriff GIS entstand jedoch erst auf dem ersten großen GIS-Symposium, dass von Roger Tomlinson im Jahre 1970 organisiert wurde (Longley 2001, Stahl 1997).

Die Frage, ob bei der Vielzahl multidisziplinärer Anwendungen Geographische Informationssysteme mehr als nur ein Werkzeug, Methode oder Wissenschaft darstellen, wird stets diskutiert. Goodchild (1993) nennt GIS ein Phänomen „*An encouraging recent trend has been the willingness of a broad spectrum of geographers to see GIS not as a tool that they can use in their own research, but as a phenomenon on which they can reflect and comment.*“ (Goodchild 1993:444). Wright et al. (1997) bezeichnet in diesem Zusammenhang GIS als ein Kontinuum aus Werkzeug und Wissenschaft. „*The science based on GIS...may advance the tools and toolmaking of GIS, as well as scientific research questions*“ (Wright et al. 1997:360). Neben den elementaren Bestandteilen der Erfassung, Verwaltung und Darstellung der Daten, kommt vor allem der Analyse eine wichtige Funktion zu. So unterscheiden sich nach Strobl (1992) Geographische Informationssysteme durch die analytische Funktionalität von vielen anderen Informationssystemen. Das zentrale Kriterium eines GIS ist die enge Integration von geometrischen und thematischen Attributen räumlicher Objekte. Hier kommen als wesentliche Werkzeuge neben der Datenerfassung und der kartographischen Bearbeitung Module zur Analyse und Modellierung zur Anwendung (Blaschke 1997).

Die analytische Auswertung im GIS bildet einen konstituierenden Bestandteil und ein wesentliches Spezifikum im GIS (Strobl 1992: 47) und verhilft, neue Informationen für entscheidungsunterstützende Abläufe aus vorliegenden Geodaten zu generieren. So basiert die Motivation zur Verwendung von GIS vorwiegend auf deren analytischem Potenzial.

„*Spatial analysis is the crux of GIS, the means of adding value to geographic data, and of turning data into useful information.*“ (Longley 2001)

Dabei ist die Analyse mehr als eine reine Überlagerung (Verschneidung) von verschiedenen Datenschichten, die als erste Stufe oder Vorstufe der Analyse bezeichnet werden kann (Blaschke 1997). Generell kann die Analyse

1. die räumliche Datenmanipulation
2. die räumliche Datenanalyse
3. die räumlich-statistische Datenanalyse
4. die räumliche Modellierung

umfassen.

In der Praxis sind diese Methoden nicht einfach voneinander zu unterscheiden, häufig sind alle vier Ansätze enthalten. In diesem Sinn werden räumliche Daten schrittweise verarbeitet und im weiteren Sinne Transformationen unterzogen (Strobl 1994-2003) (Abb. 2)

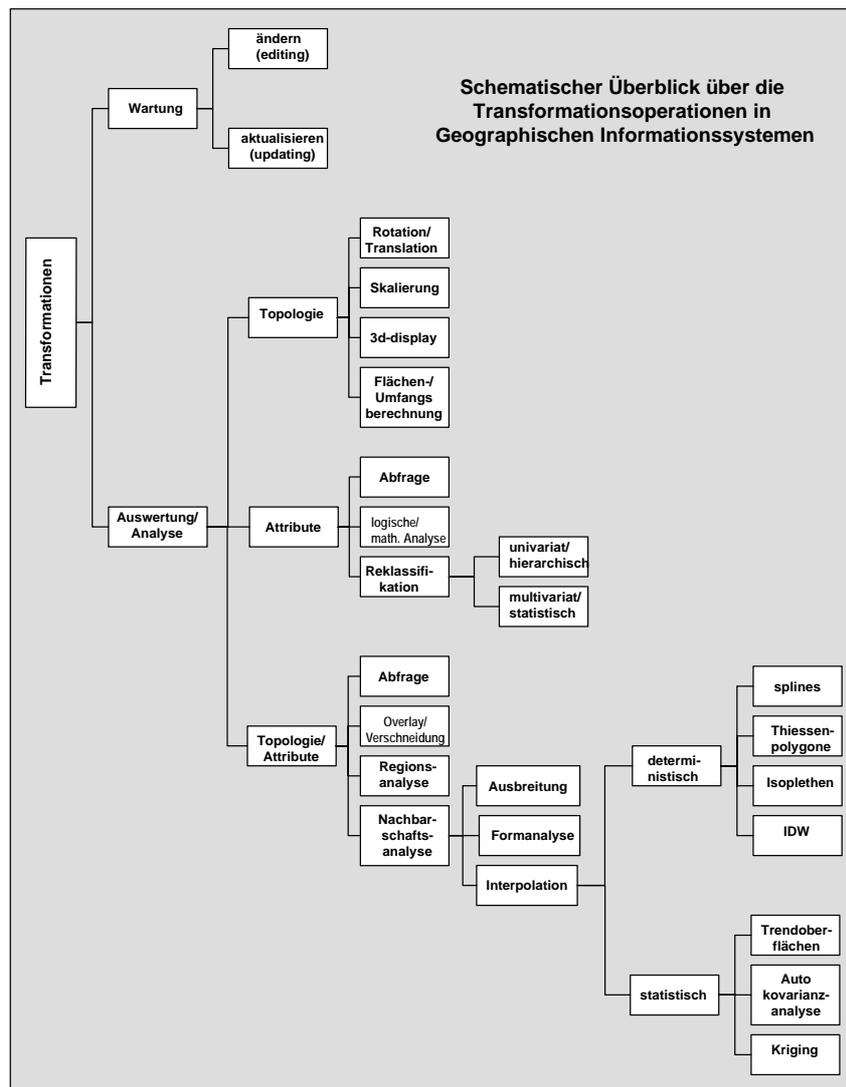


Abb. 2: Schematischer Überblick über Transformationsoperationen in GIS (Burrough 1986, Blaschke, 1997)

Deren Anwendung in der Praxis des Öffentlichen Gesundheitswesens verhilft zu einem besseren Verständnis von Krankheitsmustern, aber auch zu möglichen Assoziationen zwischen Krankheit und anderen räumlichen Phänomenen (Curtis et al. 2002). Deskriptive und explorative Techniken werfen neue Fragestellungen auf, so dass schließlich verschiedenste räum-

liche statistische Analysemethoden als Basis der räumlichen Modellierung zur Problemlösung eingesetzt werden (O'Sullivan 2003).

2.1 Beschreibung der Datentypen

Cressie (1993) hat eine Unterteilung der räumlichen Statistik in drei Untergebiete durchgeführt, die sich über verschiedene Typen von Daten definieren:

1. *Räumlich kontinuierliche Daten*, d.h. Daten, die theoretisch an jedem Punkt einer Region gemessen werden, jedoch nicht an jeder Position zur Verfügung stehen oder nur an einigen Positionen gemessen werden. Die wichtigste Fragestellung bei solchen Daten besteht darin, Prognosen abseits der Messpunkte zu erhalten und den Einfluss anderer Größen, sogenannter Kovariablen, auf die Untersuchungsgröße zu analysieren.
2. *Punktprozessdaten*, d.h. Daten, die den Ort eines Ereignisses beschreiben. Dies kann z.B. bei epidemiologischen Fragestellungen der Ort des Auftretens einer Krankheit sein. Die Analyse zielt auf die Identifikation besonderer Krankheitsmuster oder Krankheitshäufungen ab.
3. *Regionaldaten*, d.h. Daten, die sich auf größere Gebiete beziehen, z.B. administrative Einheiten. In der Regel handelt es sich dabei um Durchschnitts- oder Summenwerte dieser Regionen, z.B. die durchschnittliche Anzahl von Erkrankungen bezogen auf 100.000 Einwohner in einem Landkreis oder Postleitzahlbezirk. Wichtige Fragen betreffen die kartographische Darstellung von Regionaldaten sowie die Identifikation systematischer Unterschiede zwischen den Regionen und deren Erklärung durch eventuelle Kovariablen (Dreesmann 2004).

Bei der Schätzung räumlich kontinuierlicher Daten (1) spielen Interpolationsmethoden (z.B. Kriging) eine herausragende Rolle. Mit Hilfe solcher Verfahren werden Werte einer Variablen an nicht beprobten Orten aus erhobenen Daten geschätzt mit dem Ziel, aus einigen wenigen punktuellen Messungen ein möglichst gutes Bild und Verständnis der räumlichen Variation thematischer Oberflächen in einem Untersuchungsgebiet zu erlangen. „*Spatial interpolation is the prediction of values of a variable distributed in space at unsampled locations from measurements made at regularly or irregularly distributed sampled locations within the same area*“ (Burrough 1998). Das Ergebnis einer solchen räumlichen Interpolation ist eine diskrete, stetige Oberfläche, ein Raster, welches räumliche Variationen im Raum darstellt.

Der Hauptgrund für die Analyse von *Punktprozessdaten* (2) ist der Versuch, ein besseres Verständnis für den Prozess zu erlangen, der die Punktverteilung hervorgebracht hat. Man spricht in diesem Zusammenhang von der sogenannten Point Pattern Analysis - also der Analyse von räumlichen Punktverteilungen. Bei der Analyse der Punkteverteilung wird meistens ein theoretisches Punktmuster definiert, mit dem das beobachtete Punktmuster (reales Punktmuster) verglichen wird. Das theoretische Punktmuster wird als strikte räumliche Zufallsverteilung (complete spatial randomness CSR) bezeichnet und entsteht durch den sogenannten *homogeneous planar poisson point process*, welcher zwei Bedingungen impliziert (Lorup 2004):

1. Jeder Standort im Untersuchungsgebiet besitzt die gleiche Chance mit einem Punkt besetzt zu werden (Gleichheitsbedingung)
2. Die Besetzung des Standortes mit einem Punkt beeinflusst in keiner Weise die Besetzung eines anderen Standortes mit einem Punkt (Unabhängigkeitsbedingung)

Das Modell der räumlichen Zufallsverteilung kann mathematisch durch eine Poisson Wahrscheinlichkeitsverteilung beschrieben werden.

$$p(x) = \frac{(e^{-\lambda} \lambda^x)}{x!} \text{ für } x = 0, 1, 2, \dots$$

λ = zu erwartende Anzahl Punkte

e = Eulerzahl, eine mathematische Konstante (2,718282)

Diese theoretische räumliche Zufallsverteilung wird als Idealverteilung definiert, von der die wirklichen bzw. beobachteten Verteilungen statistisch signifikant abweichen können. Abweichungen erfolgen entweder in Richtung einer räumlich konzentrierten Verteilung (*concentrated or cluster pattern*) oder in Richtung einer räumlich gleichmäßigen Verteilung (*regular pattern*). In Kap. 2.2 werden verschiedene Techniken der Point Pattern Analysis vorgestellt.

Die Analyse der Regionaldaten (3) spielt eine bedeutende Rolle in vielen Fragen der Epidemiologie und der medizinischen Versorgungsforschung, aber auch in der Demographie oder Ökonomie, wenn Daten auf Ebene der administrativen Regionen vorliegen. Hier eignen sie sich insbesondere dafür, anhand der Visualisierung von beispielsweise Prävalenzen¹ oder Inzidenzen² einen ersten Überblick über Krankheitsverteilungen zu erlangen. Des Weiteren kann anhand verschiedener räumlich-statistischer Verfahren (z.B. Moran's I) ermittelt werden, inwieweit benachbarte Regionen miteinander räumlich korrelieren. Die Datenverfügbarkeit unterscheidet sich in Deutschland erheblich. So stehen verschiedene Statistiken zur Verfügung, z.B. zu meldepflichtigen Infektionskrankheiten (Robert Koch Institut Berlin), die sich zumindest zu einer kartographischen Darstellung und meist zur weiteren räumlich-statistischen Analyse eignen. Die Möglichkeiten bei kleinräumigen Untersuchungen sind hingegen aufgrund der Datenverfügbarkeit eingeschränkt. Überdies besteht insbesondere bei seltenen Erkrankungen das sogenannte Cluster oft aus weniger als zehn beobachteten Fällen. Eindeutige Falldefinitionen fehlen meist (Zöllner 2004).

2.2 Darstellung wichtiger Techniken

Unabhängig vom Typ der Daten, ob räumlich kontinuierliche Daten, Punktprozessdaten oder Regionaldaten, werden mit der Analyse jeweils ähnliche Ziele verfolgt (Dreesmann 2004):

¹ Prävalenz ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person an einem definierten Stichtag an einer betrachteten Krankheit erkrankt ist. Liegt an diesem Stichtag eine Population der Größe N vor und sind in dieser Gruppe M Personen erkrankt, so ergibt sich die Prävalenz P als Quotient der Anzahl M der Personen mit Krankheit zu dieser Populationsgröße N durch $P = M/N$.

² Die Inzidenz gibt die Wahrscheinlichkeit an, dass einen zufällig ausgewählte Person der Population innerhalb einer zeitlich begrenzten Periode (z.B. ein Jahr) an einer Krankheit neu erkranken wird. Liegt zu Beginn dieser Periode eine gesunde Population der Größe N_0 vor, und erkranken während der Periode I Personen neu, so errechnet sich die Inzidenz als Quotient $CI = I/N_0$.

1. die Aufbereitung der Daten für graphische Darstellungen sowie die Erkennung und Beschreibung von Strukturen in den Daten,
2. die Identifikation der Abweichungen von regelmäßigen Strukturen und die Bewertung, ob solche Abweichungen systematisch oder lediglich zufallsbedingt sind,
3. die Prognose von Werten dort, wo keine Messwerte zur Verfügung stehen und
4. die Anpassung statistischer Modelle an die Daten unter Berücksichtigung von Kovariablen.

Aus der Vielzahl räumlich-statistischer Analysemethoden werden im Folgenden diejenigen ausgewählt, die in enger Verbindung mit GIS stehen. Die Gliederung in Tab. 1 anhand der Datenstruktur ist allgemein üblich für die Diskussion räumlicher Statistik (Bailey 1998).

Tab. 1: Auswahl räumlich-statistischer Analysemethoden (in Anlehnung an Bailey 1998)

Datenstruktur	Dimension	
	Univariat	Multivariat
Lokale Daten (Punktprozessdaten)	Nächste-Nachbar-Methoden (Nearest Neighbour Methods) K-Funktion	Bivariate K-Funktionen Raum-Zeit-Interaktionen
	Punktdichteschätzung (Kernel density estimation) Kernel Regression Bayesian smoothing – ICM	Angepasste Punktdichteschätzungen (adaptive Kernel density estimation) Bayesian smoothing – ICM
Attributdaten (Regionaldaten, räumlich kontinuierliche Daten)	Räumliche Autokorrelation Variogramme	Multivariate räumliche Korrelation
	Trendoberflächenanalyse Kriging	Räumliche Regression Co-Kriging Raumzeitliche Modelle
		Räumlich lineare Modellierung (spatial general linear modelling) Cluster Analyse Canonical Correlation Multidimensional Scaling

Eine weitere Unterscheidung erfolgt anhand der Dimensionalität. Die Analyse von Punktprozessdaten, darunter die Analyse von räumlichen Punkteverteilungen, wird unter dem Begriff *point pattern analysis* (PPA) zusammengefasst. Die Ursprünge der Techniken, die heute in der statistischen Analyse von Punktmustern verwendet werden, sind vor mehr als 60 Jahren in der Pflanzenökologie entstanden (Lorup 2004). Danach wurden sie in verschiedensten Fachbereichen verwendet, darunter in der Archäologie und Anthropologie (Hodder & Orton 1976), aber auch in der Astronomie und in der Werkstoffkunde. In den 1980er Jahren haben insbesondere Statistiker komplexere PPA-Techniken entwickelt (Ripley 1981, Diggle 1985). Da viele PPA-Techniken sehr rechenintensiv und damit abhängig von Entwicklungen im Computersektor sind, erschienen die ersten PPA-Softwareprodukte erst in den 90er Jahren. Dabei sind zwei Entwicklungstypen erkennbar:

1. die Integration in vorhandene GIS-Software in Form von Extensions oder Modulen (z.B. ArcGIS extensions).
2. eine Stand-alone Software, die über eine Vielzahl von PPA-Methoden verfügt (z.B. CrimeStat 2.0).

Der Vorteil einer stand-alone-Software ist die Vielfalt der PPA-Techniken, die in einem solchen Umfang bisher in keinem GIS implementiert ist. Voraussetzung der Anwendung von PPA-Methoden ist die Existenz von x- und y-Koordinaten, die entweder anhand von Vermessungsinstrumenten (z.B. GPS) oder über die Geocodierung von Adressen erstellt werden. „*This is a branch of statistics which involves the statistical modelling of geographical phenomena (such as cases of disease and an underlying population) as a set of georeferenced points (usually residential addresses.*“ (Cockings et al. 2004: 170). Cockings et al. (2004) differenziert die point pattern analysis (PPA) von GIS, obwohl beide die Analyse von räumlichen Daten gemeinsam haben. „*Whilst there are similarities and overlaps between point pattern analysis and GIS, especially in that they are both concerned with the analysis of spatial data, we argue that there are conceptual differences between the two, even if these are difficult to define with clarity*“ (Cockings et al. 2004: 170). Unterschiede bestehen in Hinblick auf die Existenz statistischer Modelle, die zum Test von Hypothesen eingesetzt werden können. „*Possibly, the most significant distinction is that point pattern analysis employs specific statistical models which are fitted to the data to generate and test hypothesis; in this sense point pattern analysis has a much former statistical basis than GIS.*“ (Cockings et al. 2004: 170). Allgemeine Übersichten zur PPA-Analyse sind von Diggle (1993), Gatrell und Bailey (1996) und Gatrell et al. (1996) veröffentlicht worden, spezielle Anwendungsbeispiele von PPA-Methoden von Bhopal et al. (1992) sowie Diggle und Rowlingson (1994).

Im Folgenden sollen nun die im Rahmen dieser Studie relevanten räumlich-statistischen Analysemethoden, insbesondere die der PPA-Techniken, näher erläutert werden.³

2.2.1 Nächste Nachbar Statistik und K-Funktion

Handelt es sich um nur eine untersuchte Variable, kann das beobachtete Muster durch einen übergreifenden räumlichen Trend (*Kernel density estimation*) der Punktdaten oder anhand der Nachbarschaftsanalyse erklärt werden. Bei letzterer handelt es sich um eine Methode, räumliche Muster lokaler Punktdaten anhand der beobachteten Distanzen der jeweils nächsten Nachbarn verglichen mit den theoretisch erwarteten Distanzen unter Annahme einer räumlichen Zufallsverteilung zu untersuchen (Upton & Fingleton 1985).

„...nearest neighbour methods involve a method of exploring pattern in locational data by comparing graphically the observed distribution functions of event-to-event or random point-to-event nearest neighbour distances, either with each other or with those that may be theoretically expected from various hypothesized models, in particular that of spatial randomness.“ (Bailey 1998: 25, nach Upton 1985)

Die Distanz d_{ij} zwischen jedem Punktpaar (U_j, V_i) und (U_j, V_j) im Untersuchungsgebiet wird mit dem Satz von Pythagoras folgendermaßen ermittelt:

³ Für eine vollständige Beschreibung aller in Tab. 1 genannten Verfahren siehe Bailey (1998)

$$d_{ij} = \sqrt{(U_i - U_j)^2 + (V_i - V_j)^2}$$

Die durchschnittliche Nächste-Nachbar Distanz (*mean nearest neighbour distance*) berechnet sich demnach wie folgt:

$$\bar{d} = \left(\sum_{i=1}^N d_{ij} / N \right)$$

für N = Anzahl der Punkte im Untersuchungsgebiet; d_{ij} Distanz von Punkt i zu seinem nächsten Nachbarn j .

Die durchschnittliche (beobachtete) Nächste-Nachbar Distanz wird nun mit der durchschnittlichen Nächste-Nachbar Distanz einer räumlichen Zufallsverteilung verglichen, die sich folgendermaßen berechnet:

$$\delta = 1/(2\sqrt{\lambda})$$

λ = beobachtbare Punktdichte im Untersuchungsgebiet, d.h. $\lambda = N/A$, wobei N die Anzahl der Punkte im Untersuchungsgebiet und A die Fläche des Untersuchungsgebietes ist.

Der Vergleich der durchschnittlichen beobachteten Nächste-Nachbar Distanz mit der durchschnittlichen zu erwartenden Nächste-Nachbar-Distanz einer räumlichen Zufallsverteilung führt zur Formel der Nächste-Nachbar-Statistik R:

$$R = \bar{d} / \delta = 2\bar{d}\sqrt{(N/A)}$$

\bar{d} = durchschnittliche Nächste-Nachbar Distanz; δ = Nächste-Nachbar Distanz einer räumlichen Zufallsverteilung; N = Anzahl der Punkte im Untersuchungsgebiet; A = Fläche des Untersuchungsgebietes

Die Statistik R kann Werte zwischen 0 für eine perfekt räumlich konzentrierte Verteilung (d.h. alle Punkte liegen am gleichen Ort genau übereinander) und 2,149 für eine perfekt räumlich gleichmäßige Verteilung (d.h. alle Punkte liegen auf den Schnittpunkten eines dreieckigen Gitternetzes) annehmen. Sind die Punkte zufällig verteilt ist $R \sim 1$, bei einer eher konzentrierten Verteilung ist $\bar{d} < \delta$ und der Wert der R-Statistik $< 1,0$. Hingegen erwartet man bei einer eher gleichmäßigen Verteilung, dass $\bar{d} > \delta$, d.h. die R-Statistik $> 1,0$. (Lorup 2004)

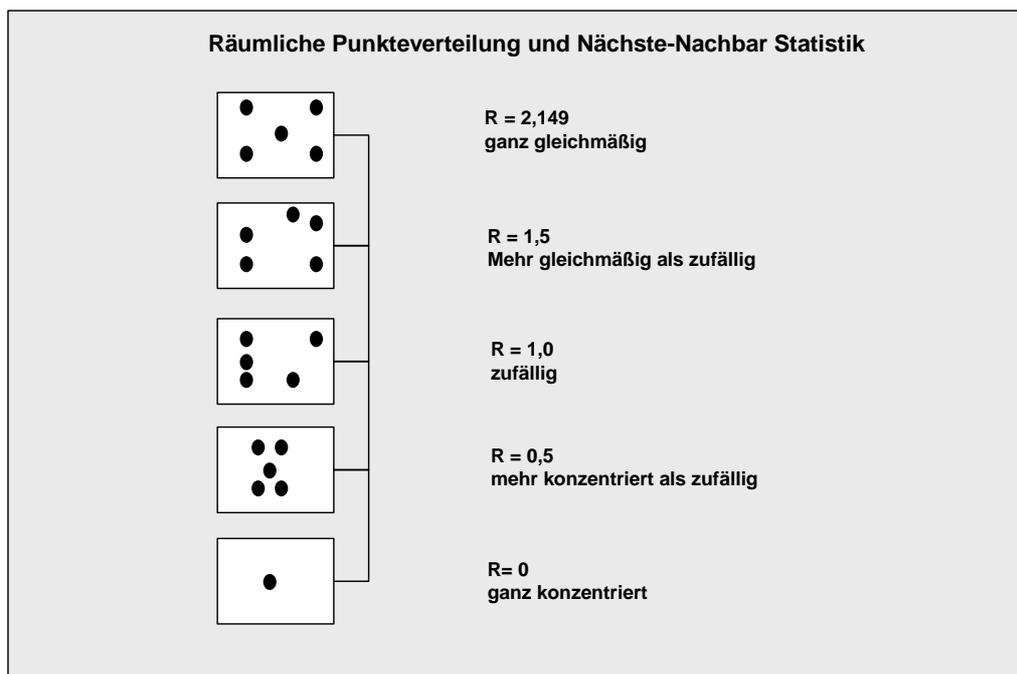


Abb.: Räumliche Punkteverteilung und korrespondierende Nächste-Nachbar Statistik (in Anlehnung an Lorup 2004)

Im Rahmen dieser Studie wurden schwerpunktmäßig Nächste-Nachbarschaftsmethoden angewendet. Deren genaue Erläuterung im Kontext der Studie sowie die Beschreibung der auf der Nächste-Nachbarstatistik aufbauenden Techniken erfolgt in Kap. 4.3.3.2.

Die K-Funktion oder K-Statistik wird auch als Analyse/ Statistik zweiter Ordnung (*second-order analysis*) bezeichnet und geht auf Ripley (1976, 1977) zurück. Eine solche Analyse gibt Auskunft über lokale Konzentrationen im Untersuchungsgebiet. Sind N Punkte in einem Untersuchungsgebiet gemäß den Bedingungen einer räumlichen Zufallsverteilung (CSR) verteilt, dann wird die zu erwartende Anzahl der Punkte innerhalb eines Kreises, der auf das Untersuchungsgebiet gelegt wurde, wie folgt ermittelt:

$$E(\# \text{ unter CSR}) = (N/A)\pi r^2$$

A = Fläche des Untersuchungsgebietes; r = Kreisradius, E = erwartete Anzahl Punkte unter Annahme einer räumlichen Zufallsverteilung

Beträgt etwa die Fläche des Kreises $\frac{1}{4}$ des Untersuchungsgebietes, dann liegen durchschnittlich ungefähr $\frac{1}{4}$ aller Punkte (+/- Stichprobenfehler) einer räumlichen Zufallsverteilung innerhalb dieser Kreisfläche. Ist die durchschnittliche Anzahl der Punkte im Kreis größer, als aufgrund einer räumlichen Zufallsverteilung zu erwarten wäre, ist dies ein Hinweis auf eine Konzentration. Die Distanzen zwischen den Punkten sind in diesem Falle kürzer, als aufgrund einer räumlichen Zufallsverteilung zu erwarten wäre. Ist hingegen die durchschnittliche Anzahl der Punkte kleiner, geht der Trend in Richtung einer gleichmäßigen Verteilung (Lorup 2004). Die K-Funktion betrachtet im Gegensatz zur Nächste-Nachbarstatistik alle Distanzen zwischen Punkten, die innerhalb einer bestimmten Entfernung liegen, nicht nur die des nächsten Nachbarn (Bailey 1998).

Die K-Funktion kann graphisch interpretiert werden, indem ein Kreis mit dem Radius „ ds “ über jeden Punkt „ i “ im Untersuchungsgebiet gelegt wird, wobei der Punkt „ i “ jeweils den Mittelpunkt des Kreises bildet. Nun wird die Anzahl aller Punkte „ j “, welche innerhalb des

selben Kreises liegen, bestimmt. Die Anzahl aller anderen Punkte „ j “ wird für jeden Punkt „ i “ bestimmt. Folgende Formel lässt sich daraus ableiten:

$$K(d_s) = (A/N^2) \sum_i \sum_j I(di_j)$$

Die K-Funktion wird für verschiedene Kreisradien (d_s) berechnet, normalerweise beginnend bei kleinstem Radius bis hin zum größten Radius, der das gesamte Untersuchungsgebiet abdeckt. Trägt man die Werte der K-Funktion auf die y-Achse, die Kreisradien auf die x-Achse eines Koordinatensystems auf, so ist das Resultat eine exponentiell ansteigende Kurve. Um eine lineare Kurve zu erhalten, wird folgende Formel nach Cressie (1991) angewendet.

$$L(d_s) = \sqrt{\frac{K(d_s)}{\pi}} - d_s$$

Diese Formel hat im Gegensatz zu der von Ripley den Vorteil, dass eine strikte räumliche Zufallsverteilung (CSR) eine Parallele zur x-Achse ist (Lorup 2004)

Sowohl die Nächste-Nachbar Statistik als auch die K-Funktion leiden unter dem sogenannten *edge-Effekt*, d.h. eine leichte Verfälschung des Ergebnisses aufgrund der Tatsache, dass Punkte, die außerhalb des Untersuchungsgebietes liegen und die nächsten Nachbarn für Punkte innerhalb des Untersuchungsgebietes sind, unberücksichtigt bleiben.

2.2.2 Punktdichteschätzungen

Punktdichteschätzungen dienen zur Feststellung von Konzentrationen oder Risikobereichen. Generell wird zwischen einfacher Dichteschätzung (*simple density*) und *Kernel*-Dichteschätzung unterschieden. In der einfachen Dichteschätzung werden Punkte, die innerhalb eines vorgegebenen Suchradius liegen, aufsummiert und durch die Größe des Suchradius dividiert.

$$\lambda_p = \frac{\text{Anzahl}[S \in C(p, r)]}{\pi r^2}$$

für $C(p, r)$ = Kreis des Radius r im Zentrum eines Untersuchungsraums p

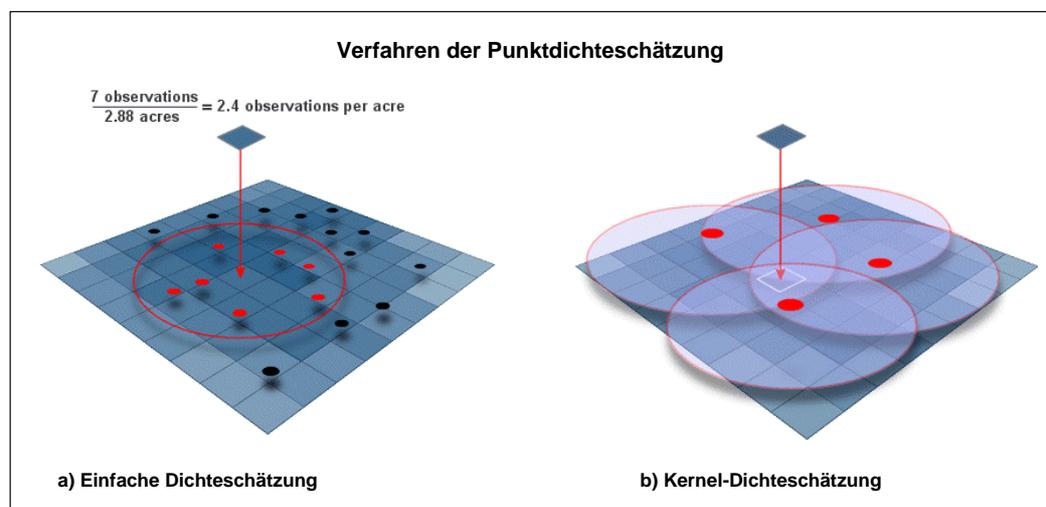


Abb. 3: Verfahren der Punktdichteschätzung (vgl. McCoy 2002)

Das Kernel-Verfahren wurde Ende der 50er Jahre entwickelt (Rosenblatt 1956, Whittle, 1958, Parzen, 1962). Es arbeitet mit einem ähnlichen Verfahren wie dem der einfachen Dichteschätzung, mit der Ausnahme, dass Punkte nahe des Zentrums des Suchradius einer Rasterzelle höher gewichtet werden, als solche nahe dessen äußerer Grenze. Für jeden Punkt wird eine symmetrische Kernel Funktion gerechnet, die eine Normalverteilung darstellt (Abb. 4).

$$g(x_j) = \sum \left\{ [W_i \cdot I_i] \cdot \frac{1}{h^2 \cdot 2\pi} \cdot e^{-\left[\frac{d_{ij}^2}{2h^2}\right]} \right\}$$

d_{ij} = Distanz zwischen einem Ereignisort und irgendeinem Referenzpunkt in der Region

h = die Standardabweichung der Normalverteilung (der Suchradius)

W_i = Gewichtung und I_i = Intensität

Die zugrunde liegende Dichtefunktion wird durch das Aufsummieren der individuellen Kernel-Funktionen geschätzt, um eine kumulative Dichtefunktion zu erhalten (Abb. 4).

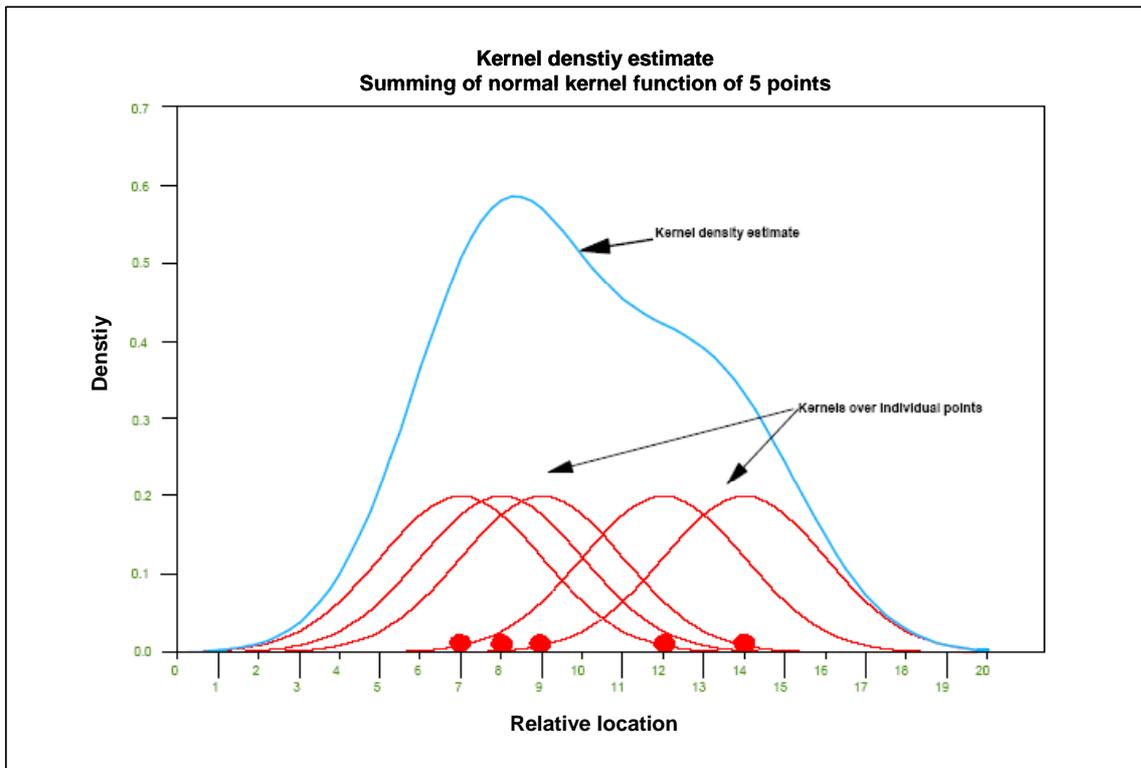


Abb. 4: Prinzip der Kernel density Schätzung (nach Levine 2002)

Das Ergebnis ist eine interpolierte geglättete Rasteroberfläche (Bailey 1998, O'Sullivan 2003, Mitchell 1999). „Here the smoothed value at any point is essentially estimated by a weighted average of all other values, with the weights arising from a probability distribution centred at that point and referred to as the kernel“ (Bailey 1998: 27).

Das Ausmaß der Glättung wird über den Suchradius eingestellt. Es gibt seither eine Vielzahl Kernel Funktionen, die von der Funktion der Normalverteilung (Kelsall & Diggle 1995) - als üblichste Variante - abweichen, wie die Dreiecksfunktion (*triangular function*) (Burt & Barber 1996) und die biquadratische Funktion (*quartic function*) (Bailey and Gatrell 1995). Eine Er-

weiterung dieses Kernel-density-Verfahrens ist die sogenannte angepasste Kernel-density-Schätzung, bei der der Suchradius automatisch im Untersuchungsgebiet variiert wird, um den Effekt einer weiteren Variablen, z.B. Bevölkerungszahl, zu erklären. Kernel-density-Methoden sind mittlerweile in vielen GIS-Software-Produkten implementiert (z.B. ArcGIS-Extension spatial analyst).

2.2.3 Räumliche Autokorrelation

Eine Möglichkeit, räumliche Ähnlichkeiten anhand von Attributdaten zu untersuchen, ist die Methode der räumlichen Autokorrelation. Diese findet häufig Anwendung bei der Analyse von Regionaldaten. Denn während bei geostatistischen Daten die Korrelation als Funktion der Distanz zweier Messpunkte analysiert wird, hat sich für Regionaldaten ein Nachbarschaftskonzept etabliert, welches nur noch berücksichtigt, ob zwei Regionen zueinander benachbart sind oder nicht. Wenn die Werte eines räumlichen Prozesses in benachbarten Regionen miteinander korrelieren, korreliert der Prozess sozusagen mit sich selber. Daher hat sich der Begriff „räumliche Autokorrelation“ eingebürgert (Dreesmann 2004). Diese Autokorrelation kann beispielsweise durch Morans I oder Gearys C gemessen werden.

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

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

N = Anzahl Fälle, z_i = Variablenwert an einem bestimmten Ort i ,

z_j = Variablenwert an einem bestimmten anderen Ort, wobei $i \neq j$,

w_{ij} = ein Gewichtungsfaktor der von der räumlichen Beziehung zwischen dem Ort i und dem Ort j abhängig ist.

Morans I entspricht dem Konzept des gewöhnlichen Korrelationskoeffizienten, allerdings ist der Wertebereich nicht exakt auf das Intervall (-1, 1) begrenzt. Ein Wert um 0 bedeutet, dass keine Korrelation zwischen benachbarten Regionen existiert. Je größer der Wert von Morans I, umso stärker ist die Korrelation, d.h. um so mehr tendieren benachbarte Regionen dazu, ähnliche Werte aufzuweisen. Ein negativer Morans I bedeutet, dass benachbarte Regionen dazu neigen, gegenläufige Werte anzunehmen. Gearys C kann hingegen nur positive Werte annehmen. Wenn keine räumliche Autokorrelation existiert, liegt C bei 1. Je kleiner C wird, umso stärker ist die räumliche Autokorrelation (Dreesmann 2004, Bailey 1998).

Räumliche Autokorrelationen sind mittlerweile in einigen GIS-Softwareprodukten integriert. Darüber hinaus eignen sich statistische Programme, die in der Lage sind, räumliche Daten zu verarbeiten (S-Plus, CrimeStat 2.0, SAS). Der Test auf räumliche Autokorrelation ist allerdings nicht geeignet, die Lokalisation einzelner Cluster in der Karte zu identifizieren (Lawson 2001).

3. Das Anwendungsbeispiel: Legionellen und Legionellose

3.1 Legionellen

Legionellen sind gram-negative, nicht sporenbildende Bakterien der Familie Legionellaceae. Sie entwickeln sich in natürlichen und künstlichen Feuchtzonen, wie Fließgewässer, stehendes Wasser, Abwässer, Thermalwässer, artesische Brunnen, Trinkwasserleitungen, Hähne, Duschköpfe, Kühleinrichtungen mit Wasser, Klimaanlage, Verdampfungsgeräte, Zierbrunnen, Sprudelbäder (z.B. Whirlpools), geschlossene Wasserkreisläufe sowie industrielle Einrichtungen mit Wassersprühern (Tiefenbrunner 2002, Pagano 1994, Pleischel o.A.) (Abb. 5).

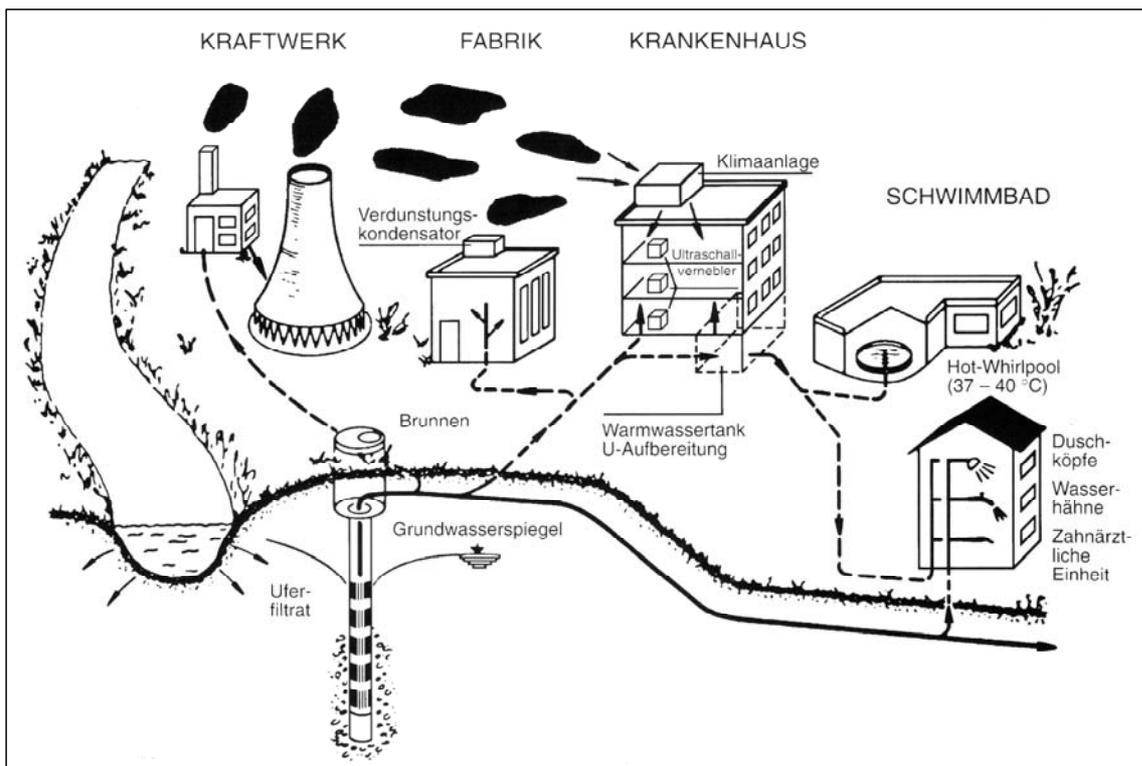


Abb. 5: Legionellenausbreitungsmöglichkeiten in wassertechnischen Systemen (Exner 1997a)

Hier finden sie zwischen 25-50 Grad Celsius optimale Bedingungen für ihre Vermehrung. Die Übertragung der schweren Verlaufsform (Legionellose) erfolgt durch kontaminierte Aerosole. Neben der Menge der aufgenommenen Keime und der Virulenz der Stämme spielt die individuelle Exposition eine wesentliche Rolle. So können die Infektionen im Krankenhaus (nosokomial) oder ambulant (reiseassoziiert sowie im häuslichen oder beruflichen Umfeld) erworben werden (RKI 2003). Besondere Risikopatienten für eine Infektion sind abwehrgeschwächte Menschen mit chronischen Erkrankungen im Allgemeinen und insbesondere chronischen Lungenerkrankungen (RKI 2004).

Bei der Legionellose handelt es sich um eine Atemwegserkrankung, die sich in zwei Formen äußern kann: Das Pontiac-Fieber, welches einem Grippe-ähnlichen Infekt gleicht und hauptsächlich durch Fieber, Husten, Muskelschmerzen charakterisiert ist, sowie die meist schwerer verlaufende Legionärskrankheit, die mit einer Lungenentzündung einhergeht (RKI 2003).

Entdeckt wurden Legionellen im Juli 1976, als in einer Gruppe von Veteranen der American Legion, die sich für ihr jährliches Treffen in Philadelphia versammelt hatten, eine Epidemie

von akuter Lungenentzündung ausbrach. Von den 4.400 Teilnehmern erkrankten 182 schwer, 29 starben (Letalität 16%). Der auslösende Erreger wurde etwa sechs Monate später identifiziert und erhielt den Namen *Legionella pneumophila*. Ursache der Infektion war die Klimaanlage eines Hotels, in dem die Kongressteilnehmer wohnten. Bereits früher wurden Legionellen bei einer Epidemie im Jahre 1968 in Pontiac (Michigan) als verantwortliche Keime entdeckt. Die Infektion, die keine Todesopfer forderte, manifestierte sich mit hohem Fieber, deshalb der Name Pontiac-Fieber.

Im Krankenhaus sind Legionellosen auch ohne zusätzliche Maßnahmen nicht häufiger nachzuweisen als in anderen Gebäuden vergleichbarer Größe, allerdings ist die Anzahl der Personen, die als Zielgruppe der schweren Verlaufsform gelten, höher. Alle Legionellen sind potenziell humanpathogen, wobei *Legionella pneumophila* mit einem Anteil von etwa 90% die bedeutendste Art ist. Neben *Legionella pneumophila* wurden 19 weitere Spezies aufgrund der Isolierung aus klinischem Material als Ursache menschlicher Infektionen beschrieben. Erkrankungen des Menschen treten weltweit sporadisch oder im Rahmen von Ausbrüchen auf.

Die Häufigkeit von Legionella-Infektionen ist nicht genau bekannt. Man geht von einer beträchtlichen Dunkelziffer aus, da einerseits häufig keine oder nur eine unzureichende Abklärung von Lungenentzündungen erfolgt, andererseits viele diagnostizierte Legionellosen nicht gemeldet oder publiziert werden. Epidemiologische Daten aus durchschnittlich 25 europäischen Ländern (*European working group for legionella infections EWGLI*) geben eine ansteigende Infektionsrate von 4,14 Fällen im Jahr 1993 auf 10,1 gemeldete Fälle auf eine Million Einwohner im Jahr 2002 an (Joseph 2004) (Tab. 2).

Tab. 2: Gemeldete Legionelleninfektionen aus 25 europäischen Ländern von 1993 bis 2002

Jahr	Anzahl Fälle	Anzahl berichtender Staaten	Bevölkerung (in Millionen)	Infektionsrate (pro Million der Bevölkerung)
1993	1.242	19	300	4,14
1994	1.161	20	346	3,35
1995	1.255	24	339	3,70
1996	1.563	24	350	4,46
1997	1.360	24	351	3,87
1998	1.442	28	333	4,33
1999	2.136	28	398	5,38
2000	2.156	28	400	5,38
2001	3.470	29	455	7,60
2002	4.696	32	467	10,1

Insgesamt wurden in diesem Zeitraum 20.481 Legionellosen registriert. Die Anzahl der absoluten Fälle stieg von 1.242 im Jahr 1993 auf 4.696 Fälle im Jahr 2002 an, gleichbedeutend eines Zuwachses von 74% in dieser 10-jährigen Periode. Über 10.000 dieser Fälle traten in den Jahren 2000 bis 2002 auf. Zwar ist dieser Anstieg u.a. auf eine erhöhte Meldeaktivität zurückzuführen, die Entwicklung der Infektionsrate weist jedoch auf eine deutliche Zunahme der Legionellosen hin. In den USA werden 25% der jährlichen Legionellosen (8.000 bis 18.000 Fälle) im Krankenhaus erworben (Kool et al. 1999). In Deutschland rechnet man schätzungsweise mit 6.000 bis 10.000 Legionella-Pneumonien pro Jahr. Bei 1 – 5 % der in Krankenhäusern behandelten Lungenentzündungen (Pneumonien) wird eine Legionellose diagnostiziert. Die Aufnahme der Erreger erfolgt in der Regel durch Einatmen erregerhaltigen

Wassers als Aerosol (Duschen, Klimaanlage, Whirlpool etc.). So konnte beispielsweise eine kontaminierte Warmwasserversorgung in Krankenhäusern mit der Übertragung von Legionellen in Verbindung gebracht werden. Die Kalt- und Warmwasserversorgung, d.h. das System zur Verteilung und Abgabe von Wasser für den menschlichen Gebrauch, ist sicherlich die weitverbreitetste Möglichkeit in einem Krankenhaus, Legionellen zu vermehren und in kontaminierten Aerosolen an Patienten, Besucher oder Personal weiterzugeben (Tiefenbrunner 2002). Ein häufiger Infektionsweg ist dabei die Aspiration. Die Inkubationszeit beträgt 2-10 Tage. Meistens sind immungeschwächte Personen betroffen, wobei auch Rauchen und Alkoholgenuss mögliche fördernde Faktoren darstellen. Die ersten Symptome einer Legionella-Pneumonie beginnen mit allgemeinem Krankheitsgefühl und Fieber bis über 40,5°C. Die Pneumonie verläuft oft schwerwiegend, und die Heilung kann langwierig sein. Die Letalität liegt bei ca. 15% (bei Immundefizienz bis 80%). Es gibt mehrere Methoden, die Legionellen labortechnisch zu diagnostizieren. Dazu zählen der kulturelle und der direkte Erregernachweis, der Nachweis des Legionella-Antigens im Urin mittels ELISA und der Legionellen-Antikörpernachweis (mindestens vierfacher Titeranstieg in zwei Proben, z.B. im indirekten Immunfluoreszenztest), wobei dieser oft nur retrospektiv Hinweise auf eine Infektion gibt, da ein Titeranstieg verzögert, oft erst 6 – 8 Wochen nach der Infektion erfolgt. Eine Meldepflicht nach dem § 7 IfSG (Infektionsschutzgesetz) besteht bei direktem oder indirektem Nachweis einer akuten Legionelleninfektion (zum Ganzen Exner 1990, 1997a, 1997b, Yu 1995).

3.2 Legionellose in Deutschland

In Deutschland werden die Daten zu den Legionelleninfektionen vom zuständigen Gesundheitsamt über die Landesstelle an das Robert Koch Institut als übergeordnete Bundesbehörde weitergeleitet. In Abb. 6 sind zunächst die absolute Anzahl der Legionellenerkrankungen gemäß der Referenzdefinition⁴ je Kreis oder kreisfreier Stadt der BRD dargestellt. Insgesamt wurden im Jahr 2003 gemäß der Referenzdefinition 395 Fälle gemeldet. Gegenüber dem Vorjahr (413 Fälle) war die Zahl damit leicht rückläufig. Ein tatsächlicher Erkrankungsrückgang kann aber angesichts der insgesamt geringen Fallzahlen nicht abgeleitet werden. Vielmehr ist nach wie vor von einer hohen Untererfassung auszugehen (RKI 2004). Im Jahresverlauf konnte – wie auch im Vorjahr – eine leichte Zunahme der Erkrankungszahlen in den Sommer- und Herbstmonaten festgestellt werden (Abb. 6):

⁴ Referenzdefinition: einheitliche Kombination von Falldefinitions-kategorien. Für die meisten nach §§6 und 7 des Infektionsschutzgesetzes (IfSG) meldepflichtigen Krankheiten und Erregernachweise umfasst die Referenzdefinition die klinisch begründeten Kategorien „klinisch und durch labordiagnostischen Nachweis bestätigt“ und „klinisch-epidemiologisch bestätigt“ (RKI 2004)

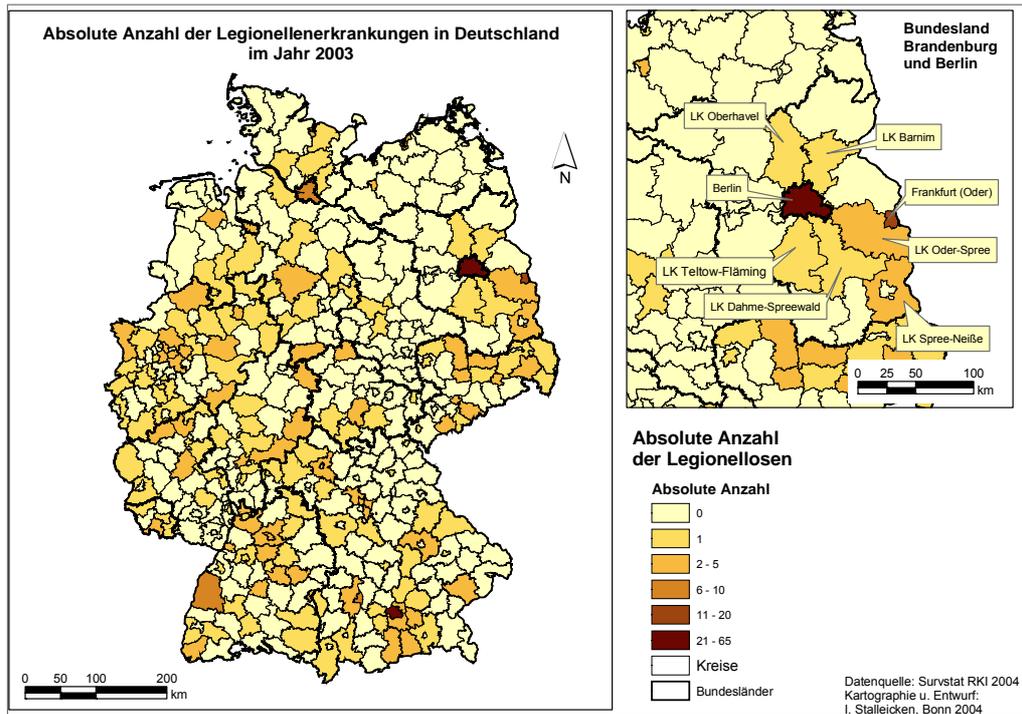


Abb. 6: Absolute Anzahl der Legionellenerkrankungen im Jahr 2003

Neben der absoluten Anzahl der Legionellosen in der BRD können die Inzidenzen pro Kreis oder kreisfreier Stadt berechnet werden (Abb. 7). Bei einer bundesweiten Inzidenz von 0,5 Erkrankungen pro 100.000 Einwohner waren im Land Berlin - wie bereits in den Vorjahren - vergleichsweise deutlich höhere Erkrankungszahlen zu verzeichnen (1,9 pro 100.000 Einwohner) als in den anderen Bundesländern, bei denen die Inzidenzen zwischen 0,1 (Sachsen-Anhalt) und 0,8 (Brandenburg) Erkrankungen pro 100.000 Einwohner lagen (Abb. 7).

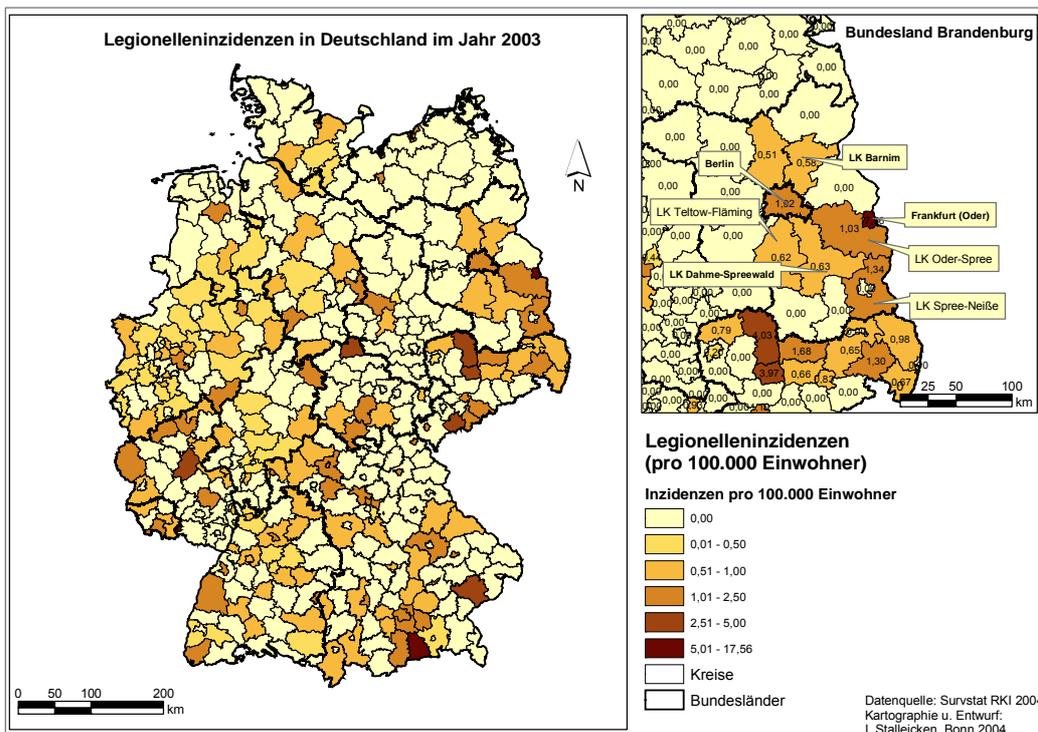


Abb. 7: Legionelleninfektionen in Deutschland – Meldungen des Robert Koch Institutes -

Betrachtet man die Inzidenzen in den Kreisen des Bundeslandes Brandenburg und des Landes Berlin genauer, fällt eine maximale Inzidenz für die Stadt Frankfurt (Oder) auf, die mit 17,56 weit über dem bundesdeutschen Durchschnitt liegt (Abb. 8).

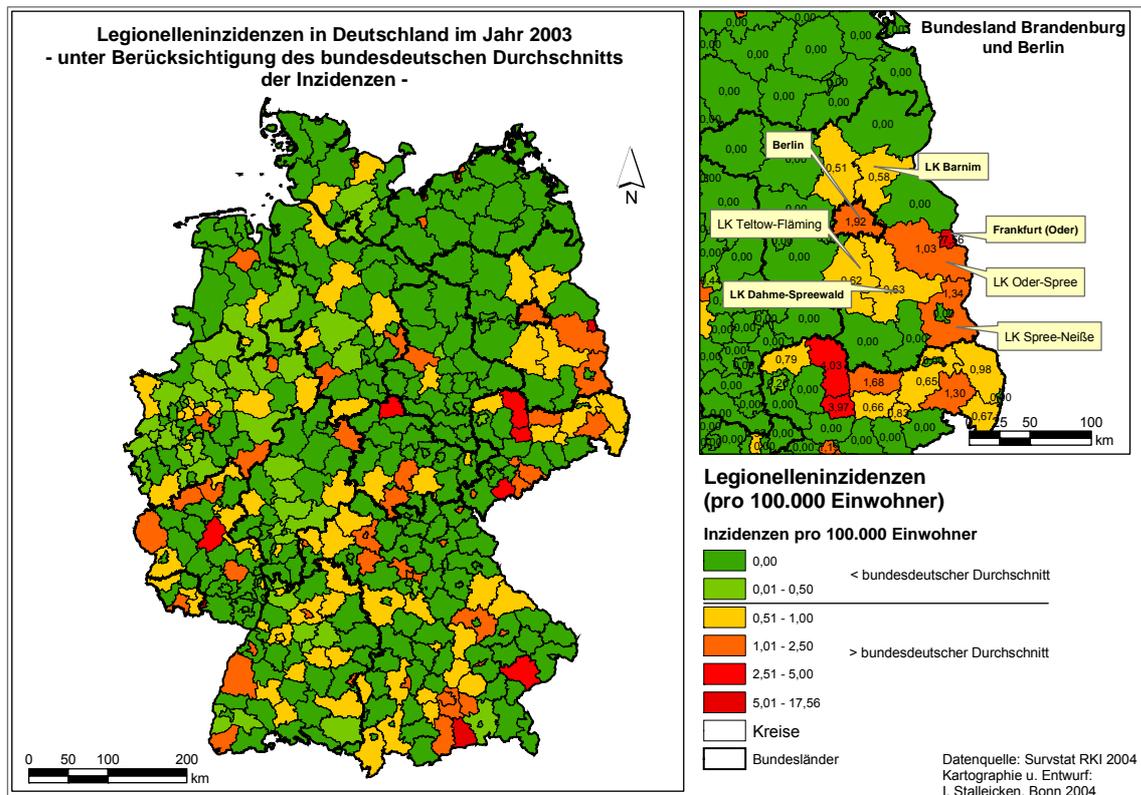


Abb. 8: Legionellenerkrankungen in Deutschland unter Berücksichtigung des bundesdeutschen Durchschnitts

Insgesamt wurden hier 12 Erkrankungen gemeldet, die alle auf zwei Ausbrüche im Klinikum des Landes Brandenburg zurückzuführen sind. In Abb. 8 sind die Kreise, die über dem bundesdeutschen Durchschnitt der Legionellenerkrankungen von 0,5 pro 100.000 Einwohner liegen, in den Farben Orange bis Rot dargestellt.

Es fällt auf, dass in wenigen Kreisen und kreisfreien Städten Deutschlands der bundesdeutsche Durchschnitt überschritten wurde. Eine Häufung der Kreise, in denen dieser bundesdeutsche Durchschnitt überschritten wurde, ist in Teilen der Bundesländer Brandenburg und Sachsen erkennbar, darunter die Landkreise Oder-Spree, Dahme-Spreewald, Spree-Neiße des Bundeslandes Brandenburg sowie die Landkreise Torgau-Oschatz, Döbeln, Riesa-Großenhain etc. des Bundeslandes Sachsen. Vereinzelt Erkrankungshäufungen ergaben sich in Teilen des Bundeslandes Bayern, darunter in den Kreisen Miesbach, Bad Tölz-Wolfratshausen, dem Landkreis München und der Stadt München.

Um eine Veränderung der Legionelleninzidenzen erkennen zu können, wurde die Differenz der Legionelleninzidenzen von 2001 bis 2003 betrachtet (Abb. 9)

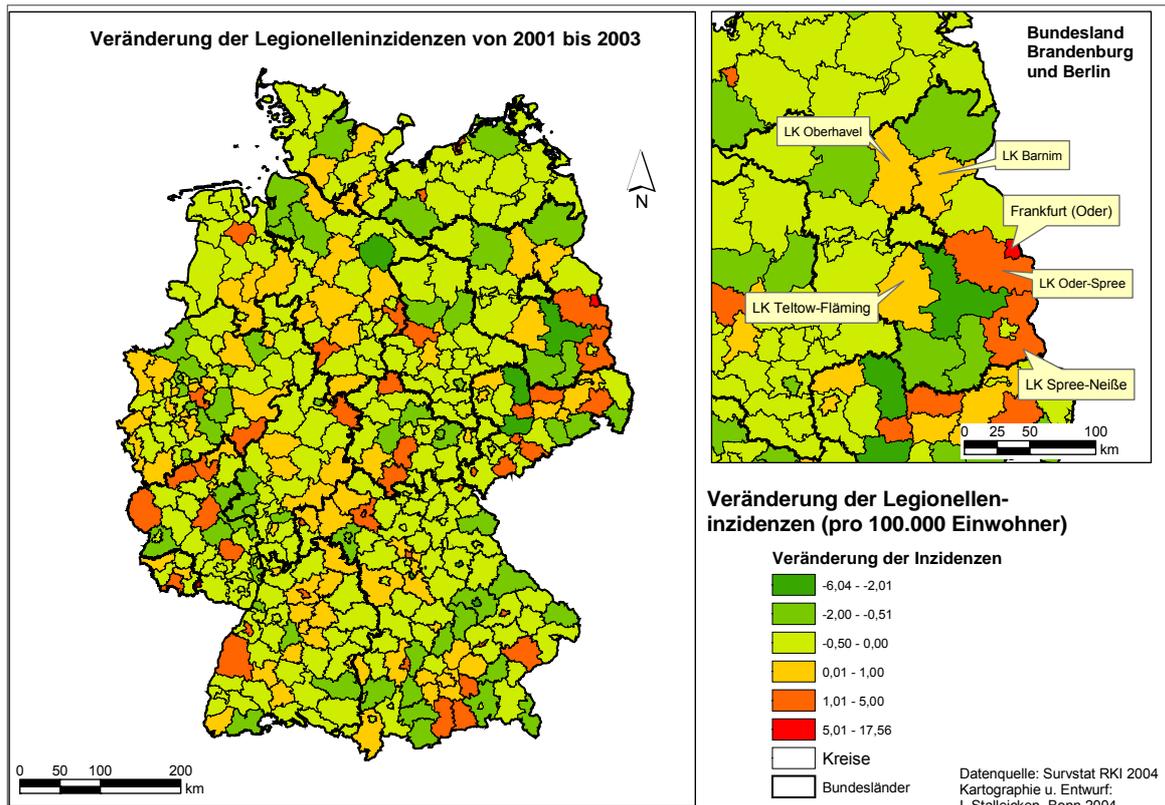


Abb. 9: Veränderung der Legionelleninzidenzen von 2001 bis 2003 in Deutschland

In der Mehrzahl der Kreise der BRD sind die Inzidenzen auf einem gleichen Niveau geblieben. Nur wenige Kreise verfügen über eine erhöhte Inzidenz gegenüber dem Jahr 2001. Dies sind im Land Brandenburg die Kreise Oder-Spree, Spree-Neiße und die Stadt Frankfurt (Oder) sowie im Bundesland Bayern die Landkreise Miesbach, Tölz-Wolfratshausen, Ebersberg und weitere benachbarte Kreise. Die räumliche Darstellung der Veränderung der Inzidenzen lässt darüber hinaus keine regionale Häufung erkennen, was insbesondere in der geringen absoluten Fallzahl begründet liegt.

4. Methodik

4.1 Datenerhebung

Die für eine mögliche Infektion mit Legionellen relevanten Informationen der Patienten wurden mittels Fragebogen zum Zeitpunkt der Aufnahme in das Klinikum erfasst. Dabei handelt es sich um allgemeine Patientendaten (Adresse mit Angabe von Straße und Hausnummer) sowie um Angaben zu möglicherweise relevanten Risikofaktoren vor Aufnahme in das Klinikum. Hierzu zählen Reisen, der Besuch von Schwimmbädern oder sonstigen Sportstätten, der Aufenthalt in der Nähe von Luftbefeuchtern, Rückkühlwerken etc. Außerdem wurde der Serostatus der Patienten bei der Aufnahme erfasst.

Die räumliche und statistische Analyse der erhobenen Daten auf der Basis der Adressen (Hausnummer-genau) von etwa 14.000 Patienten sollte Aufschlüsse über mögliche umwelt-assoziierte Risiken erbringen. Es sollte ermittelt werden, ob Patienten schon vor ihrer Aufnahme mit Legionellen infiziert wurden und wo möglicherweise eine derartige Infektion stattgefunden haben kann. Zudem sollte im Rahmen dieser Studie herausgefunden werden, ob die Patienten in bestimmten Bereichen des Klinikum-Einzugsgebietes einem überdurchschnittlich hohen Infektionsrisiko ausgesetzt sind. Ebenso wichtig war die Frage, ob es auffällige Erkrankungshäufungen im Einzugsgebiet des Klinikums gibt und wenn ja, welche Ursache diese Häufungen haben können.

4.1.1 Serologische Tests

Bei Aufnahme eines jeden Patienten wurden von einem beauftragten Labor serologische Tests auf Legionellen-Antikörper und -Antigen durchgeführt. Die Diagnostik umfasst:

- *Legionella pneumophila* IgM Serum-Antikörper
- *Legionella pneumophila* IgG Serum-Antikörper
- *Legionella pneumophila* Antigen

Die Ergebnisse der positiven Befunde wurden semi-quantitativ angegeben: -, (+), +, ++, +++.

Wenn ein fremder Erreger in den Organismus gelangt, reagiert der Körper als erstes mit der Produktion von Immunglobulin M (IgM). Dessen Produktion sinkt einige Wochen nach Beginn der Infektion ab. Dann sind zum Schutz des Organismus die Immunglobulin G (IgG) Antikörper gebildet worden. IgG ist das eigentliche Immunglobulin des Blutes, zeigt daher auch die höchste Serum-Konzentration aller Immunglobuline. Es hat die stärkste Abwehrwirkung gegen fremde Substanzen im Blut, braucht allerdings relativ lange (ca. 3 Wochen), bis es nach einer Infektion in ausreichenden Mengen gebildet wird.

Nach einer erfolgten Infektion mit Legionellen werden zunächst Legionella-spezifische Antikörper der Klasse IgM später der Klasse IgG gebildet. Nach einer Legionelleninfektion resultiert jedoch keine lebenslange Immunität. Mit zeitlichem Abstand von der Infektion sinken die nachgewiesenen Antikörper-Titer langsam ab. Aus der Höhe des Titers und dem Nachweis der einzelnen Klassen kann daher näherungsweise eine Aussage zum Infektionszeitpunkt abgeleitet werden. Der Serostatus der Patienten ist in diesem Sinne detaillierter ausgewertet worden (-,(+), +, ++, +++).

Der Antigen-Nachweis (AG) aus Urin sollte im Gegensatz zur retrospektiven Aussage der Serologie das Vorliegen einer aktuellen *Legionella-pneumophila*- Infektion der Serogruppe 1 zum Aufnahmezeitpunkt untersuchen.

4.1.2 Der Fragebogen

Mit Hilfe eines Fragebogens wurden im Klinikum seit August 2003 folgende Informationen erfasst, die für die Infektion mit Legionellen von Bedeutung sein können:

1. Reisen in den letzten 12 Tagen vor Aufnahme in das Krankenhaus mit Übernachtung in auswärtigen Einrichtungen (z.B. Hotel)
2. Regelmäßiger Besuch von Schwimmbädern oder sonstigen Sportstätten
3. Verwendung von Luftbefeuchtern im häuslichen Bereich
4. Existenz von Rückkühlwerken mit Aerosolbildung im Wohn- und Arbeitsumfeld
5. Auftreten von Atemwegsinfektionen im letzten Jahr (gehäufte grippale Infekte oder Lungenentzündung)
6. Gehäuft auftretende Temperaturerhöhung im letzten halben Jahr
7. Datum des letzten Krankenhausaufenthaltes

Die Befragung dauerte 12 Monate. Es wurden in diesem Zeitraum mehr als 14.000 Patienten befragt. Nicht alle eingewiesenen Patienten waren in der Lage einen Fragebogen auszufüllen (aufgrund von Bewusstlosigkeit etc.).

4.2 Datenerfassung

Die Daten der Fragebögen wurden in einer Datenbank (MS Access®) erfasst. Hierzu wurde ein Eingabeformular erstellt, um Eingabefehler weitestgehend zu reduzieren (Abb. 10).

The screenshot shows a web-based form titled "Patientenbefragung zur Legionellen-Studie". It is divided into several sections:

- Patientennummer:** 540176.3
- Aufnahmedatum:** 20030406
- Name:** Mustermann
- Vorname:** Susanne
- Infektionsserologie:**
 - IgM: negativ
 - IgG: ++
 - Antigen: positiv
 - Stations-ID: CC2
 - Letzter Krankenhausaufenthalt: Jahr 2003, Monat 7
- Fragebogen:**
 - 1. Reisen in den letzten 12 Tagen vor Aufnahme in das Krankenhaus mit Übernachtung in anderen Einrichtungen (z.B. Hotel)
 - 2. Regelmäßiger Besuch von Schwimmbädern oder Sportstätten
 - 3. Werden Luftbefeuchter im häuslichen Bereich verwendet?
 - 4. Sind Rückkühlwerke mit Aerosolbildung im Wohn- oder Arbeitsumfeld bekannt?
 - 5. Atemwegsinfektionen im letzten Jahr (gehäufte grippale Infekte oder Lungenentzündung)
 - 6. Gehäuft auftretende Temperaturerhöhungen im letzten halben Jahr?
- Bemerkungen:** (Empty text area)

Abb. 10: Eingabeformular der Datenbank „Legionellen FO“

Anhand einer eindeutigen Verknüpfung der Patientennummer zur Tabelle „Patientenadressen“ konnten Dateninkonsistenzen und Eingabefehler identifiziert und eliminiert werden. Über Abfragen konnten spezifische Informationen ermittelt werden, die in beiden Tabellen vorlagen. Die Abfragen wurden aus der Datenbank exportiert und zur weiteren Analyse in ein Tabellenkalkulations- (MS Excel 97[®]) und ein Statistikprogramm (SPSS 12[®]) überführt.

4.3 Datenanalyse

Die anschließende Datenanalyse erfolgte auf drei Wegen (Abb. 11):

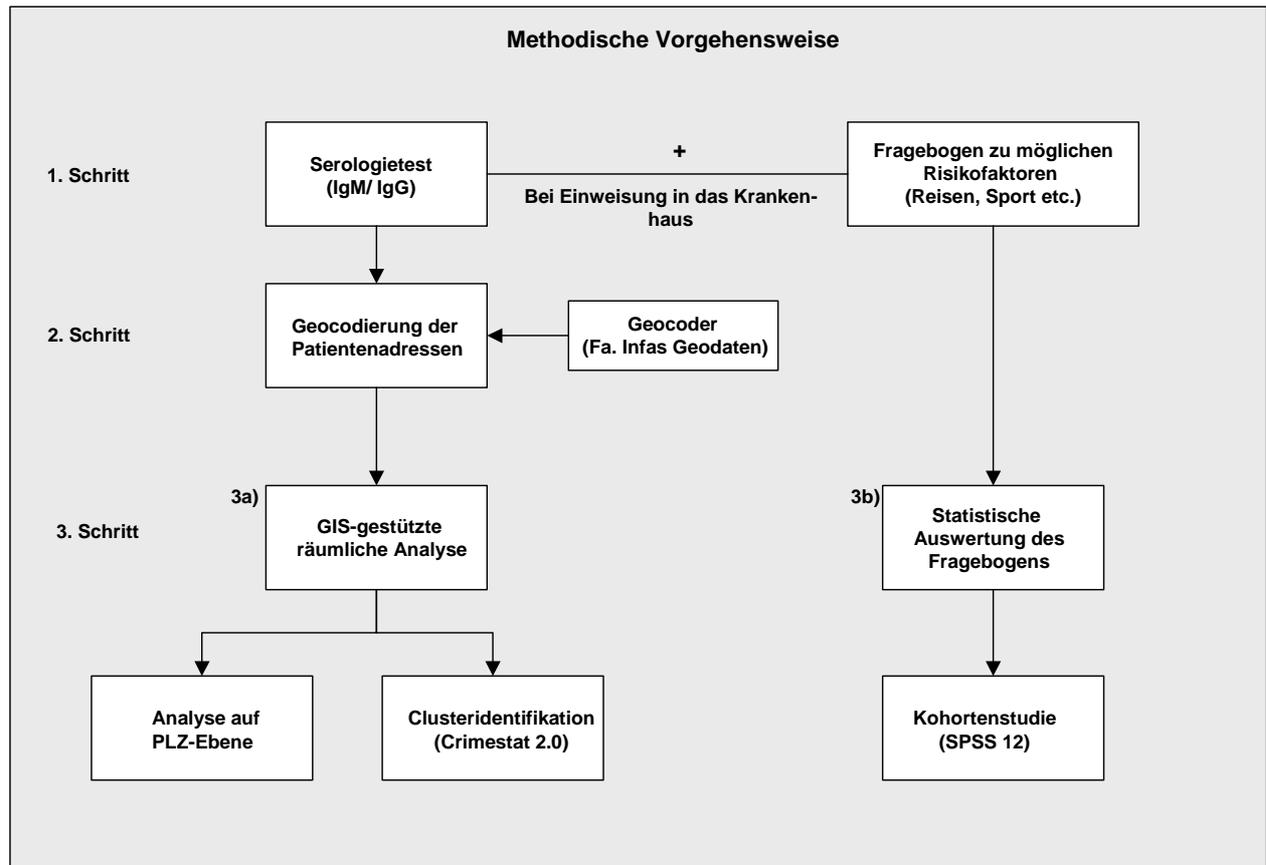


Abb. 11: Methodische Vorgehensweise

1. Deskriptive statistische Analyse der allgemeinen Patientendaten im Untersuchungszeitraum, der Ergebnisse des serologischen Tests sowie der Ergebnisse der Patientenbefragung (siehe 3b);
2. Analytisch-epidemiologische Untersuchung (Kohortenstudie) anhand der Ergebnisse der serologischen Tests und der Befragung der Patienten mit SPSS 12[®] sowie Epi Info 2000[®];
3. Räumlich-statistische Analyse der geocodierten Patientenadressen mit Hilfe Geographische Informationssysteme, d.h. eine Analyse auf der Ebene der Postleitzahlregionen sowie die Clusteridentifikation unter Verwendung der Koordinaten von Straße und Hausnummer (ArcGIS 8[®] sowie Crimestat 2.0[®]) (siehe 3a).

4.3.1 Deskriptive statistische Analyse

Die deskriptive statistische Datenanalyse erfolgte in SPSS 12 auf Grundlage der Abfragen der Access-Datenbank. Hier wurden zum einen allgemeine Patientendaten, z.B. die Anzahl der Patienten über den Untersuchungszeitraum etc., zum anderen die Ergebnisse der Befragung der Patienten unter Berücksichtigung der Ergebnisse der serologischen Tests analysiert (Abb. 12)

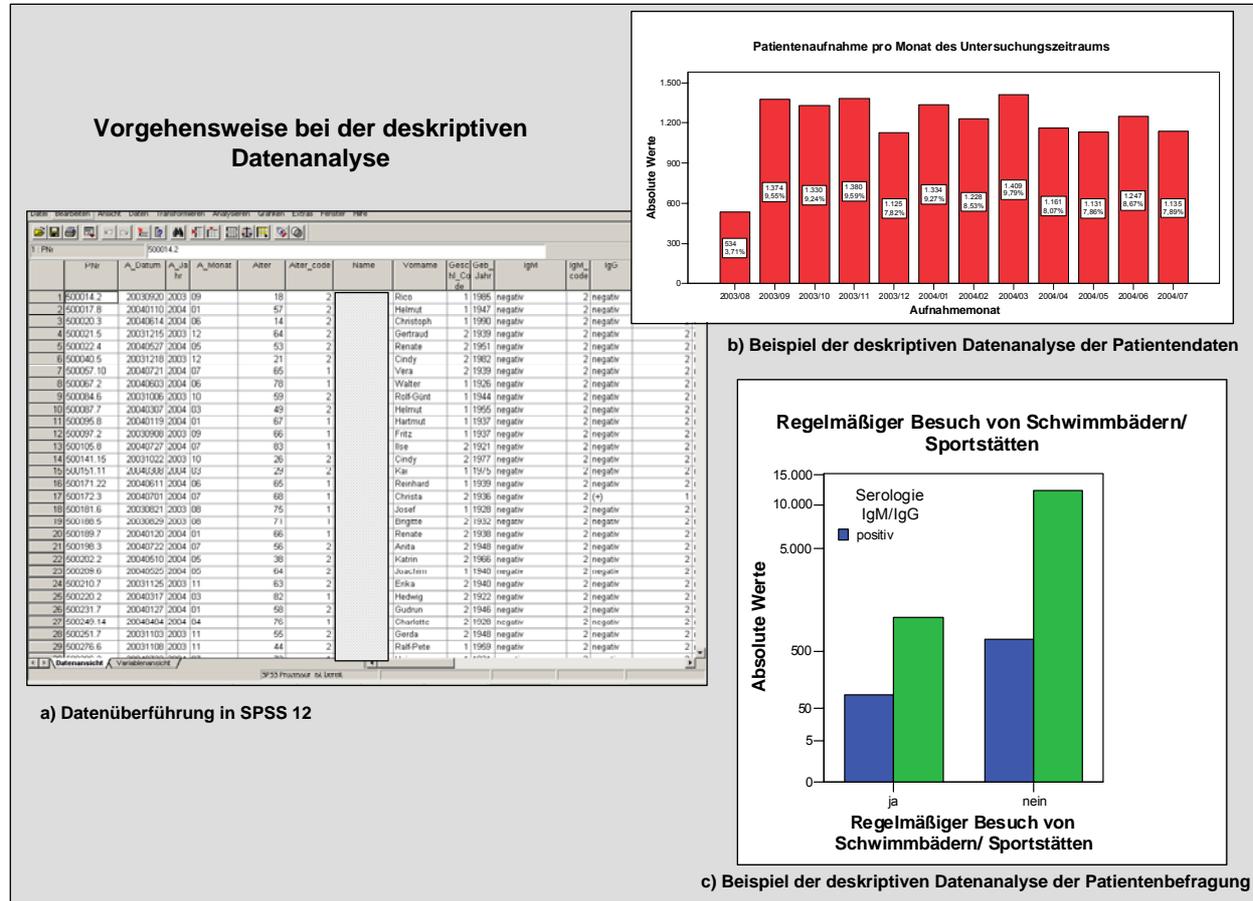


Abb. 12: Vorgehensweise bei der deskriptiven Datenanalyse

4.3.2 Analytisch-epidemiologische Kohortenstudie

Im nächsten Schritt wurde geprüft, ob der Anteil IgM- und IgG-positiver Patienten in Abhängigkeit von der Exposition gegenüber verschiedenen Risikofaktoren variiert mit dem Ziel, die Anzahl möglicher Risikofaktoren im Rahmen der räumlich-statistischen Analyse zu begrenzen. Dazu diente im Rahmen einer Kohortenstudie die Ermittlung des relativen Risikos (Verhältnis des Risikos bei den Exponierten zum Risiko bei Nicht-Exponierten).⁵

⁵ Das relative Risiko ist ein Parameter, der Werte zwischen 0 und $+\infty$ annehmen kann. Wenn $R = 1$ gilt, ist das Erkrankungsrisiko bei Exposition und Nicht-Exposition gleich groß, und man erkennt keinen Einfluss der Exposition auf die Erkrankungswahrscheinlichkeit. Ist $R > 1$, so ist das Erkrankungsrisiko unter Exposition größer als unter Nicht-Exposition und man erkennt einen schädigenden Einfluss. Im Falle von $R < 1$, wäre die Situation eines schützenden Einflusses anzunehmen (Kreienbock 1995), der ggf. für Präventionsstrategien genutzt werden kann.

Die Berechnung des relativen Risikos wurde sowohl in SPSS 12[®] als auch in Epi Info 2000[®] durchgeführt; die Ergebnisse wurden zur Absicherung miteinander verglichen. Die Risikovariablen (hier Serologie positiv/negativ), die angibt, ob ein bestimmtes Ereignis eintritt oder nicht, wurde jeweils in Abhängigkeit von einer unabhängigen (ursächlichen) und ebenfalls dichotomen Variablen (hier z.B. Reiseaktivitäten, regelmäßiger Besuch von Schwimmbädern/ Sportanlagen, Verwendung von Luftbefeuchtern etc.) untersucht (Abb. 13).

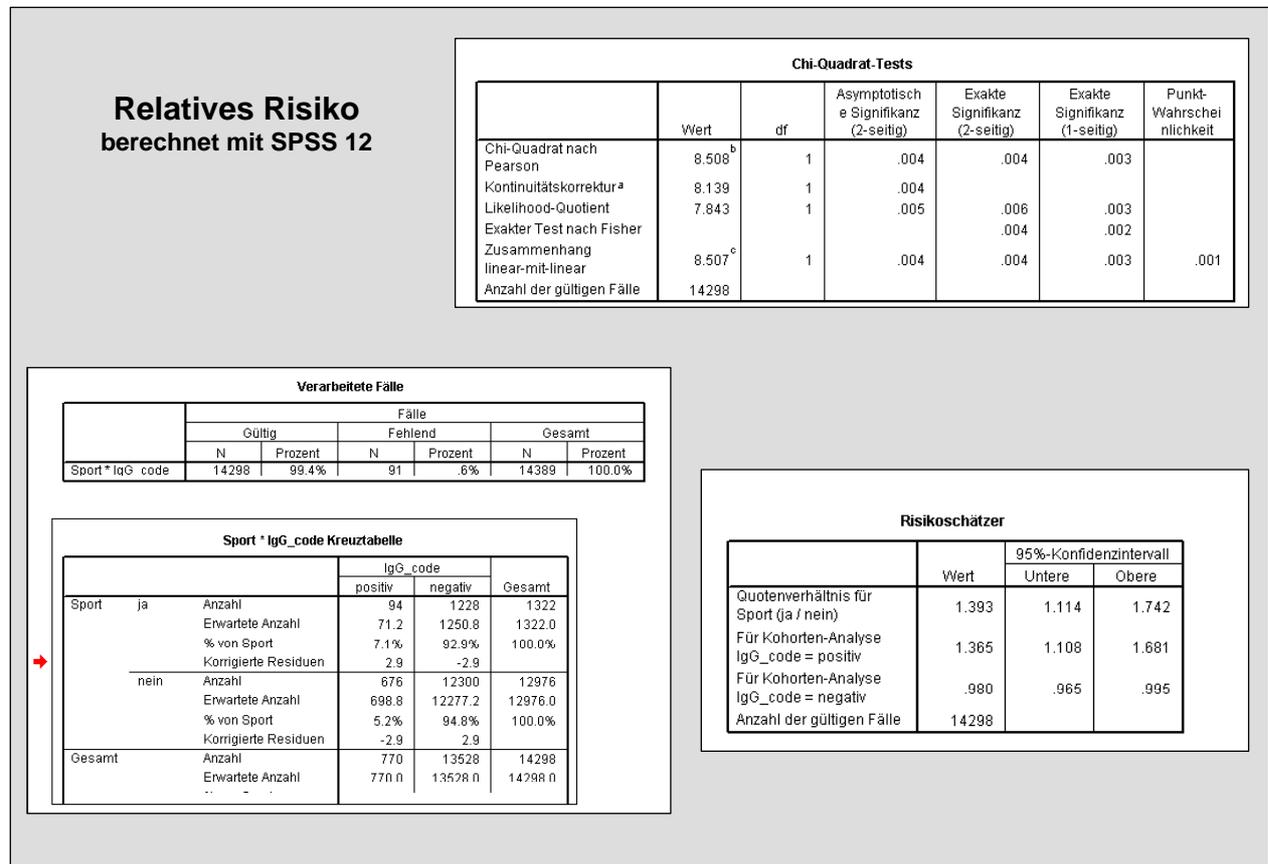


Abb. 13: Risikoanalyse in SPSS 12

Kohortenstudien untersuchen eine festgelegte Fallgruppe über einen gewissen Zeitraum hinweg und stellen fest, bei welchen Probanden das Ereignis eintritt und bei welchen nicht und ob sich das Risiko des Eintretens zwischen den Kategorien einer unabhängigen Variablen unterscheidet. (Bühl 2004).

Bei Kohortenstudien wird für die beiden Kategorien der unabhängigen Variablen die Inzidenzrate/ Prävalenzrate bestimmt. Der Quotient aus den beiden Raten wird als relatives Risiko berechnet.

SPSS 12[®] berechnet das relative Risiko, das 95%-Konfidenzintervall sowie die Signifikanz (Test nach Fisher, zweiseitig signifikant bei $p < 0,05$).

Neben Expositionsangaben anhand der Informationen des Fragebogens wurden weitere Angaben als Risikofaktoren angenommen: das Alter, das Geschlecht sowie die Diagnose des Patienten (ICD 10 Code).

4.3.3 Räumlich-statistische Analyse

In der Epidemiologie werden in den letzten Jahren vermehrt räumlich-statistische Analysemethoden angewendet (Kistemann et al. 2002). Diese Verfahren werden in der Regel unterstützt durch den Einsatz Geographischer Informationssysteme (GIS). Auf diese Weise konnte eine Vielzahl von Erkrankungen und Ausbrüchen retrospektiv analysiert werden und zur Aufklärung der Erkrankungsursache ein erheblicher Beitrag geleistet werden. Aufgrund der Epidemiologie der Legionellose, insbesondere der Vermehrung und Verbreitung ihres Erregers in der technischen Umwelt in Abhängigkeit von spezifischen Bedingungen, sind räumliche Analysemethoden hier von besonderer Bedeutung.

Insbesondere im Bereich der umwelt-assoziierten Erkrankungen, zu denen auch die Legionellosen zu rechnen sind, ist die räumlich-statistische Analyse von Risikofaktoren, die mit der Krankheit in Verbindung gebracht werden können, eine Möglichkeit, Fragen des Infektionsweges sowie des Infektionsrisikos zu beantworten. In der Regel wird der Wohnort der Probanden als räumlicher Bezugspunkt gewählt. Hierfür wurden zwei Wege ausgewählt:

1. die flächenhafte Analyse auf Basis der Postleitzahlenbezirke
2. die punktbasierte Analyse auf der Basis der Geocodierung von Patientenadressen

In Abb. 14 ist die methodische Vorgehensweise der räumlich-statistischen Analyse der Risikofaktoren dargestellt:

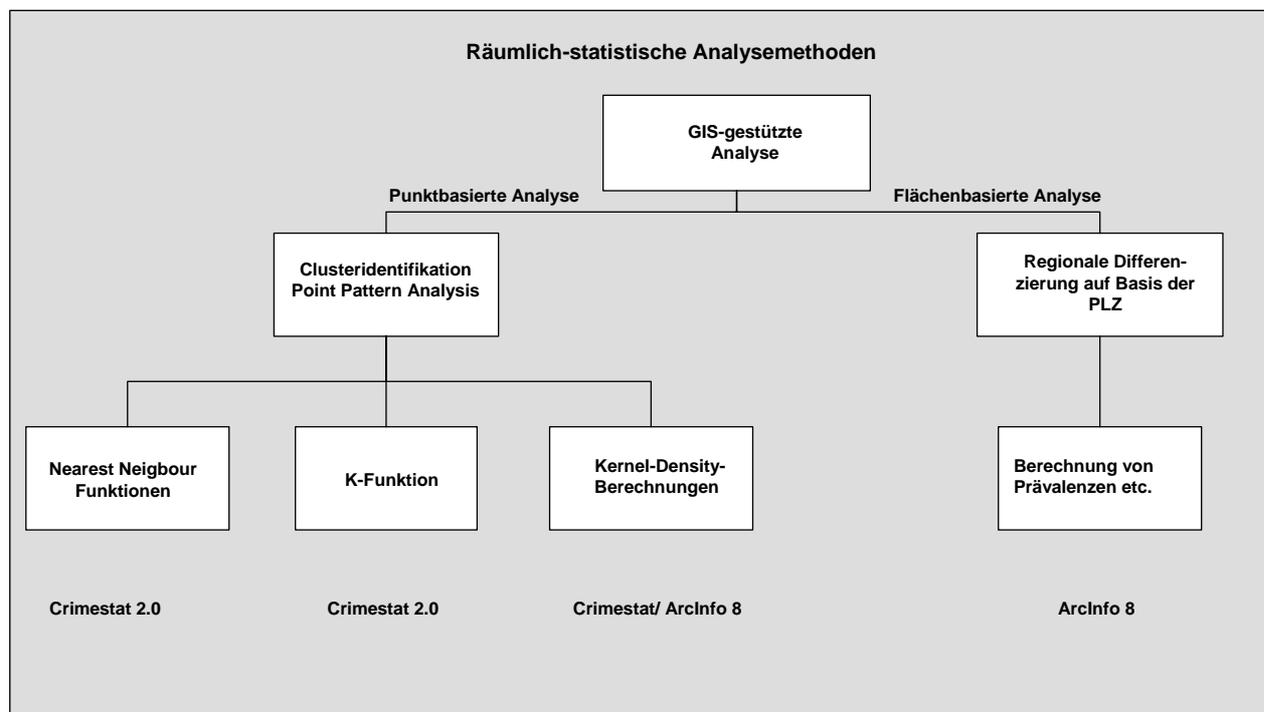


Abb. 14: Anwendung räumlicher Analysemethoden

Bei der regionalen räumlichen Analyse auf Basis der Postleitzahlbezirke wurden Prävalenzen, sowie Über- und Unterschreitungen des Erwartungswertes (arithmetischer Mittelwert der gesamten Untersuchungsgruppe) je Postleitzahlenbezirk ermittelt, um einen ersten Eindruck über eventuelle Risikogebiete zu erlangen. Die punktbasierte Analyse hingegen erfolgte insbesondere mittels der Point Pattern Analysis-Methoden. Diese wurden maßgeblich mit der Software Crimestat 2.0 durchgeführt.

4.3.3.1 Flächenhafte Analysemethoden

Diese Form der räumlichen Analyse erfolgte über eine Aggregation der Punktdaten auf Flächeneinheiten. Hierzu wurde die positive Seroprävalenz je Flächeneinheit (hier Postleitzahlbezirke) berechnet. Ein vergleichbares Verfahren wurde in der Vergangenheit bereits im Zusammenhang mit der Interpolation (Kriging) von akuten respiratorischen Erkrankungen (ARE) als Indikator von Influenza herangezogen (Stalleicken et al. 2004).

Um nun festzustellen, ob die benachbarten auf Postleitzahlebene aggregierten Werte (Prävalenzen pro 100 Patienten) räumlich miteinander korrelieren, wurde die räumliche Autokorrelation nach Moran und Geary mit der Software Crimestat 2.0 berechnet (Abb. 15).

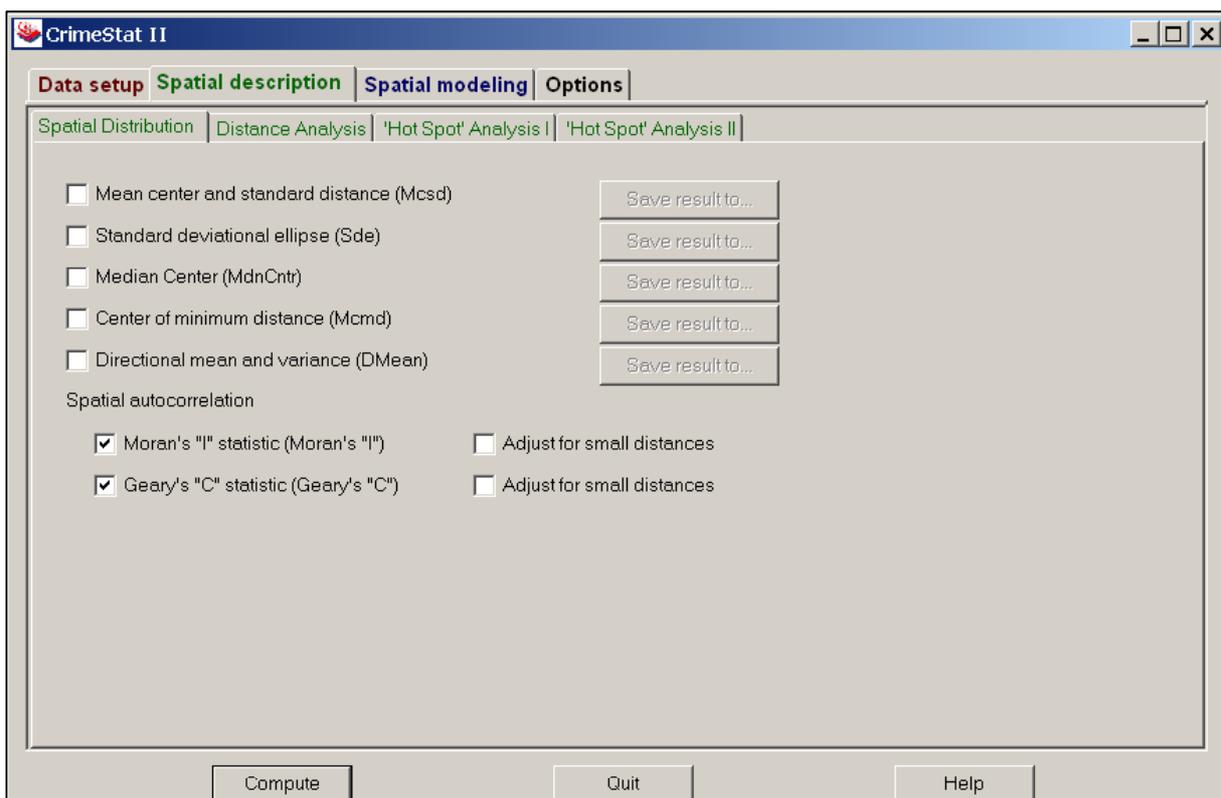


Abb. 15: Berechnung der räumlichen Autokorrelation mit Crimestat 2.0

Dazu wurden zunächst in ArcGIS sogenannte Zentroide pro Postleitzahlbezirk erstellt und diesen Zentroiden die entsprechenden Prävalenzen zugewiesen. Die Moran's Statistik schwankt zwischen -1 (sehr hohe negative räumliche Autokorrelation) und $+1$ (sehr hohe räumliche Autokorrelation). Der zu erwartende Moran's I Wert einer räumlichen Unabhängigkeit (spatial random expected I) hat einen Wert, der negativ ist, aber nahe bei 0 liegt. Liegt der berechnete Moran's I über dem zu erwartenden Moran's I Wert, dann liegt eine positive räumliche Autokorrelation vor (und umgekehrt). Ob diese positive oder negative räumliche Autokorrelation nun statistisch signifikant ist, wird mit einem statistischen Test ermittelt, indem man die von Crimestat errechnete z-Teststatistik (normality significance z) mit dem kritischen z-Wert aus einer statistischen Tafel vergleicht.

Die Geary's Statistik schwankt zwischen 0 (sehr hohe räumliche Autokorrelation) und 2 (sehr hohe negative räumliche Autokorrelation). Der zu erwartende Geary's Wert ist immer 1. Werte der Geary's Statistik zwischen 0 und 1 bedeuten eine positive räumliche Autokorrelation

und Werte zwischen 1 und 2 eine negative räumliche Autokorrelation. Auch hier kann analog zu Moran's I ein Signifikanztest durchgeführt werden (Lorup Unigis).

Zur Beurteilung, ob in Kreisen überdurchschnittlich viele bzw. unterdurchschnittlich wenige Fallzahlen im Untersuchungszeitraum auftraten, wird der Lokationsquotient (LQ) errechnet. Dieser gibt das Verhältnis der beobachteten Fälle der Raumeinheiten zu den erwarteten Fällen an und ist eines der am häufigsten angewendeten Verfahren zu Identifizierung von Extremwerten (Gesler 1986). Die erwarteten Fälle werden über den Mittelwert der Prävalenzen im gesamten Untersuchungszeitraum ermittelt. Die Berechnung des Lokationsquotienten (LQ) basiert auf folgender Formel:

$$LQ = \frac{O_i}{P_i} : \frac{O}{P}$$

wobei O_i die beobachteten Fälle der i -ten Raumeinheit sind, P_i die Bevölkerung der i -ten Raumeinheit, O die beobachteten Fälle der Gesamtregion und P die Bevölkerung der Gesamtregion. Ist LQ größer 1 dann liegt eine überdurchschnittlich hohe Zahl serologisch positiver Patienten in dieser Raumeinheit vor. Bei einem LQ von 1,1 in einem Kreis etwa liegt die Prävalenz 10% über der mittleren Prävalenz des Untersuchungsraumes. Liegt der ermittelte Wert unter 1, dann wurden weniger positive Patienten als im Durchschnitt ermittelt. Die räumliche Darstellung des LQ auf Kreisebene lässt erkennen, ob die serologisch positiven Patienten gleichmäßig über den Untersuchungsraum verteilt sind oder nicht. Ebenso wird ein Hinweis auf eventuelle Cluster gegeben.

Ein Nachteil dieser Methode ist, dass geringe absolute Patientenzahlen pro Postleitzahlbezirk zu Verzerrungen führen. Daher wurden nur solche Postleitzahlenbezirke für die Berechnung selektiert, in denen mehr als 20 Patienten pro Postleitzahlbezirk wohnten.

4.3.3.2 Point Pattern Analysis Methoden

Ziel der räumlich-statistischen Analyse ist die Untersuchung der Verteilung der Patienten mit unterschiedlichem Serostatus sowie die Identifikation von räumlichen Erkrankungsklustern und deren Nähe zu relevanten Risikofaktoren in der Umwelt, die für die Verbreitung von Legionellen von Bedeutung sind. Aus dem Bereich der GIS-gestützten räumlichen Analyse wurden vornehmlich Verfahren der Point Pattern Analysis (PPA-Methoden) angewendet (Kap. 2.2), und zwar im einzelnen:

1. zur Bestimmung der *räumlichen Punktverteilung* mit dem Schwerpunkt auf die Nächste Nachbar Statistik und der K-Statistik (Kap.2.2.1)
2. zur Bestimmung der *räumlichen Punktdichte*, u.a. nach positivem und negativem Serostatus (Erstellung von Punktdichtekarten mit Hilfe der Kernmethode u.a.) (Kap. 2.2.2).
3. zur *Identifikation räumlicher Cluster*, d.h. signifikant höherer Konzentrationen von Punkten in Teilbereichen des Untersuchungsgebietes.

Voraussetzung einer solchen räumlich-statistischen Analyse ist die so genannte Geocodierung der Patientendaten, wobei die Wohnadresse eines jeden Patienten zugrunde gelegt wurde und anhand des Straßennamens und der Hausnummer in das Geographische Infor-

mationssystem implementiert wurde (Stand-alone Software der Firma Infas Geodaten). Problematisch in diesem Zusammenhang war die erst Ende 2002 abgeschlossene Gebietsreform im Land Brandenburg. Diese führte bei der Geocodierung zu einem erhöhten Arbeitsaufwand, da viele Ortsnamen und Postleitzahlen nicht mehr übereinstimmten. Jeder Datensatz durchlief während des Geocodierungsprozesses eine Qualitätsprüfung. So wurde die Güte der Geocodierung bewertet, d.h. es besteht für jeden Datensatz eine Information darüber, ob der Datensatz korrekt ist, korrigiert wurde oder eventuell Fehler enthält, die nachträglich überprüft werden mussten:

- Qualität A: haus-, straßenabschnitts-, stimmbezirks- und PLZ-genau
- Qualität B: straßenabschnitts-, stimmbezirks- und PLZ-genau
- Qualität C: stimmbezirks- und PLZ-genau

Fehlerhafte Geocodierungen konnten durch mögliche Korrekturen während des Geocodierungsprozesses weitestgehend vermieden werden, da die nicht eindeutigen Adressen manuell nachbearbeitet werden können.

Zur **Bestimmung der räumlichen Punkteverteilung (1)** wurde die *Nächste-Nachbar Statistik* (*nearest neighbour statistic*) angewendet (Kap. 2.2.1). Bei dieser Methode werden die beobachteten Distanzen zwischen jedem Punkt und seinem nächsten Nachbarn im Untersuchungsgebiet berechnet und mit den zu erwartenden Distanzen einer räumlichen Zufallsverteilung verglichen. Die Berechnung erfolgte mit der Software Crimestat 2.0.

Eine weitere Möglichkeit zur Bestimmung der räumlichen Punkteverteilung ist die K-Funktion, die bereits ebenfalls in Kap. 2.2.1 erläutert wurde. Die Berechnung erfolgte ebenfalls mit der Software Crimestat 2.0 (Abb. 16).

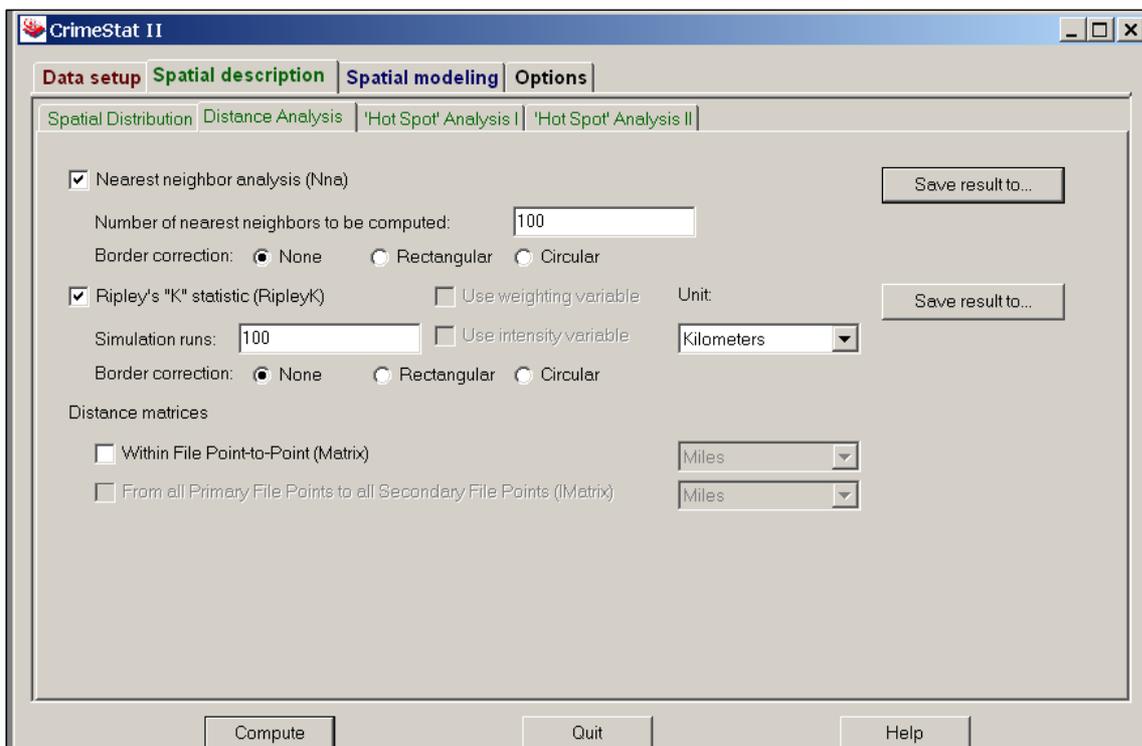


Abb. 16: Berechnung der räumlichen Punkteverteilung mit Crimestat 2.0

Bei der Berechnung der räumlichen Punkteverteilung ist zu beachten, dass die Größe des betrachteten Untersuchungsgebietes angegeben werden muss. Die Nächste-Nachbar Statistik gibt folgende zehn Parameter aus (Levine 2002):

1. die Anzahl der in die Berechnung einfließenden Punkte
2. die mittlere Nächste-Nachbar Distanz
3. die Standardabweichung der mittleren Nächste-Nachbar Distanz
4. die kürzeste Distanz
5. die längste Distanz
6. die mittlere zufällige Distanz
7. die mittlere beobachtete Distanz
8. der Nächste-Nachbar Index
9. der Standardfehler des Nächste-Nachbar Index
10. der Signifikanztest des Nächste-Nachbar Index

Die Ergebnisse können als dbf-File gespeichert werden.

Ebenso ist es möglich, den in Kap. 2.2.1 beschriebenen Edge-Effekt bei den Berechnungen zu berücksichtigen. Dies geschieht mit Hilfe der Einstellungen „rectangular“ oder „circle“, je nach Ausdehnung des Untersuchungsgebietes (Levine 2002).

Ripley's K-Statistik wurde bereits in Kap. 2.2.1 erläutert. Crimestat berechnet die L-Statistik, die der K-Statistik entspricht (Wurzelfunktion von K).

Anhand einzelner Simulationen, deren Anzahl manuell bestimmt werden kann, werden konzentrierte oder gleichmäßige Muster identifiziert. Für jede zufällige Simulation wird die L-Statistik eines jeden Intervalls berechnet. Die höchsten und niedrigsten L-Werte werden für jedes Intervall ausgewählt (bezeichnet als engl.: envelope) und mit den Werten der beobachteten L-Statistik verglichen. So können Unterschiede und Abweichungen beobachtet werden. Um festzustellen, ob das beobachtete Punktmuster einer konzentrierten, zufälligen oder gleichmäßigen Verteilung entspricht, wurde für jede Distanz der Wert für $L(t)$ mit $L(t)_{\min}$ und $L(t)_{\max}$ (Konfidenzband) verglichen. Liegt $L(t)$ oberhalb des Konfidenzbandes, dann entspricht das beobachtete Punktmuster einer konzentrierten Verteilung und umgekehrt (Lorup Unigis 2004).

Die **Bestimmung der räumlichen Punktdichte (2)** erfolgt anhand der Kernel density Interpolation. Diese wurde bereits in Kap. 2.2.2 eingehend beschrieben. Die Interpolation wurde in ArcInfo 8 durchgeführt. Ebenso wurden Kernel density Berechnungen im Rahmen der Clusteridentifikation benötigt. In ArcInfo 8 wurde mit Hilfe der Spatial Analyst Extension unter Variation der Parametereingaben für die IgG sowie IgM positiven Patienten Punktdichtekarten erstellt (Abb. 17)

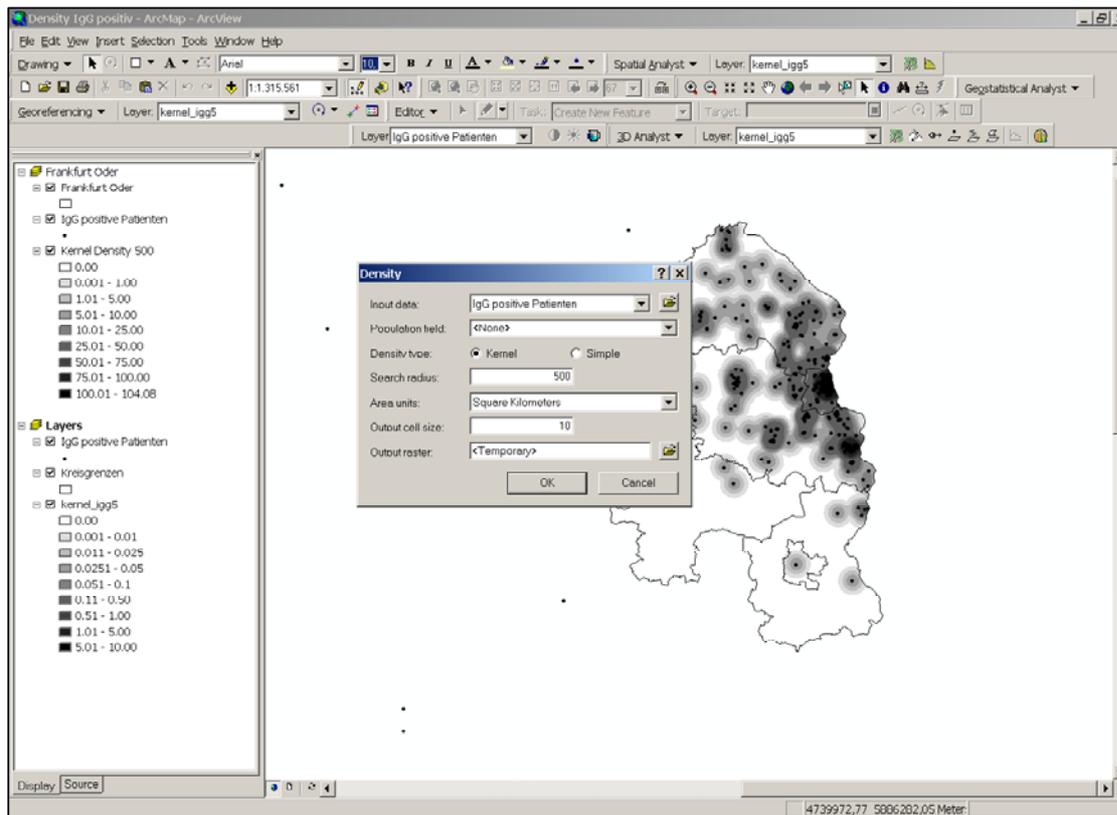


Abb. 17: Beispiel der Kernel Density Berechnung in ArcGIS 8

Im Rahmen der Clusteridentifikation wird in Crimestat 2.0 die Kernel density Interpolation integriert. Die Vorgehensweise wird im Folgenden erläutert.

Die **Identifikation von Clustern (3)** (engl. Hot Spot Analysis) erfolgt auf der Basis der Nächste-Nachbar Statistik. Sie ermittelt, inwieweit Punktkonzentrationen signifikant höher sind, als aufgrund einer zufälligen Verteilung der Punkte im Untersuchungsgebiet zu erwarten wäre. Diese Methode wird als *Nearest Neighbour Hierarchical Clustering Method (Nnh)* bezeichnet und identifiziert Häufungen von Ereignissen, die räumlich nah sind. Es handelt sich hierbei um eine hierarchische Clustermethode (Sneath 1957, McQuitty 1960, Sokal & Sneath 1963, King 1967, Sokal & Michener 1958, Ward 1963, Hartigan 1975), die Punktcluster solange gruppiert, bis nur noch ein einziges großes Cluster übrig bleibt (Abb. 18). Die Crimestat Nnh-Routine verwendet die Nächste-Nachbar Methode, die eine Distanz mit einem bestimmten Schwellenwert/ Grenzwert (*threshold distance*) basierend auf einer zufällig erwarteten Nächste-Nachbar Distanz definiert, und diese Distanz mit den Distanzen der Punktpaare vergleicht.

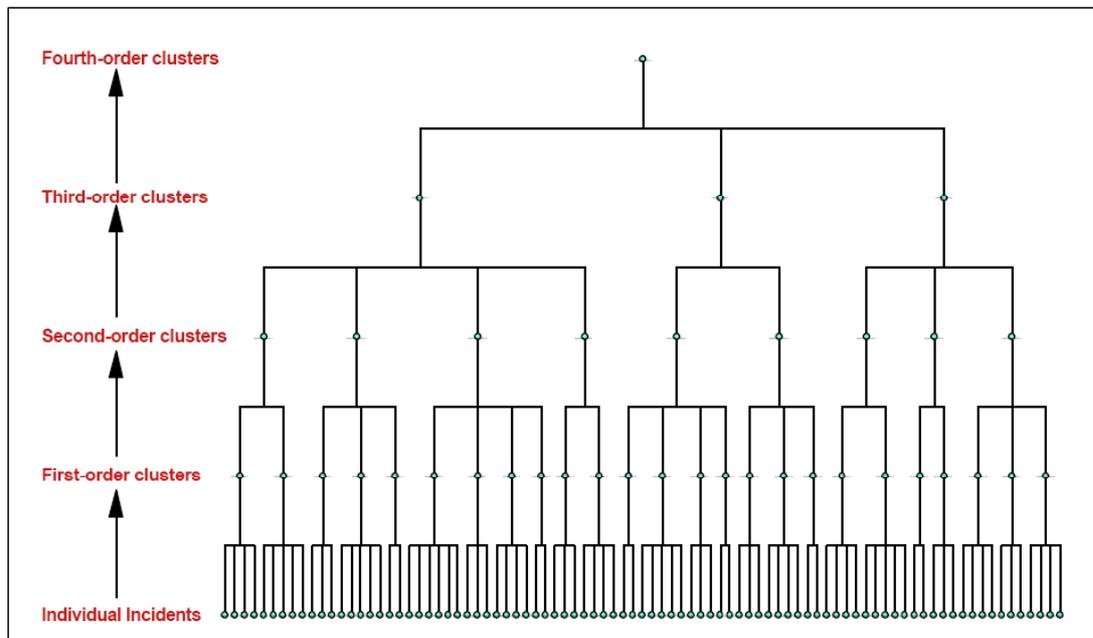


Abb. 18: Darstellung der hierarchischen Clustermethode (Levine 2002)

Nur solche Punkte, deren Distanz geringer ist als die der Schwellenwertdistanz (*engl.: threshold distance*), werden als Cluster identifiziert. Beispiele für die Anwendung dieser Methode sind in Abb. 19 ersichtlich:

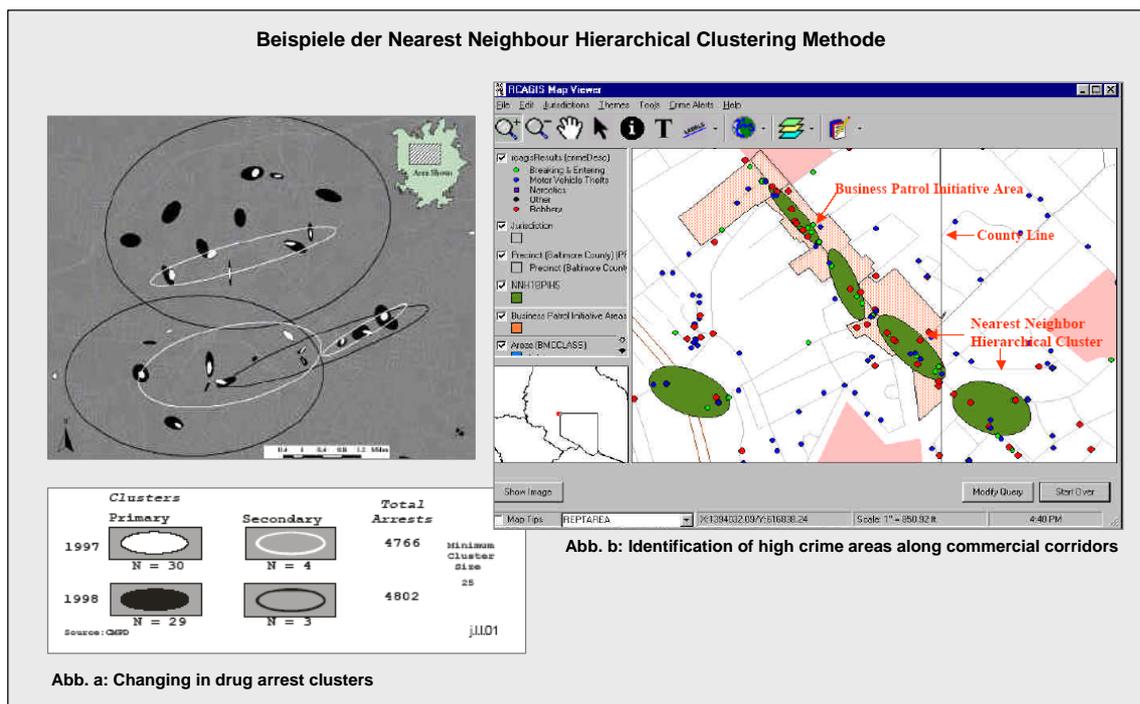


Abb. 19: Beispiele der Nearest neighbour hierarchical clustering Methode

Zwei wichtige Kriterien müssen bei dieser Clustermethode festgelegt werden:

1. Nächste-Nachbar Distanz: Ein wichtiger Eingabeparameter ist das Konfidenzintervall der zufällig erwarteten Nächsten-Nachbar Distanz für Nächste-Nachbarn erster hierarchischer Ordnung
2. Clustergröße: Ebenso wichtig ist die Eingabe der Mindestanzahl an Punkten, die ein Cluster definieren sollen. Crimestat bildet auf Basis dieser Eingabe so genannte Standardabweichungsellipsen, die 50%, 90% oder 99% der Punkte fassen.

Ein wichtiger Vorteil dieser Methode ist, dass sie sich besonders für kleinere Untersuchungsgebiete eignet. Darüber hinaus können Verbindungen zwischen Clustern anhand der hierarchischen Ordnung erkannt werden. Zur Evaluierung der Clusteridentifikation ist es möglich, Monte Carlo Simulationen der primären Cluster zu berechnen. Dazu werden vom Programm n Fälle einem Rechteck mit derselben Größe des Untersuchungsgebietes zufällig zugewiesen und damit die Anzahl Cluster entsprechend der festgesetzten Parameter (z.B. *threshold distance*, Mindestanzahl Punkte pro Cluster) berechnet. Das Ergebnis ist eine Clusteranzahl, die auf Basis einer zufälligen Verteilung zustande gekommen ist. Angenommen es handelt sich beispielsweise um eine Stichprobe von $n=1181$ Raubüberfällen. Bei der Nnh-Methode werden 69 primäre Cluster identifiziert. Die Simulation errechnet auf Basis einer Zufallsverteilung maximal 7 Cluster (95% Perzentil =3). Das Ergebnis der Simulation bestätigt damit, dass 3 Cluster zufällig bestehen, alle übrigen jedoch nicht dem Zufall unterliegen (Levine 2002).

Von besonderer Bedeutung für diese Studie ist allerdings die Möglichkeit, eine zweite Variable für die Bildung von Clustern zu berücksichtigen. So ist die Identifikation der Cluster von Patienten mit Legionelleninfektionen abhängig von der absoluten Patientenzahl im Untersuchungsgebiet. Es liegt nahe, dass Regionen mit vielen Patienten ebenfalls über eine erhöhte Anzahl Legionelleninfektionen verfügen. Um dies bei der Clusteridentifikation zu berücksichtigen, wird eine zweite Variable, nämlich die absolute Patientenzahl, in die Berechnungen integriert.

Mit Hilfe einer Erweiterung der *Nearest neighbour hierarchical clustering* Methode, der so genannten *Risk-adjusted nearest neighbour clustering Methode* (Rnnh-Methode) wird die Schwellenwertdistanz unter Berücksichtigung der absoluten Patientenzahl (secondary file) angepasst, d.h. diese Distanz variiert im gesamten Untersuchungsgebiet.⁶ „*Cirme Stat includes a Risk-adjusted Nearest Neighbour Hierarchical clustering routine (or RNNH) that defines Clusters of points that are closer than what would be expected on the basis of a baseline population...It is a risk measure rather than a volume measure.*“ (Levine 2002: 236). Dies geschieht mit Hilfe eines Interpolationsalgorithmus, welcher auf der Kernel-density-Methode beruht. Für jede Rasterzelle wird die erwartete Anzahl Patienten ermittelt und deren absolute Dichte (Patienten pro Rasterzelle) berechnet. Diese wird auf die Anzahl der serologisch positiven Patienten (primary file) umgerechnet, d.h. die Verteilung der serologisch positiven Patienten ist somit proportional zur absoluten Patientenzahl. Es ist somit möglich, die Verteilung serologisch positiver Patienten mit der Verteilung der absoluten Patientenzahl zu vergleichen. Die Berechnung der RNNH-Methode erfolgt in folgenden Schritten:

1. es werden zwei Datensätze benötigt, die Ereignisvariable (Legionelleninfektionen) und die Basisvariable (absolute Patientenzahl)
2. ein Raster muss definiert werden
3. das Untersuchungsgebiet muss definiert werden (anderenfalls wird die Größe des Rasters zugrunde gelegt)

⁶ Für eine detaillierte Darstellung beider Methoden siehe Levine (2002)

4. die Festlegung der Parameter für die Kernel density Interpolation (Methode der Interpolation, Suchradius, eventuell Gewichtungen)

Aus der Berechnung der Rnnh-Methode können drei Ergebnisse produziert werden: Die Lage der Cluster (dbf-file), das Zentrum des Clusters sowie die Standardabweichungsellipse des Clusters.

Insbesondere für den Bereich Public Health ist diese Methode von besonderer Bedeutung. In Abb. 20 ist ein Beispiel für die Anwendung der Rnnh-Methode im Rahmen der Untersuchung von Tuberkulose-Erkrankungen dargestellt:

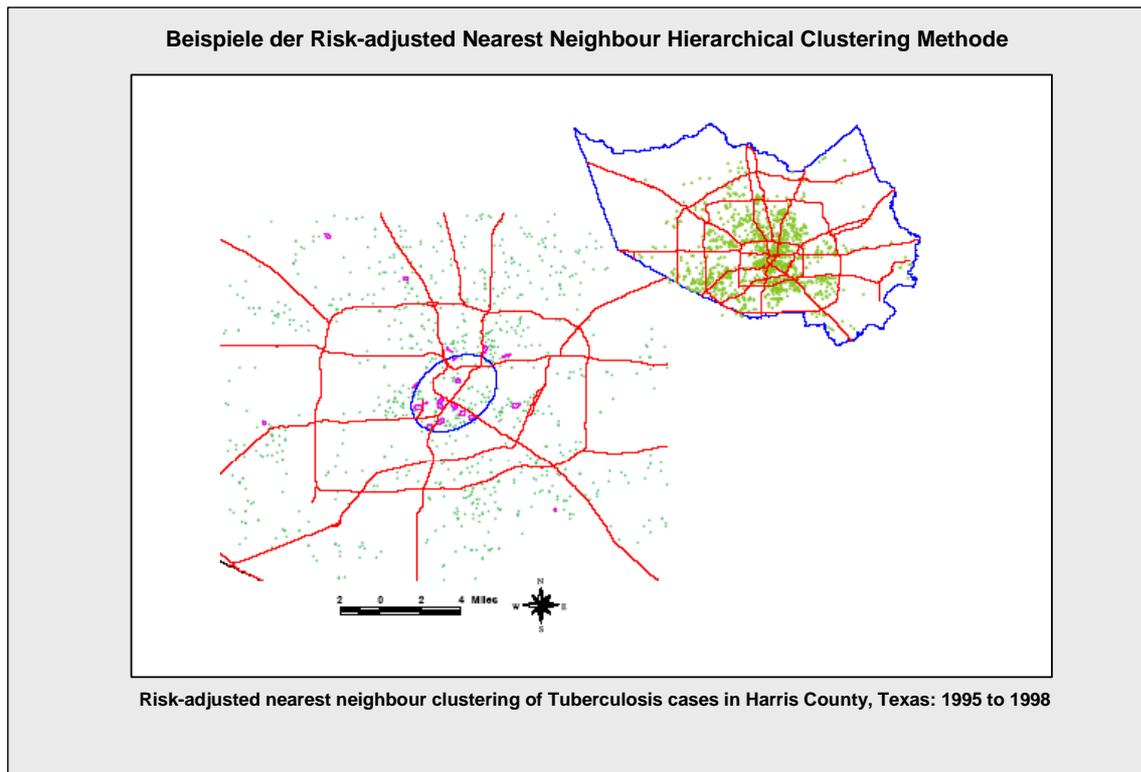


Abb. 20: Beispiel der Risk-adjusted nearest neighbour hierarchical clustering Methode

Die Risk-adjusted nearest neighbour clustering Methode wurde zur Identifikation der Häufung von Tuberkuloseerkrankungen verwendet. Basisdaten waren in diesem Fall Bevölkerungsdaten der Stadt Houston. In den folgenden Kapitel sollen nun die Ergebnisse der erläuterten Methoden vorgestellt werden. Es werden zunächst die für das Verständnis der räumlichen Analyse bedeutsamsten Ergebnisse der Befragung der Patienten dargestellt. Aufbauend darauf, wird die Kohortenstudie durchgeführt, deren Ergebnisse eine Identifikation von Patienten mit einer erhöhten Exposition gegenüber bestimmten Risikofaktoren anhand der Befragung und der serologischen Tests ermöglicht. Anschließend folgt eine ausführliche räumlich-statistische Analyse auf der Basis der Geocodierung der Patientenadressen unter Berücksichtigung der Erkenntnisse aus der statistischen Analyse.

5. Ergebnisse der Datenauswertung

5.1 Deskriptive Datenanalyse

Bevor eine Risikoanalyse auf Grundlage der Befragungsergebnisse der serologischen Tests und auf deren Grundlage eine räumlich-statistische Analyse durchgeführt wird, soll die deskriptive Auswertung einen Überblick über das Datenmaterial geben und erste Ansätze der Interpretation liefern. Insbesondere die deskriptive Analyse der Patientenbefragung ermöglicht im Zusammenspiel mit den serologischen Ergebnissen erste Erkenntnisse im Hinblick auf die Risikoanalyse.

5.1.1 Deskriptive Analyse der Patientendaten im Untersuchungszeitraum

Es wurden insgesamt 14.389 Patienten im Rahmen der Untersuchung erfasst. Diese absolute Anzahl der Patienten verteilt sich über den Untersuchungszeitraum wie folgt (Abb. 21):

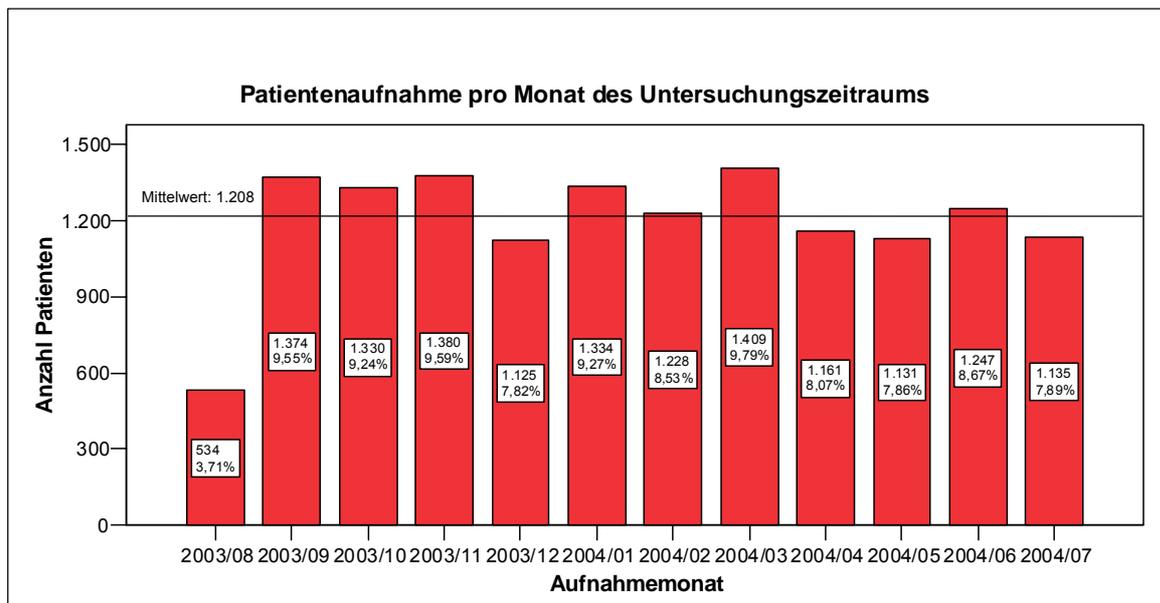


Abb. 21: Patientenaufnahme pro Monat des Untersuchungszeitraums (absolut und prozentual)

Die Befragung der Patienten begann im August 2003. In diesem Monat wurden noch nicht viele Patienten befragt und serologisch untersucht, daher liegt hier die Patientenzahl deutlich unter dem monatlichen Durchschnitt von ca. 1.208 Patienten. Betrachtet man die Altersverteilung der Patienten, so ist die Altersklasse von 66 bis 75 Jahren am häufigsten vertreten (Abb. 22).

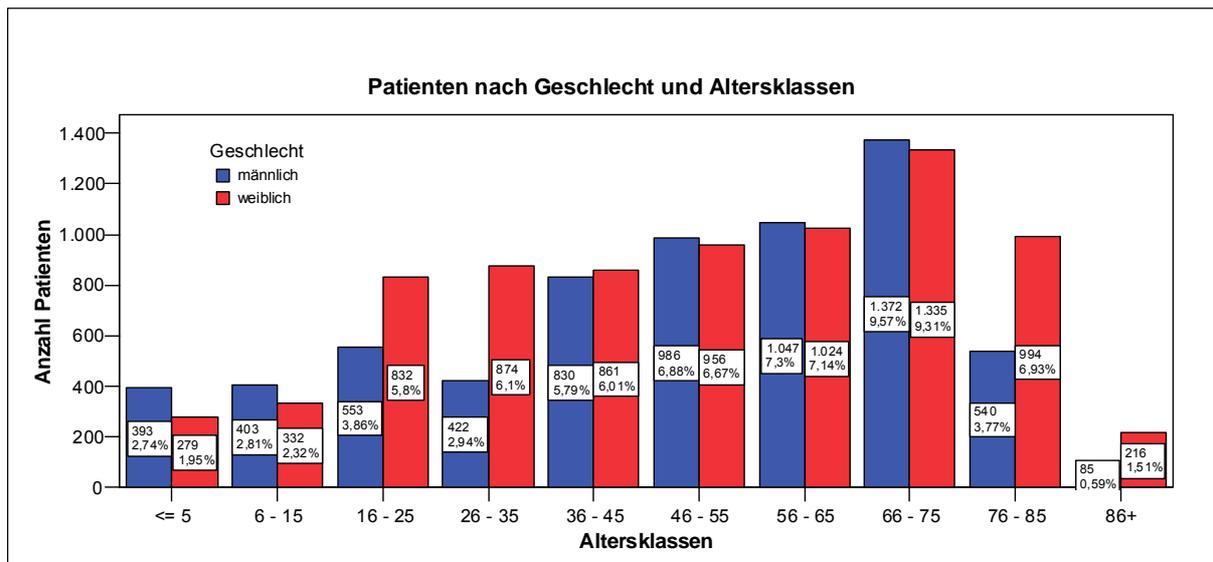


Abb. 22: Patienten nach Geschlecht und Altersklassen (absolut und prozentual)

In der Altersklasse zwischen 76 und 85 Jahren erkennt man darüber hinaus einen deutlichen Überhang des weiblichen Geschlechts, ebenso in den Altersklassen von 16 bis 35 Jahren. In den übrigen Altersklassen überwiegt das männliche Geschlecht.

5.1.2 Deskriptive Analyse des Serostatus der Patienten

Insgesamt wurden 50 IgM-positive Patienten, 777 IgG-positive Patienten ermittelt, wovon 13 sowohl IgM- als auch IgG-positiv waren (Abb. 23).

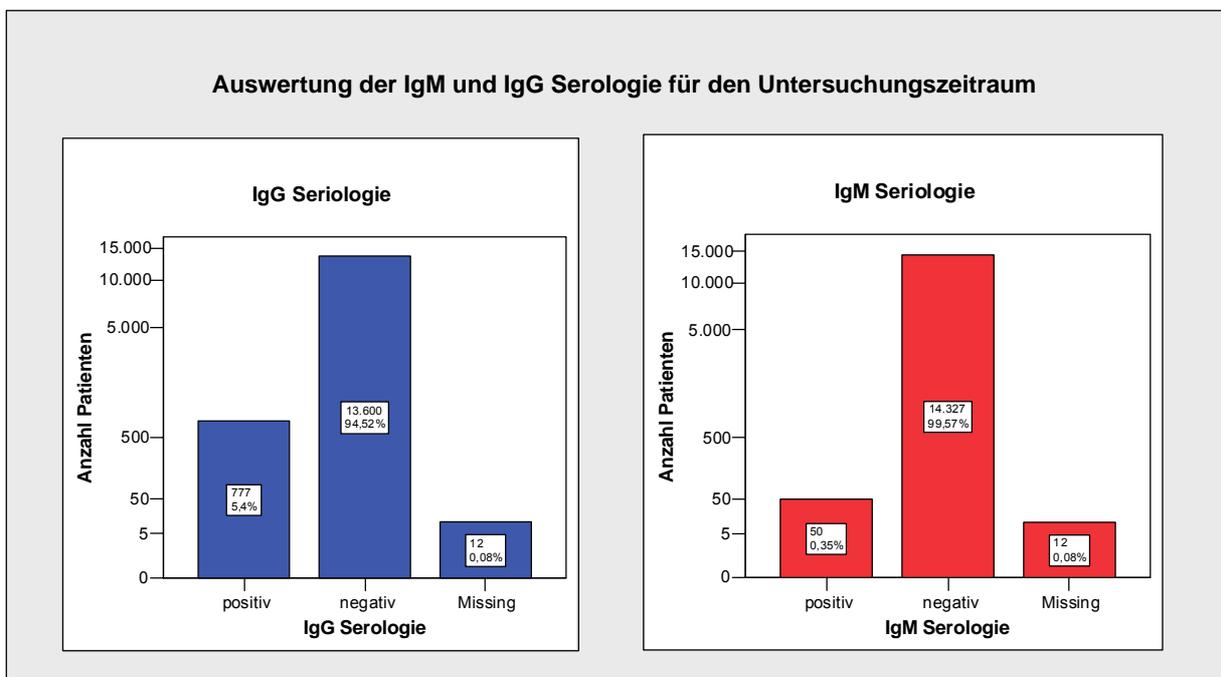


Abb. 23: Auswertung der IgM und IgG Serologie (exponentielle Darstellung)

Bei einem Patienten innerhalb des Untersuchungszeitraum wurde bei Aufnahme Legionella Antigen im Urin nachgewiesen, d.h. dieser Patienten war bei Aufnahme mit Legionellen infiziert. Zwei weitere Patienten wurden jeweils unmittelbar vor Beginn und nach Beendigung

des Untersuchungszeitraums mit Aufnahme in das Krankenhaus Antigen positiv diagnostiziert. Für weitere Berechnungen werden diese beiden Patienten nicht berücksichtigt, da sie außerhalb des definierten Untersuchungszeitraumes aufgenommen wurden.

Wenn in den Fragebögen darauf verwiesen wurde, dass ein serologischer Test nicht durchführbar war oder der Patient sich diesem Test nicht unterziehen wollte, wurde dies in der statistischen Analyse berücksichtigt (missing values).

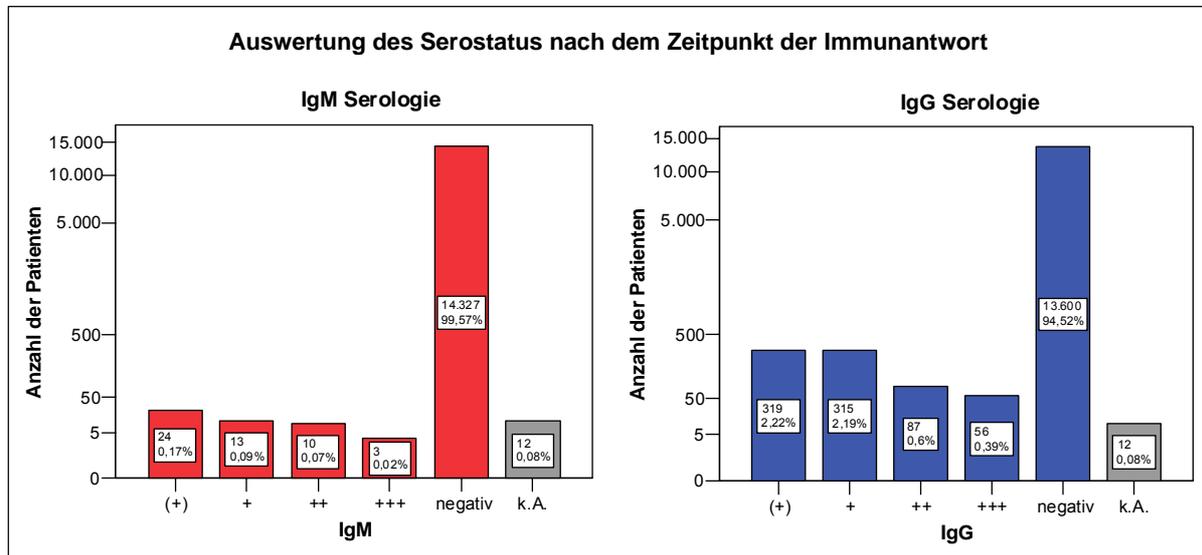


Abb. 24: Auswertung des Serostatus nach dem Zeitpunkt der Immunantwort (exponentielle Darstellung)

Der Zeitpunkt der Infektion bei den IgG-positiven Patienten lag in der Mehrzahl der Fälle weiter zurück (Abb. 24). So wurden bei 319 der 777 positiven Patienten lediglich schwach positive Ergebnisse detektiert, was auf eine länger zurückliegende Infektion mit bereits abfallenden Titern hinweist, wenn nicht durch gleichzeitig nachweisbares IgM auf eine beginnende Immunantwort geschlossen werden kann. Auch die IgM-Antwort zeigte nur bei 13 Patienten stark positive Ergebnisse, was unter Berücksichtigung des diagnostischen Fensters auf eine erst kürzlich erfolgte Infektion hinweist. Dieser geringe Anteil korrespondiert mit dem ebenfalls extrem niedrigen Anteil von Antigen-positiven Patienten, d.h. solche Patienten die zu diesem Zeitpunkt mit Legionellen infiziert sind.

Die Prävalenz pro Monat, berechnet aus den absoluten Anzahlen IgM- bzw. IgG-positiver Patienten im Verhältnis zur gesamten untersuchten Patientenzahl, sieht für IgM folgendermaßen aus (Abb. 25):

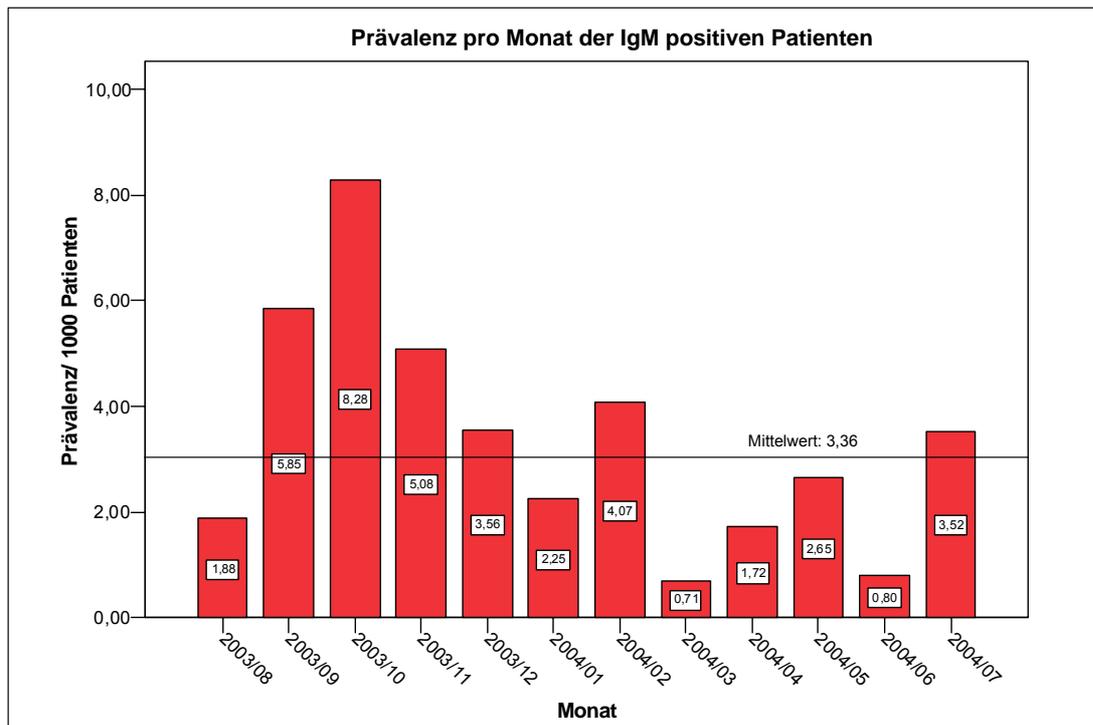


Abb. 25: Prävalenz pro Monat der IgM-positiven Patienten

Im Monat Oktober wird mit einem Wert von 8,28 pro 1.000 Patienten die mit Abstand höchste Prävalenz erreicht. Hingegen ist die Prävalenz in den Monaten März (0,71/ 1.000 Patienten) und Juni (0,80/ 1.000 Patienten) sehr gering. Es bildet sich saisonale Häufung in der kalten Jahreszeit ab. Für die IgG-positiven Patienten ist folgende monatliche Entwicklung erkennbar (Abb. 26):

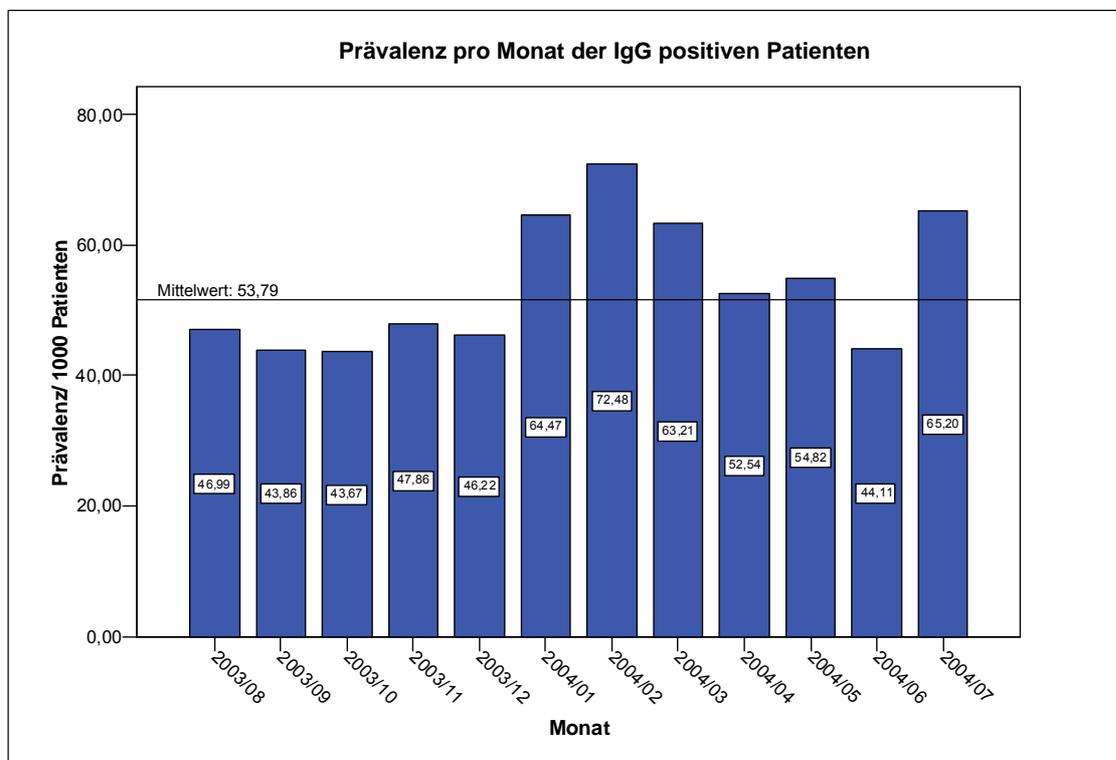


Abb. 26: Prävalenz pro Monat der IgG-positiven Patienten

Die Prävalenz pro Monat der IgG-positiven Patienten ist weitaus höher als die der IgM-positiven Patienten. So wird im Monat Februar 2004 eine Prävalenz von 72,48 erreicht. Eine stetige etwas geringere Prävalenz ist von August 2003 bis Dezember 2003 erkennbar. Auf Grund der längeren Persistenz der IgG-Antikörper ist erwartungsgemäß eine saisonale Abhängigkeit nicht zu erkennen.

Neben der Prävalenz pro Monat konnte die Prävalenz je Altersklasse ermittelt werden (Abb. 27)

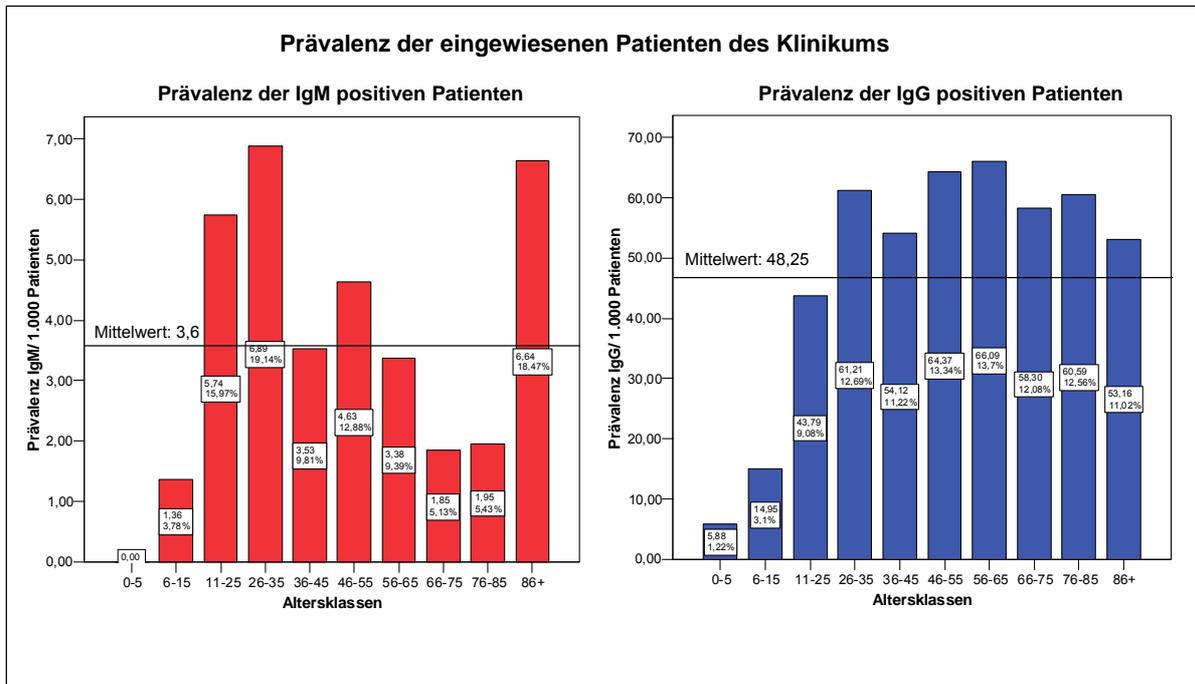


Abb. 27: Prävalenz der eingewiesenen Patienten des Klinikums

Aus diesen absoluten Fallzahlen wurde die Prävalenz sowohl für IgM als auch IgG-positive Patienten für die einzelnen Altersklassen errechnet (Abb. 27). Auffällig sind die recht großen Unterschiede der Prävalenz in den Altersklassen der IgM-positiven Patienten. Eine hohe Prävalenz erreichen die Altersklassen von 11 bis 35 Jahren sowie vor allem die über 85-Jährigen. Hinsichtlich IgG-Status sind bei den über 25jährigen Patienten keine relevanten Unterschiede zwischen den jeweiligen Altersklassen erkennbar. Bei den jüngeren Altersgruppen steigt hingegen die Prävalenz mit dem Lebensalter steil an, was im Sinne einer Kumulation stattgefundenener Immunreaktionen interpretiert werden kann.

Die Seroprävalenz wurde auch hinsichtlich der Häufigkeit der Klinikumsaufenthalte analysiert (Abb. 28).

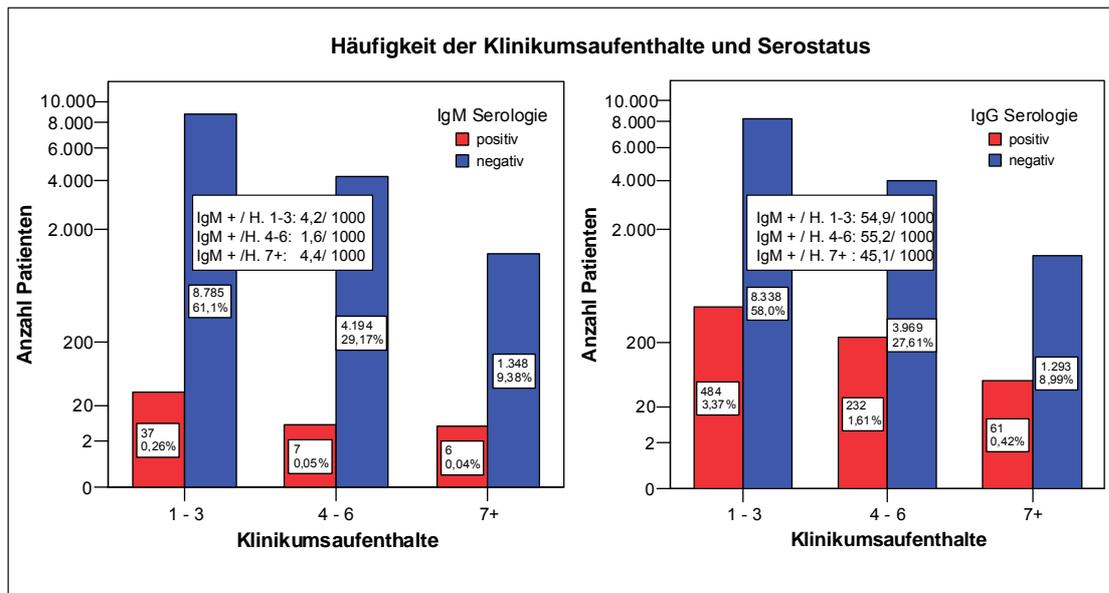


Abb. 28: Häufigkeit der Krankenhausaufenthalte und Serostatus

Die höchste absolute Anzahl der Patienten mit positiver IgM- oder IgG-Serologie wurde ein- bis dreimal in das Klinikum aufgenommen. Knapp die höchste IgM-Prävalenz (4,4/1000) haben Patienten, die mindestens sieben Mal aufgenommen wurden. Bei der IgG-Prävalenz liegt die mittlere Gruppe (4-6 Aufnahmen) knapp vorn (55,2/1000), die Unterschiede sind jedoch gering.

Die folgenden Erläuterungen zur deskriptiven Analyse der Angaben aus den Fragebögen werden nur beispielhaft dargestellt, da sie für die weitere räumlich-statistische Auswertung kaum Relevanz haben. Die Ergebnisse der Befragung wurden in der Datenbank mit den serologischen Ergebnissen verknüpft, um schließlich eine Auswertung der Befragung unter Berücksichtigung der serologischen Tests ermöglichen zu können.

Eine hier dargestellte Beispiel betrifft die zweite Frage zum regelmäßigen Besuch von Schwimmbädern oder sonstigen Sportstätten (Abb. 29).

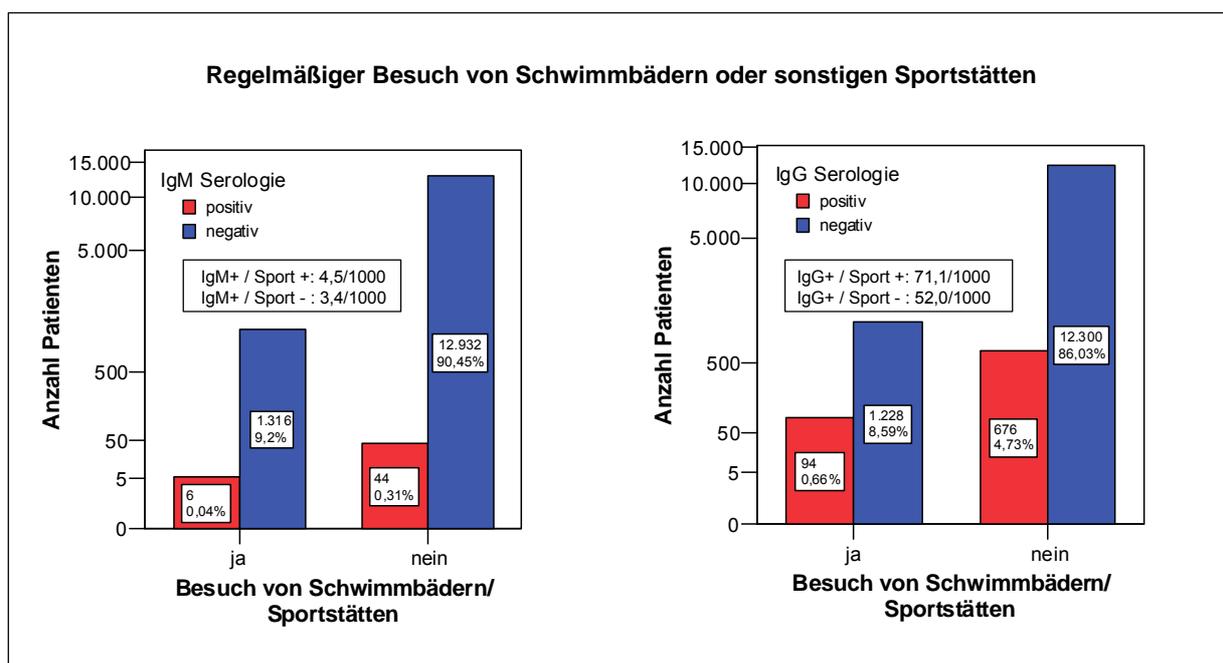


Abb. 29: Auswertung der Befragung zum Besuch von Schwimmbädern/ Sportstätten

Insgesamt sechs der 50 IgM-positiven Patienten und 0,04% der gesamten befragten Patienten haben sich regelmäßig in Schwimmbädern oder sonstigen Sportstätten aufgehalten. Bei den IgG-positiven Patienten handelt es sich um 94 (0,66% aller Patienten) von 777 Patienten, welches einen Anteil von 7,1% derer ausmacht, welche die Frage bejaht haben. Daraus lässt sich eine Prävalenz für IgM-positive Patienten mit regelmäßigen sportlichen Aktivitäten von 4,5 pro 1.000 Patienten ermitteln. Die Prävalenz für IgG positive Patienten ist mit 71,1 pro 1.000 weit höher. Bei beiden Antikörperklassen ist die Prävalenz ohne derartige regelmäßige sportliche Aktivitäten geringer. Für IgM Positive beträgt diese lediglich 3,5 pro 1.000 Patienten, bei IgG-positiven 52,0 pro 1.000 Patienten.

Im Rahmen des Gutachtens wurde eine ausführliche deskriptive Datenauswertung durchgeführt (siehe Kistemann et al. 2004)

5.2 Analytisch-epidemiologische Untersuchung

In den folgenden Unterkapiteln werden die Ergebnisse der Risikoberechnung dargestellt und erläutert. Dabei handelt es sich bei den Expositionsfaktoren um die genannten Expositionen, die im Fragebogen behandelt wurden, sowie um weitere in der Literatur bekannte Expositionsfaktoren, wie Alter und Geschlecht. Über diese Faktoren hinaus wurde die Diagnose der Patienten (ermittelbar anhand des ICD 10 Code) bei Einlieferung in das Klinikum als ein weiterer Expositionsfaktor in die Risikoberechnung integriert.

5.2.1 Risikoberechnung anhand der Ergebnisse der Befragung

In der folgenden Risikoberechnung soll nun festgestellt werden, inwieweit die im Fragebogen angesprochenen Faktoren, wie Reiseaktivität, Sportaktivitäten, Verwendung von Luftbefeuchtern etc. zu einem erhöhten Infektionsrisiko mit Legionellen führen. In Kap. 4.3.2 wurde bereits die Vorgehensweise und Methodik zur Berechnung des relativen Risikos erläutert. Die Ergebnisse der Risikoberechnung, ebenso die obere und untere Grenze des 95%-Konfidenzintervalls des relativen Risikos und das Ergebnis des Signifikanztests nach Fisher sind in den folgenden Tabellen dargestellt. Die Tab. 3 stellt die beschriebene Risikoberechnung zunächst für die IgM-Serologie dar:

Tab. 3: Berechnung des relativen Risikos für die Ergebnisse der IgM Serologie

	Unabhängige Variable	Relatives Risiko	95% Konfidenzintervall rel. Risiko		Signifikanz
			Untere	Obere	
1	Reisen in den letzten 12 Tagen	0,66	0,16	2,71	0,77
2	Besuch von Schwimmbädern/ Sportstätten	1,34	0,57	3,14	0,46
3	Verwendung von Luftbefeuchtern	0,91	0,28	2,90	0,99
4	Existenz von Rückkühlwerken	3,29	0,46	23,57	0,27
5	Auftreten von Atemwegsinfektionen	1,34	0,60	2,96	0,49
6	Gehäuft auftretende Temperaturerhöhung	1,36	0,33	5,57	0,66

Die Interpretation dieser Ergebnisse gibt keinen signifikanten Hinweis auf ein erhöhtes Risiko für die genannten Expositionen. Zwar ist bei Frage 2, 4, 5 und 6 das relative Risiko erhöht,

allerdings zeigen die untere Grenze des 95%-Konfidenzintervalls (jeweils <1) und der Fisher-Test (jeweils $p > 0,1$), dass es sich hierbei nicht um signifikante Ergebnisse handelt.

In Tab. 4 werden die Ergebnisse der Berechnung des relativen Risikos für die IgG-Serologie dargestellt:

Tab. 4: Berechnung des relativen Risikos für die IgG Serologie

	Unabhängige Variable	Relatives Risiko	95% Konfidenzintervall rel. Risiko		Signifikanz
			Untere	Obere	
1	Reisen in den letzten 12 Tagen	0,96	0,71	1,29	0,87
2	Besuch von Schwimmbädern/ Sportstätten	1,37	1,11	1,68	0,004
3	Verwendung von Luftbefeuchtern	1,16	0,89	1,50	0,26
4	Existenz von Rückkühlwerken	1,06	0,45	2,48	0,81
5	Auftreten von Atemwegsinfektionen	0,90	0,72	1,14	0,40
6	Gehäuft auftretende Temperaturerhöhung	0,83	0,53	1,30	0,44

Die Risikoberechnung der Tab. 4 zeigt, dass mit Ausnahme einer Exposition kein erhöhtes Risiko für eine Legionelleninfektion in Hinblick auf die verschiedenen Expositionsfaktoren besteht. Lediglich der Besuch von Schwimmbädern oder sonstigen Sportstätten zeigt ein erhöhtes Risiko, welches auch durch das 95%-Konfidenzintervall abgesichert ist. Das Ergebnis ist mit $p = 0,004$ hoch signifikant.

Die Berechnung des relativen Risikos wurde auch für IgM- und IgG-Serologie gemeinsam durchgeführt. Die Ergebnisse sind in Tab. 5 dargestellt.

Tab. 5: Berechnung des relativen Risikos für beide Antikörperklassen

	Unabhängige Variable	Relatives Risiko	95% Konfidenzintervall rel. Risiko		Signifikanz
			Untere	Obere	
1	Reisen in den letzten 12 Tagen	0,96	0,72	1,28	0,81
2	Besuch von Schwimmbädern/ Sportstätten	1,36	1,11	1,66	0,004
3	Verwendung von Luftbefeuchtern	1,12	0,87	1,45	0,38
4	Existenz von Rückkühlwerken	1,01	0,43	2,37	0,99
5	Auftreten von Atemwegsinfektionen	0,92	0,73	1,15	0,48
6	Gehäuft auftretende Temperaturerhöhung	0,83	0,54	1,28	0,45

Wiederum kann der Besuch von Schwimmbädern und sonstigen Sportstätten als ein Risikofaktor erklärt werden. Dies wird durch das 95%-Konfidenzintervall bestätigt. Das Ergebnis ist mit 0,004 hoch signifikant. Alle übrigen Ergebnisse lassen keine Aussage über ein erhöhtes Risiko durch die genannten Expositionen zu. Zwar ist das relative Risiko leicht erhöht, jedoch

liegt die untere Grenze des Konfidenzintervalls jeweils unter 1, und das Ergebnis ist nicht signifikant.

Über die Expositionsfaktoren, die in der Befragung ermittelt wurden, hinaus sind das Alter der Patienten sowie das Geschlecht als weitere mögliche Risikofaktoren überprüft worden. So wurde jeweils für die IgG Serologie, die IgM Serologie und für beide zusammen eine Risikoberechnung durchgeführt. Die Ergebnisse dieser Risikoberechnung sind in Tab. 6 dargestellt.

Das relative Risiko der unabhängigen Variable Geschlecht liegt bei IgG sowie IgM+IgG über 1, ebenso das gesamte 95%-Konfidenzintervall. Das Ergebnis ist hoch signifikant. Dies bedeutet, dass die weiblichen Patienten einem schwach erhöhten Risiko unterliegen.

Bei der Betrachtung des Alters als Expositionsfaktor liefert die Risikoberechnung kein eindeutiges Ergebnis. Die IgM Serologie lässt sich sogar dahingehend interpretieren, dass jüngere Patienten einem erhöhten Risiko unterliegen. Das Ergebnis ist auch hier signifikant. Bei Betrachtung der IgG Serologie besteht hingegen eher für ältere Menschen ein erhöhtes Risiko. Anhand des 95%-Konfidenzintervalls ist zumindest eine leichte Tendenz eines solchen Risikos anzunehmen, welches ebenfalls anhand der Signifikanz von $p = 0,07$ erkennbar ist. Entkräftet wird dieses Ergebnis durch die gemeinsame Betrachtung von IgM und IgG, welches zwar ein erhöhtes relatives Risiko zeigt, jedoch mit $p = 0,16$ nicht signifikant ist. Hier kann man keine eindeutige Aussage treffen.

Tab. 6: Ergebnisse der Risikoberechnung zu den Expositionsfaktoren Geschlecht und Alter

Unabhängige Variable	Serologie	Relatives Risiko	95% Konfidenzintervall rel. Risiko		Signifikanz
			Untere	Obere	
Geschlecht weiblich	IgM	1,53	0,86	2,73	0,16
	IgG	1,26	1,10	1,45	0,001
	IgM + IgG	1,27	1,11	1,46	0,001
Alter >65 Jahre	IgM	0,50	0,25	0,99	0,05
	IgG	1,14	1,00	1,32	0,07
	IgM + IgG	1,11	1,00	1,27	0,16

Zusammenfassend ist festzustellen, dass

1. der Besuch von Schwimmbädern und sonstigen Sportstätten, sowie
2. weibliches Geschlecht

in der analytisch-epidemiologischen Untersuchung als signifikante Risikofaktoren für positive IgG-Serologie sowie positive IgG- und/oder IgM-Serologie identifiziert wurden. Alle übrigen überprüften Variablen erwiesen sich als nicht signifikant.

5.3 Räumlich-statistische Analyse der Patientendaten

Die räumliche Analyse der Patientendaten soll nun Aufschluss darüber bringen, wie die räumliche Verteilung der Patienten mit positivem Serostatus im Einzugsgebiet des Klinikums ist und ob es eventuell Cluster, d.h. Gebiete mit einer relativen Konzentration von IgM- oder IgG-positiven Patienten gibt. Anhand solcher „hotspots“ können Risikogebiete und darüber hinaus möglicherweise umweltrelevante Faktoren für ein erhöhtes Risiko einer Legionellen-Infektion identifiziert werden. Eine solche Analyse basiert auf der in Kap. 4.3.3.2 beschriebenen Geocodierung. Als Grundlage der punktbasierten Analyse, wurden die Adressen auf den jeweiligen Postleitzahlbezirk aggregiert, um somit eine flächenhafte Analyse durchführen zu können. Diese bietet einen ersten Überblick über die Verteilung der serologisch positiven Patienten.

5.3.1 Flächenhafte räumliche Analyse

Eine flächenhafte räumliche Analyse auf Ebene der Postleitzahlbezirke soll Regionen mit einem erhöhten Risiko für eine Legionellen-Infektion identifizieren. Die absolute Zahl der Patienten, die IgG positiv waren, wurde auf jeden Postleitzahlbezirk heruntergebrochen (Abb. 30). Dabei wurden nur solche Postleitzahlbezirke in die Berechnung miteinbezogen, die mindestens 20 oder mehr Patienten aufwiesen. Die Karte zeigt deutlich, dass insbesondere der Raum Frankfurt/ Oder die höchste absolute Anzahl IgG positiver Patienten aufweist.

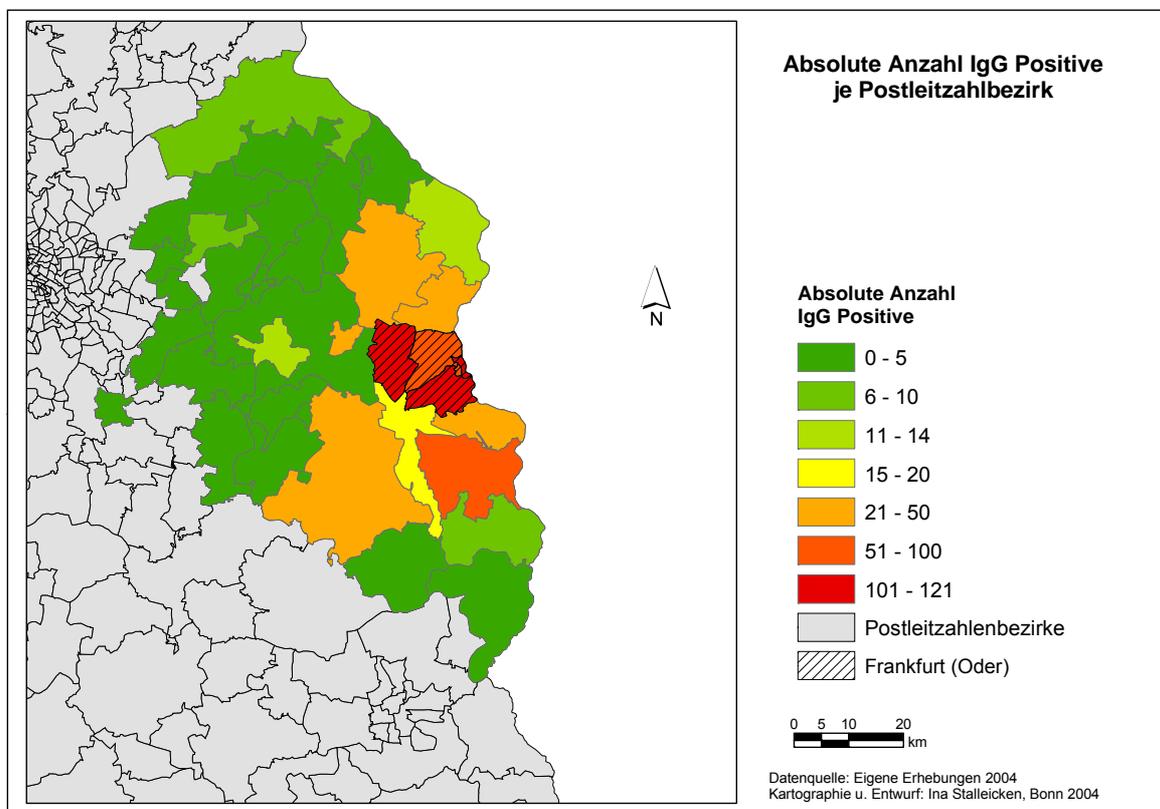


Abb. 30: Absolute Anzahl IgG positive Patienten je Postleitzahlbezirk

Um ein reelleres Bild der Anzahl IgG positiver Patienten zu erlangen, wird anhand der Berechnung der Prävalenz die absolute Anzahl der Patienten je Postleitzahlbezirk berücksichtigt. In Abb. 31 ist die Prävalenz der IgG positiven Patienten dargestellt:

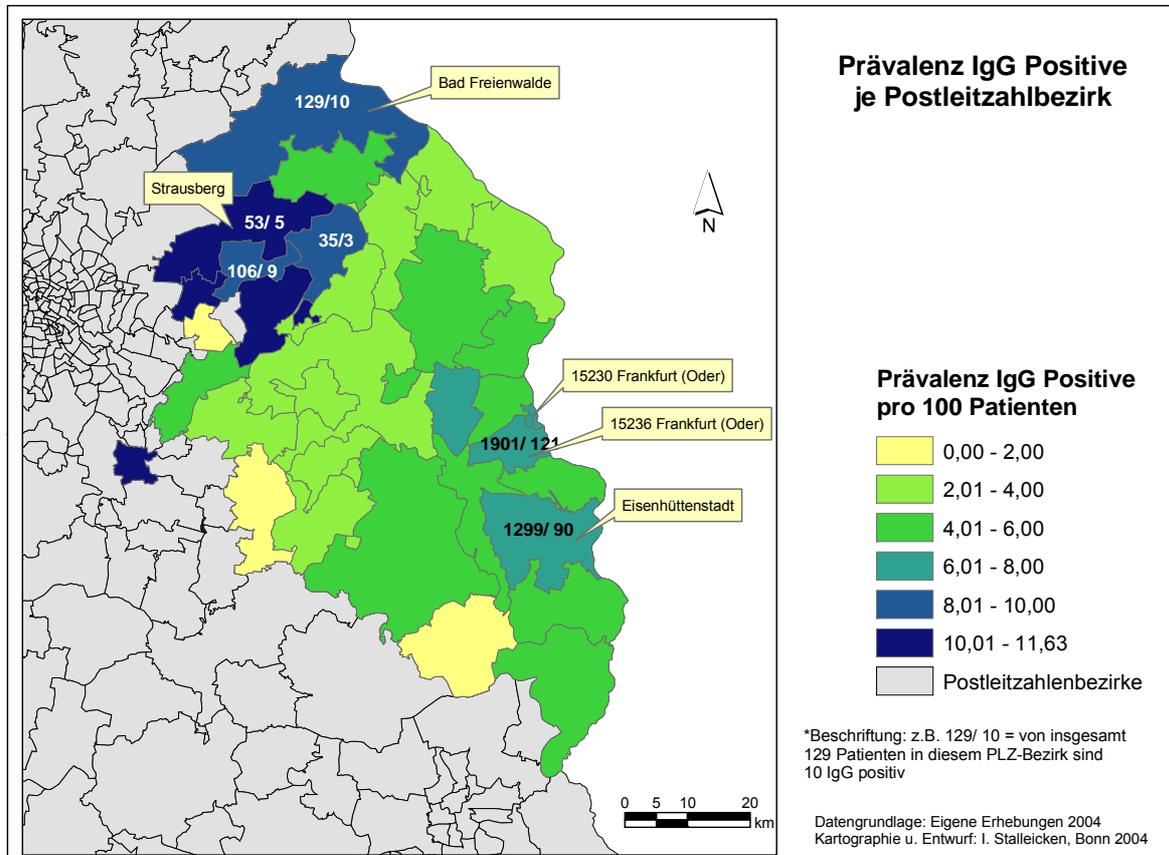


Abb. 31: Prävalenz de IgG positiven Patienten je Postleitzahlbezirk

Im nördlichen Teil des Untersuchungsgebietes fallen mehrere Postleitzahlbezirke auf, die über eine hohe Prävalenz verfügen. Darüber hinaus besitzen einzelne Postleitzahlgebiete von Frankfurt (Oder) sowie Eisenhüttenstadt eine erhöhte Prävalenz. Jedoch ist hier zu beachten, dass Gebiete mit insgesamt geringeren Patientenzahl mehr Aufmerksamkeit auf sich ziehen, als Gebiete mit einer höheren Patientenzahl. So werden beispielsweise die nördlichen Postleitzahlbezirke des Untersuchungsgebietes aufgrund geringer absoluter Patientenzahlen deutlich überrepräsentiert.

Um nun ermitteln zu können, ob es in einem Gebiet überdurchschnittlich viele Legionelleninfektionen gibt, soll der Lokationsquotient⁷ ermittelt werden. Bevor dies geschieht, soll anhand der Berechnung der Räumlichen Autokorrelation (Moran's/ Geary's) festgestellt werden, ob benachbarte Werte im Untersuchungsgebiet räumlich miteinander korrelieren (Abb. 32)

⁷ Der Lokationsquotient gibt das Verhältnis der beobachteten Fälle der Raumeinheiten zu den erwarteten Fällen (=arithmetisches Mittel des gesamten Untersuchungsraumes) an und ist eines der meist angewendeten Verfahren zur Identifikation von Extremwerten (Gesler 1986).

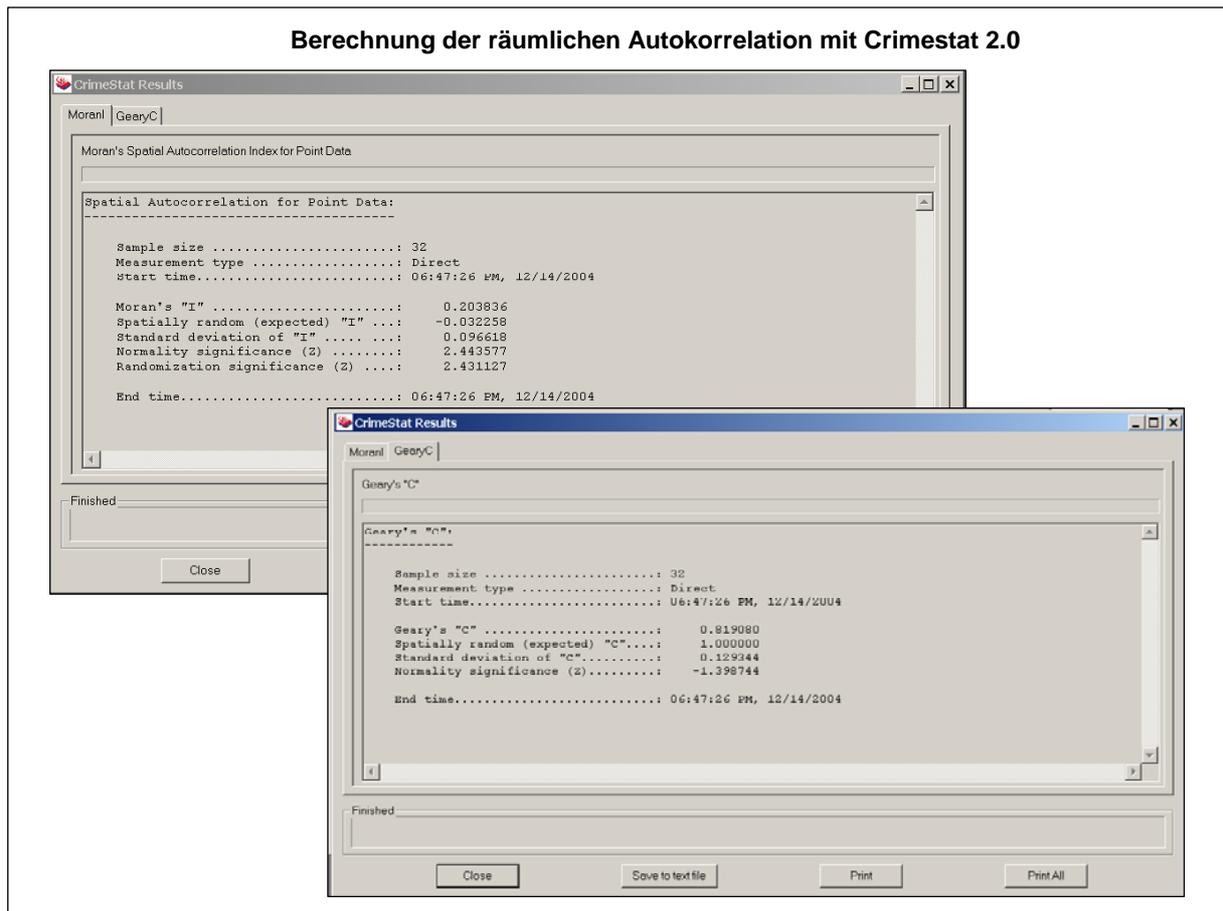


Abb. 32: Berechnung der räumlichen Autokorrelation mit Crimestat 2.0

Insgesamt wurde die räumliche Autokorrelation für 32 Postleitzahlbezirke anhand der Prävalenzen je Postleitzahlbezirk ermittelt. Der errechnete Wert der Moran's I Statistik liegt mit 0,203836 über dem zu erwartenden Wert einer räumlichen Unabhängigkeit von $-0,032258$, d.h. es kann eine positive räumliche Autokorrelation festgestellt werden. Das Ergebnis ist statistisch signifikant, da der errechnete z-Wert (normality significance z) mit 2,443577 über einem kritischen z-Wert von 2,33 mit einem 99%-Konfidenzintervall liegt. Der errechnete Wert der Geary's Statistik liegt mit 0,819080 unter dem zu erwartenden Wert einer räumlichen Unabhängigkeit von 1. Auch in diesem Fall kann eine positive räumliche Autokorrelation interpretiert werden, die jedoch nicht statistisch signifikant ist, da der errechnete z-Wert von $-1,398744$ ⁸ unter dem kritischen z-Wert eines 99%-Konfidenzintervalls von 2,33 liegt.

In Abb. 33 sind die Prävalenz der IgG-positiven Patienten und der zugehörige Lokationsquotient dargestellt. Erreicht dieser Quotient in einem Bezirk einen Wert >1 , so kann man in diesem Gebiet von einer überdurchschnittlich hohen Prävalenz ausgehen.

⁸ Der Wert der z-Teststatistik wird als Absolutwert gelesen, d.h. 1,398744

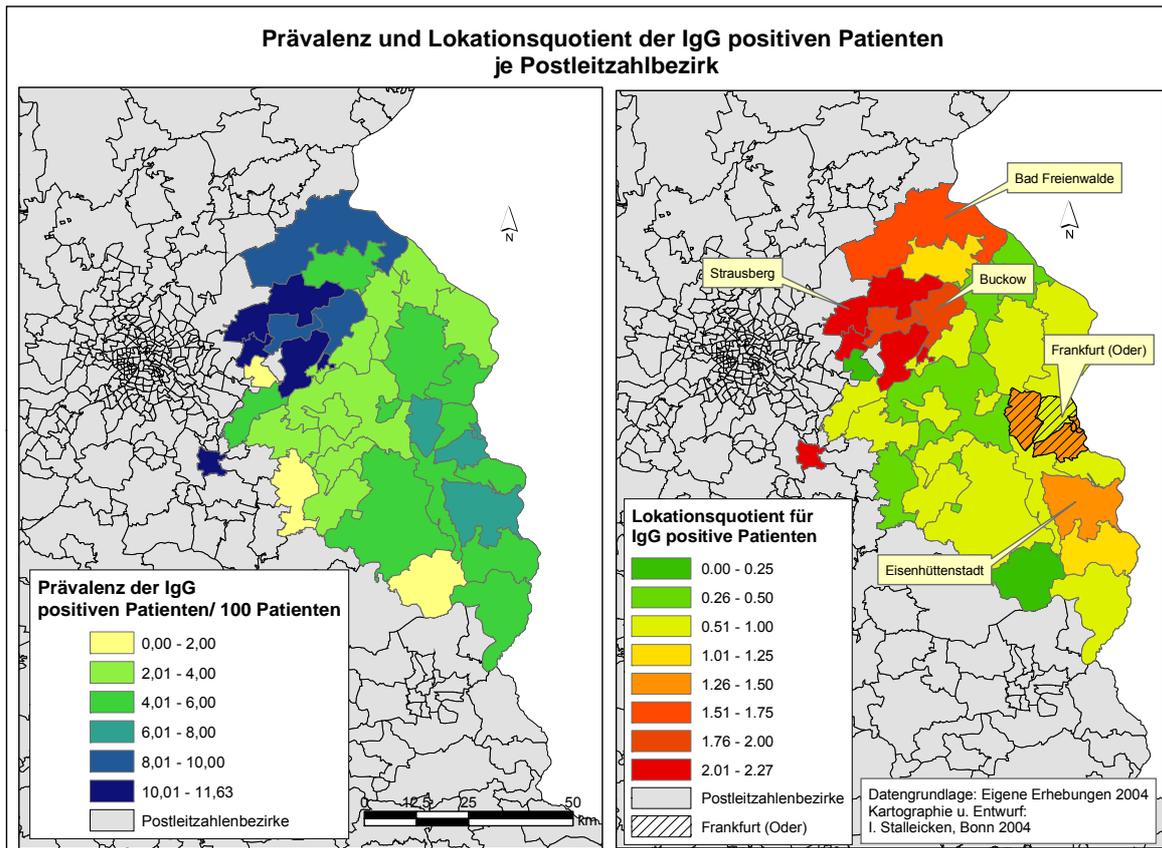


Abb. 33: Prävalenz und Lokationsquotient für die IgG-positiven Patienten

Einzelne Bezirke von Frankfurt (Oder), Eisenhüttenstadt, Buckow, aber auch Strausberg weisen eine erhöhte Prävalenz sowie einen erhöhten Lokationsquotienten auf. Die absolute Patientenzahl ist jedoch teilweise sehr niedrig. So resultiert beispielsweise die hohe Prävalenz im Bezirk Strausberg daraus, dass von insgesamt 53 untersuchten Patienten 5 IgG-positiv waren.

5.3.2 Punktbasierte räumliche Analyse der Patientendaten im Untersuchungsgebiet

Die geocodierten Patientendaten wurden in das Geographische Informationssystem implementiert (Abb. 34). Patienten mit erstem Wohnsitz im Ausland (insbesondere Polen) konnten nicht berücksichtigt werden. Insgesamt wurden 14.452 Patienten geocodiert, von denen allerdings nicht alle an der Befragung teilgenommen haben, bzw. der Fragebogen nicht ausgefüllt werden konnte.

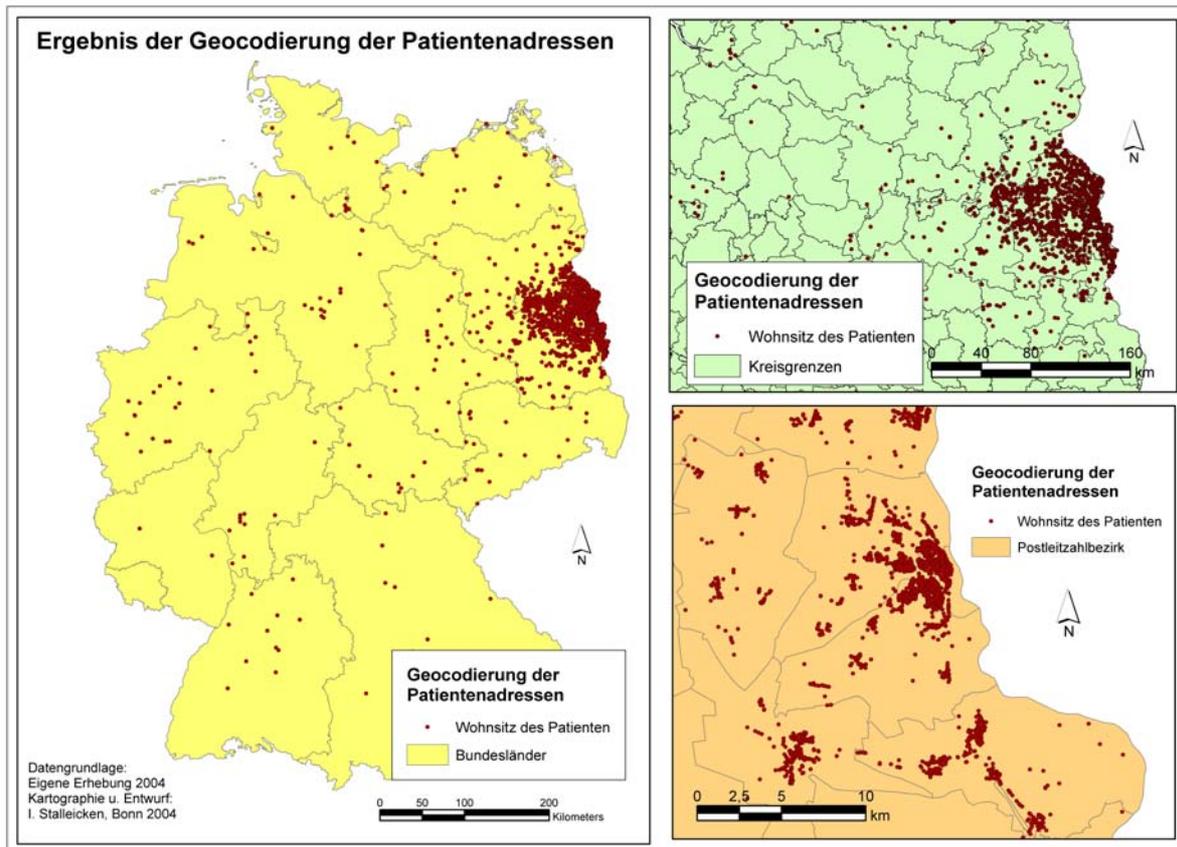


Abb. 34: Ergebnis der Geocodierung der Patientenadressen

Abb. 34 umfasst drei Karten unterschiedlichen Maßstabes:

1. die Verteilung der Patienten über die gesamte Bundesrepublik Deutschland
2. die Verteilung der Patienten im Land Brandenburg mit Darstellung der Kreisgrenzen
3. die Verteilung der Patienten im Raum Frankfurt (Oder) mit Darstellung der Postleitzahlbezirke

Das Klinikum verfügt um ein recht großes Einzugsgebiet, welches sich jedoch erwartungsgemäß auf das Land Brandenburg, insbesondere auf den Raum Frankfurt (Oder) konzentriert. Auf die Bundesländer Berlin und Brandenburg entfallen allein 14.254 Patienten. Alle übrigen bleiben bei der weiteren räumlichen Betrachtung, insbesondere bei der Analyse des Serostatus, unberücksichtigt.

Von den berücksichtigten Patienten wurden im nächsten Schritt diejenigen selektiert, bei denen bei Aufnahme im Klinikum ein positiver Serostatus (IgM und/oder IgG) nachgewiesen wurde (Abb. 35). In allen drei Karten sind Häufungen benachbarter Patienten mit positivem

Serologiestatus im Raum Frankfurt (Oder), Eisenhüttenstadt u.a.m. erkennbar. Aus dieser Darstellung darf man jedoch noch nicht auf die Existenz von Clustern schließen.

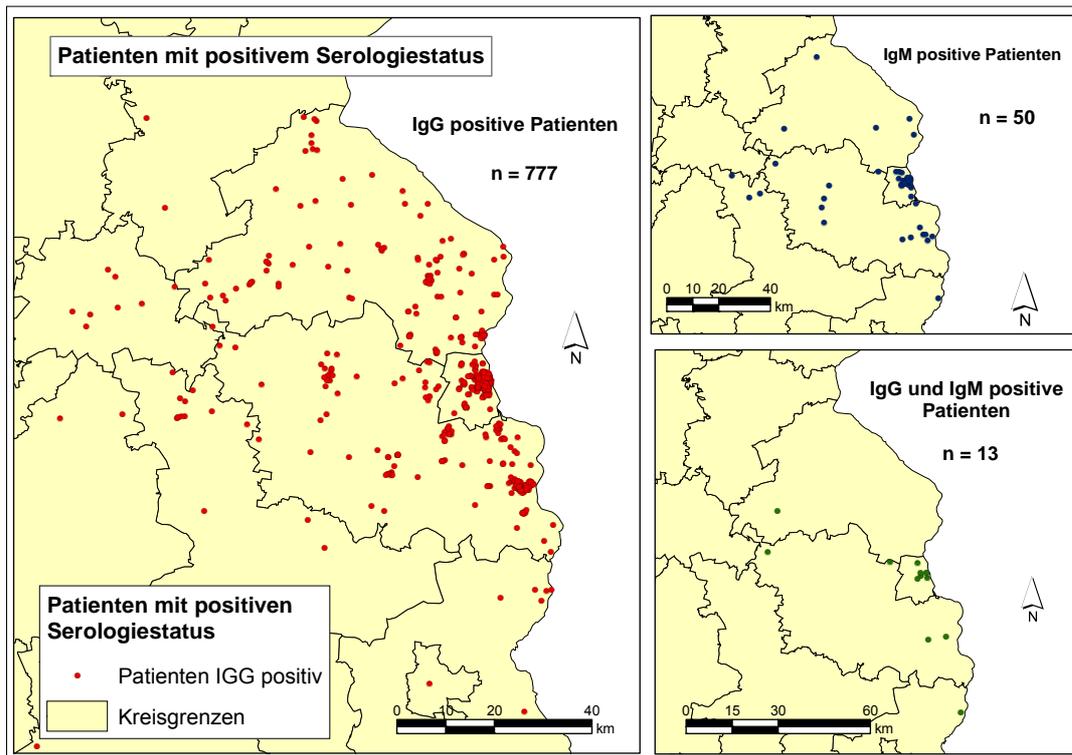


Abb. 35: Patienten mit positiven Serostatus

In Abb. 36 sind die Patienten mit positivem Serostatus nach dem Zeitpunkt der Infektion dargestellt, der wie in Kap. 4.1.1 erläutert semi-quantitativ erhoben wurde.

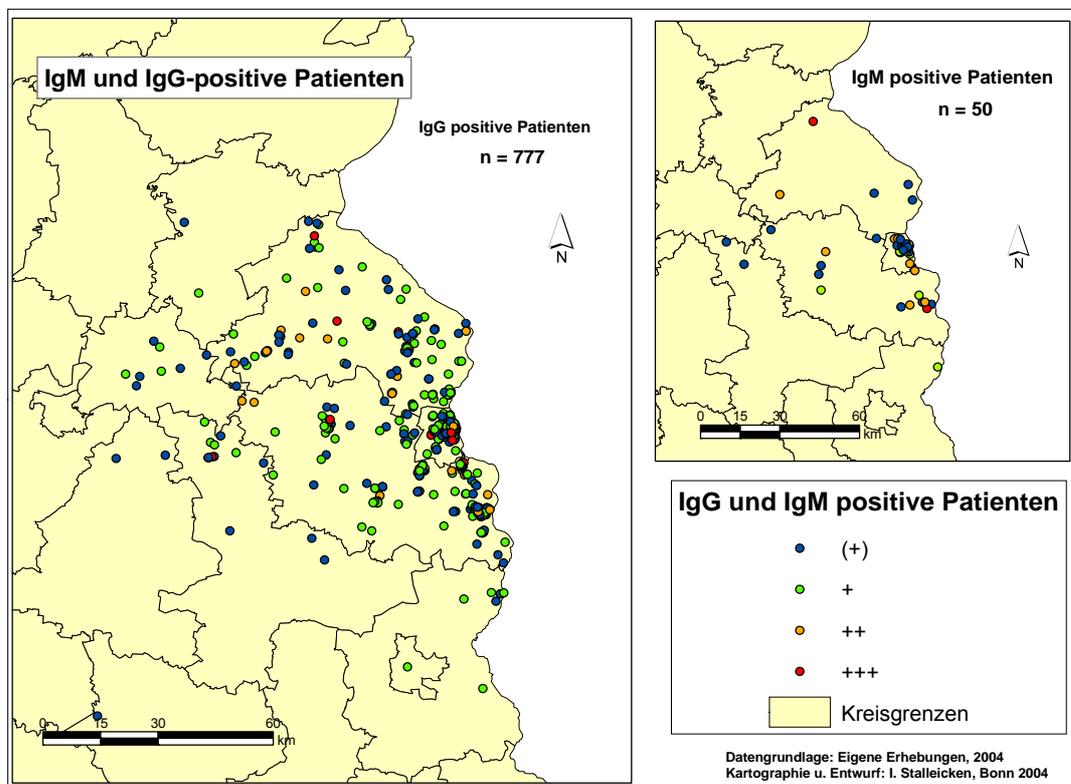


Abb. 36: IgM und IgG-positive Patienten nach Serotypen klassifiziert

Die Signatur „+++“ bedeutet in diesem Fall eine Infektion, die am wenigsten weit zurückliegt, hingegen die Signatur „(+)“ eine Infektion, die am weitesten zurückliegt. Die überwiegende Anzahl der Patienten hat sich bereits vor längerer Zeit infiziert. Dies ist sowohl für IgG als auch für IgM positive Patienten erkennbar.

Im Rahmen der statistischen Analyse wurden mit Hilfe des Fragebogens verschiedene Risikofaktoren ermittelt, die bei einer Legionelleninfektion von Bedeutung sind. In Abb. 37 sind nur solche serologisch positiven Patienten dargestellt, die angaben, sich regelmäßig in Sportanlagen und Sportstätten aufzuhalten.

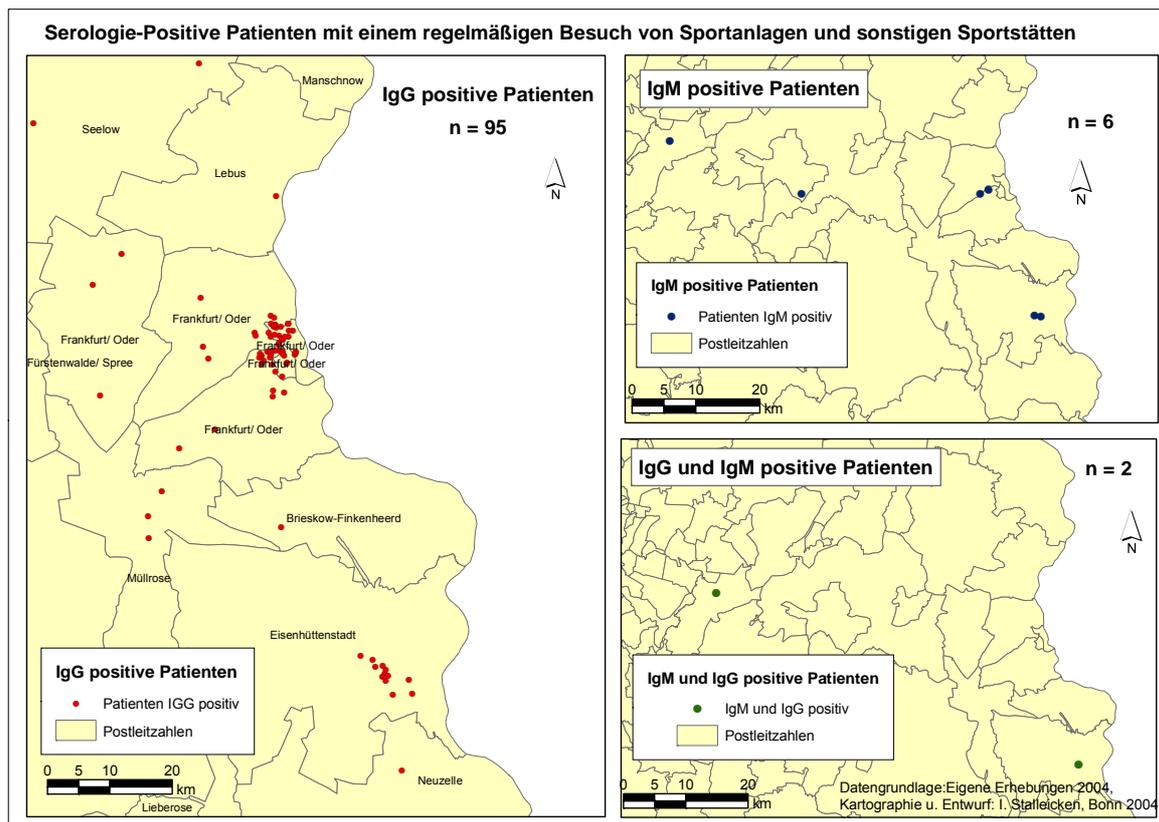


Abb. 37: Serologisch positive Patienten mit einem regelmäßigen Besuch von Sportanlagen und Sportstätten

Die Mehrzahl der IgG positiven Patienten, die regelmäßige Sportaktivitäten betreiben, kommen aus dem Raum Frankfurt (Oder) sowie Eisenhüttenstadt. Diese Tatsache gilt nicht für Patienten, die gleichzeitig IgM und IgG positiv sind.

5.3.2.1 Die Verteilung der serologisch positiver Patienten im Untersuchungsgebiet

Zur Analyse der Verteilung der Patienten im Untersuchungsgebiet wurde die Nächste-Nachbar Statistik und die K-Statistik nach Ripley herangezogen (Kap. 2.2.1). Es wurden jeweils für IgG positive und für IgM positive Patienten die räumliche Punkteverteilung bestimmt. Als Untersuchungsgebiet wurde die Größe des Landes Brandenburg zugrunde gelegt (etwa 30.611 km²), da dieses als wesentliches Einzugsgebiet des Krankenhauses gilt, oder als Alternative die maximale Ausdehnung des Punktmusters. Es wurde kein *edge-Effekt* berücksichtigt. Das Ergebnis der Berechnung der Nächste-Nachbar Statistik ist in Abb. 38 dargestellt:

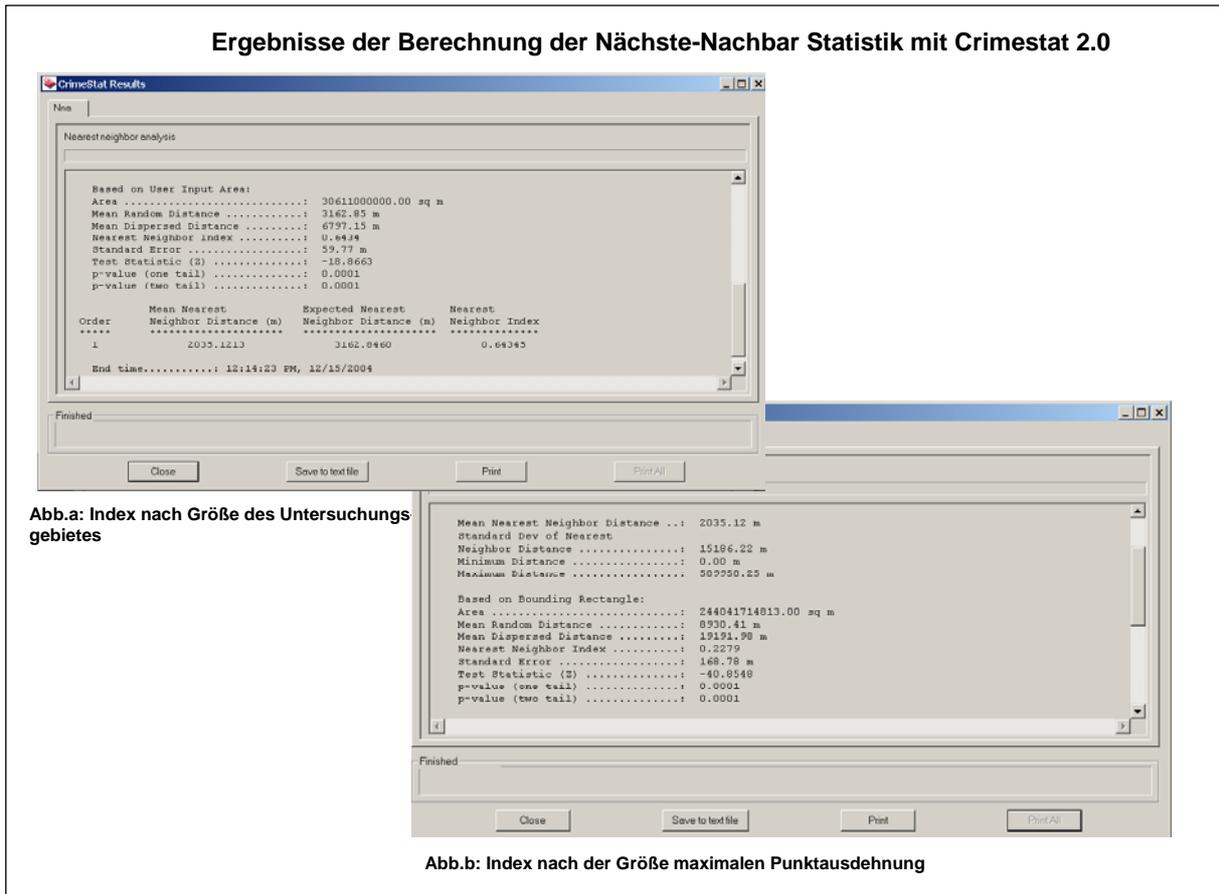


Abb. 38: Ergebnis der Berechnung der Nächste-Nachbar Statistik für IgG positive Patienten

Die Nächste-Nachbar Indizes der IgG positiven Patienten weichen je nach zugrunde gelegtem Untersuchungsgebiet leicht voneinander ab. Der Index, berechnet nach der Größe des Bundeslandes Brandenburg, beträgt 0,6434. Legt man die maximale Ausdehnung des Punktmusters (*based on bounded rectangle*) zugrunde, errechnet sich ein Index von 0,2279. Während der erste Index auf eine eher konzentrierte als zufällige Punkteverteilung hindeutet, tendiert der zweite Index zu einer konzentrierten Punkteverteilung. Es wird deutlich, wie stark der Einfluss des zugrunde gelegten Untersuchungsgebietes ist. Die Werte der Teststatistik (z) sind signifikant, da sie, absolut gesehen, über dem Wert eines 99%-Konfidenzintervalls von 2,33 liegen.⁹

Bei der Berechnung der Nächste-Nachbar Indexes der IgM positiven Patienten, wird ebenfalls die Größe des Bundeslandes Brandenburg als auch die des größten Punktmusters ermittelt (Abb. 39).

⁹ negative Werte der Teststatistik z werden als absolute Werte erkannt und dann mit dem Wert von 2,33 eines 99%-Konfidenzintervalls verglichen.

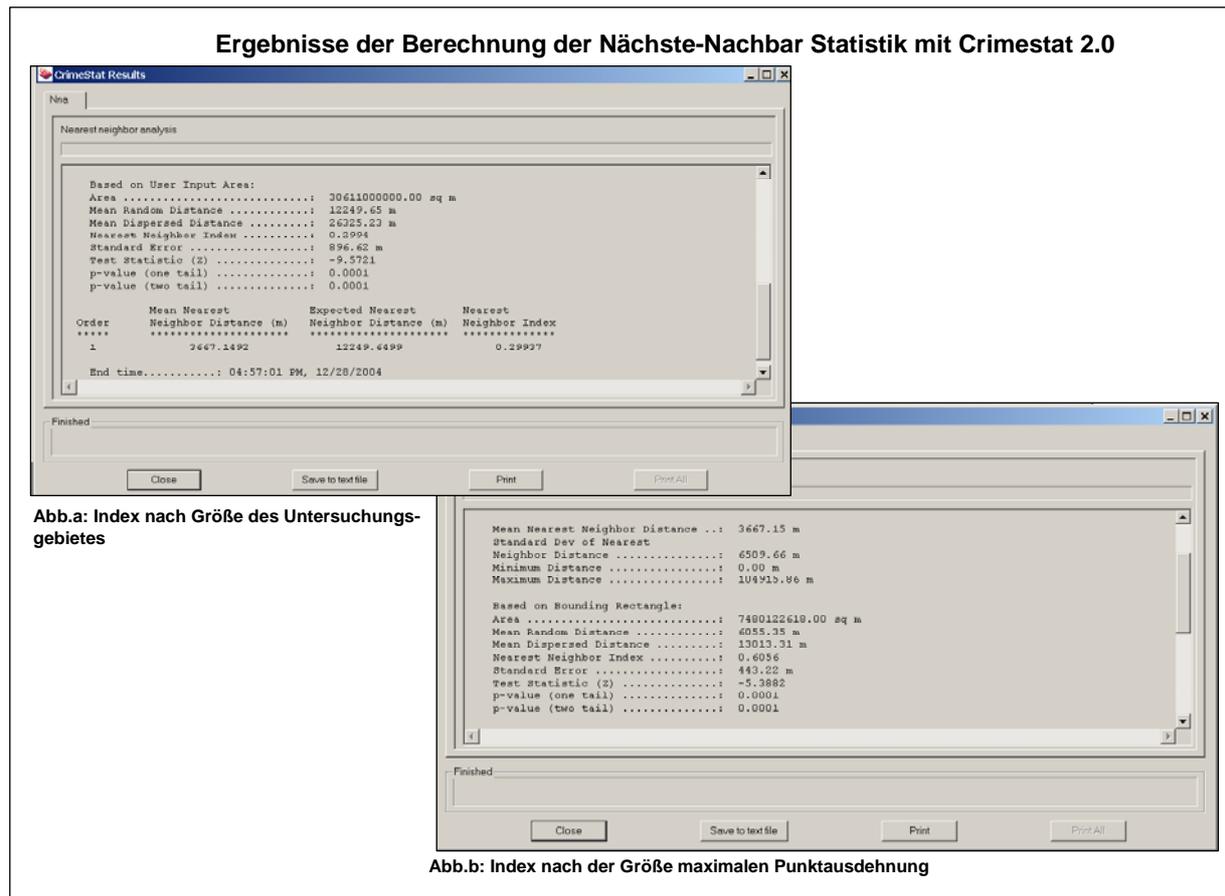


Abb. 39: Ergebnis der Berechnung der Nächste-Nachbar Statistik für IgM positive Patienten

Hier beträgt der Index 0,29937, berechnet in Abhängigkeit von der Größe des Bundeslandes Brandenburg, hingegen der Index „based on bounded rectangle“ 0,6056. Dies liegt darin begründet, dass die maximale Ausdehnung des Punktmusters in diesem Falle kleiner ist, als die der Fläche des Landes Brandenburg. Beide Ergebnisse deuten jedoch auf eine eher konzentrierte Punkteverteilung hin. Sie sind statistisch signifikant, da die berechneten Werte der z-Teststatistik über dem kritischen z-Wert eines 99%-Konfidenzintervalls von 2,33 liegen. Die Ergebnisse der K-Funktion nach Ripley für IgG positive Patienten sind in dem folgenden Diagramm (Abb. 40) dargestellt. Auch hier wurde ein Untersuchungsgebiet nach der Größe des Bundeslandes Brandenburgs zugrunde gelegt.

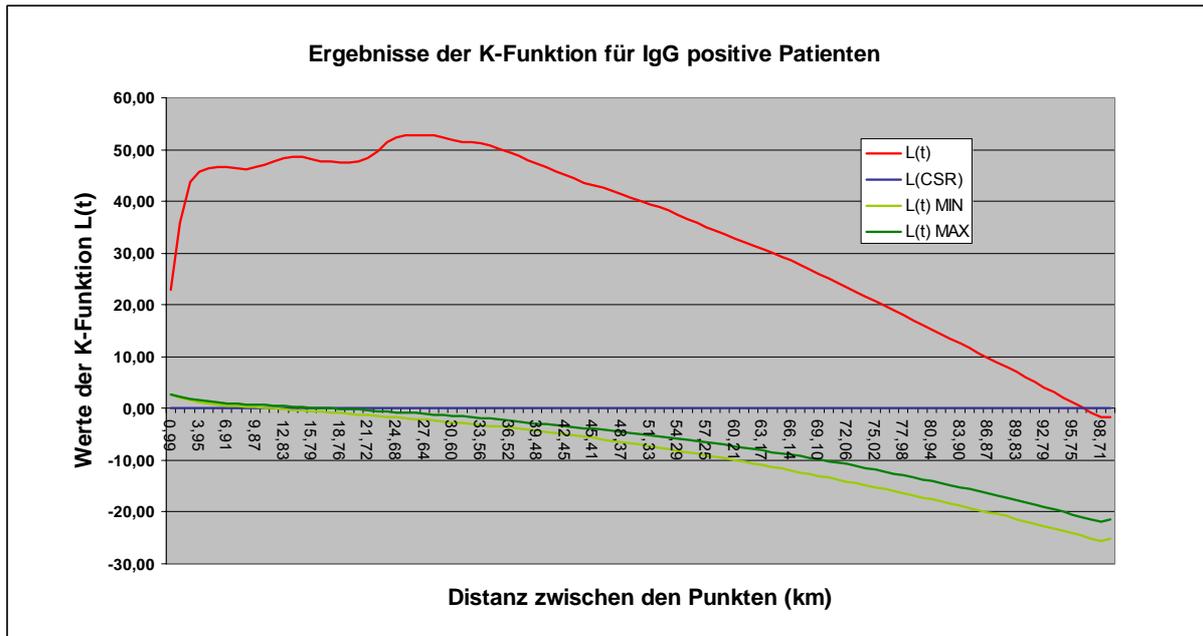


Abb. 40: Ergebnis der K-Funktion für IgG positive Patienten

Das beobachtete Punktmuster liegt oberhalb des Konfidenzbandes, was auf eine eher konzentrierte Verteilung hinweist. Diese Aussage entspricht somit der Aussage aus der Berechnung des Nächsten Nachbar Indexes.

Die Ergebnisse der K-Statistik für IgM positive Patienten sehen folgendermaßen aus (Abb. 41):

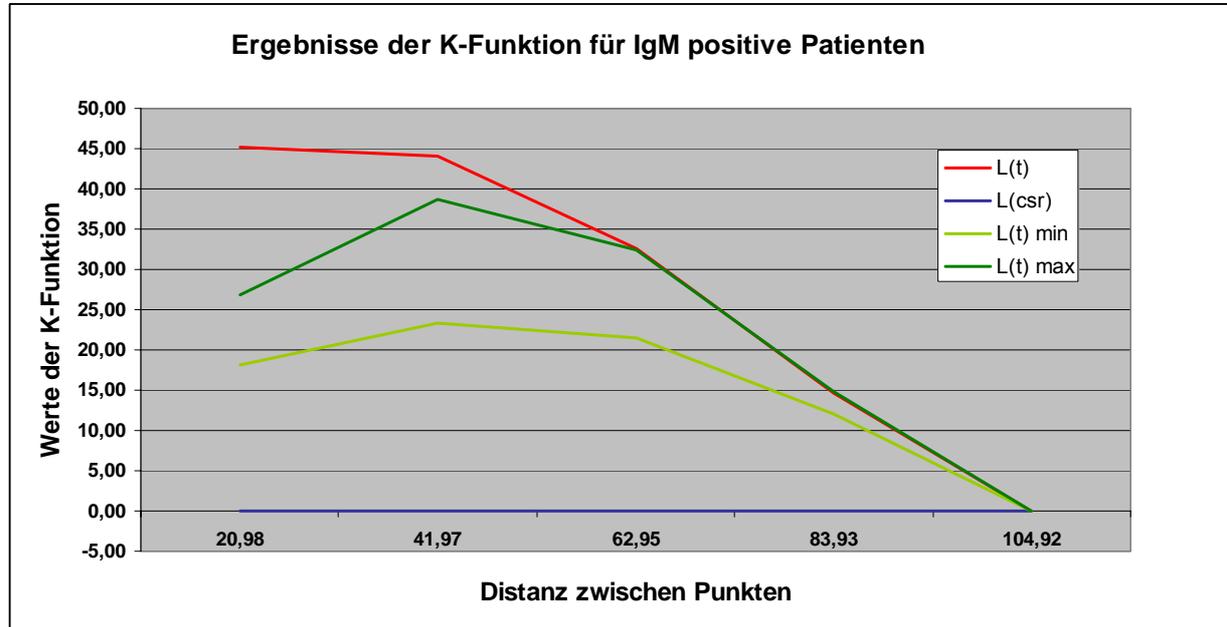


Abb. 41: Ergebnis der K-Funktion für IgM positive Patienten

Die $L(t)$ Werte der beobachteten Punkteverteilung liegen bis zu einer Distanz von etwa 65 km oberhalb des Konfidenzbandes. Über diese Entfernung hinaus gehen die Werte ähnlich denen des maximalen Konfidenzintervalls. Auch in diesem Fall deckt sich die Aussage mit der des Nächste-Nachbar Indexes, denn man kann auch in diesem Fall von einer eher konzentrierten Verteilung sprechen, wenn auch diese, beeinflusst durch die betrachtete Größe

des Untersuchungsgebietes, Tendenzen hin zu einer zufälligen Verteilung hat. Die tabellarischen Ergebnisse der Berechnungen werden ausführlich im Anhang dargestellt.

Die Aussagen der Nächste-Nachbar-Statistik und der K-Funktion decken sich bei den hier durchgeführten Berechnungen. Im folgenden Kapitel soll nun aufbauend auf die Analyse der Verteilung der Patienten im Untersuchungsgebiet, die Dichte der Patienten berechnet werden, um erste Cluster oder Hotspots zu ermitteln.

5.3.2.2 Die Dichte serologisch positiver Patienten im Untersuchungsgebiet

Die räumliche Punktdichte wurde mit Hilfe der Kerne-Density Interpolation berechnet (ArcGIS 8) (Kap. 2.2.2). Die Aussagekraft einer solchen Punktdichteberechnung ist in großem Maße von den festzulegenden Parametern, wie z.B. dem Suchradius, abhängig. Für die IgG positiven Patienten wurden verschiedene Suchradien ermittelt. Diese unterscheiden sich je nach der betrachteten Größe des Untersuchungsgebietes (Abb. 42).

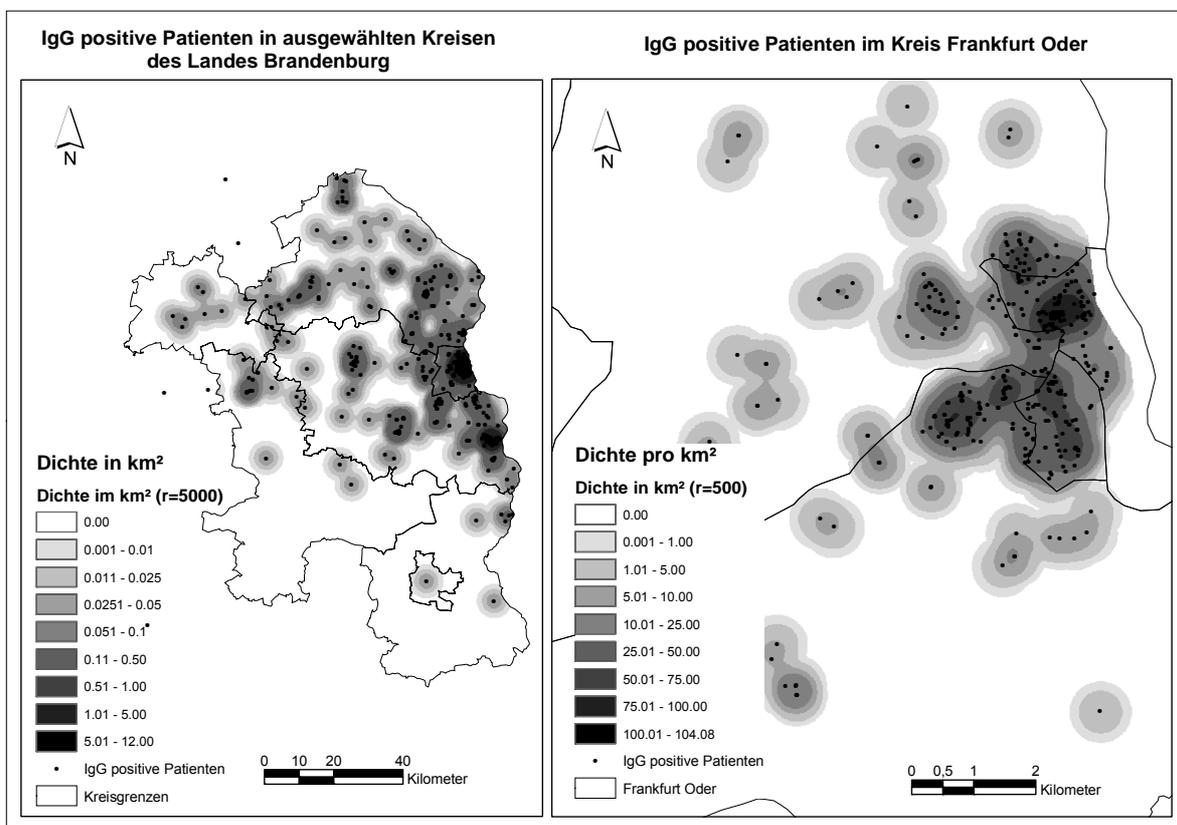


Abb. 42: Kernel-density Interpolation für IgG positive Patienten (Radius in Metern)

In der linken Karte wurde die Dichteberechnung für das Land Brandenburg durchgeführt. Entsprechend sorgfältig musste hier der Suchradius festgelegt werden (im Beispiel 5000 m): Bei zu kleinen Radien trat der sogenannte „Bullaugeneffekt“ hervor, bei zu großen Radien waren kaum Unterschiede in der Dichteberechnung erkennbar. Problematisch bei der Errechnung der Dichte IgG positiver Patienten war die enorme Konzentration im Raum Frankfurt (Oder). So schwankt die Dichte IgG positiver Patienten im gesamten Untersuchungsgebiet von weniger als 0,001 bis hin zu maximal 12 Patienten pro km². Ursache ist die insgesamt sehr hohe Patientenzahl. In der rechten Karte ist daraufhin die Dichteberechnung expli-

zit für die Stadt Frankfurt (Oder) durchgeführt worden (Suchradius 500 m). Die Ergebnisse ergeben in diesem Fall ein detaillierteres Bild hinsichtlich der Dichte IgG positiver Patienten. Die folgende Abbildung zeigt das Ergebnis der Dichteberechnung IgM positiver Patienten unter Verwendung verschiedener Suchradien (Abb. 43)

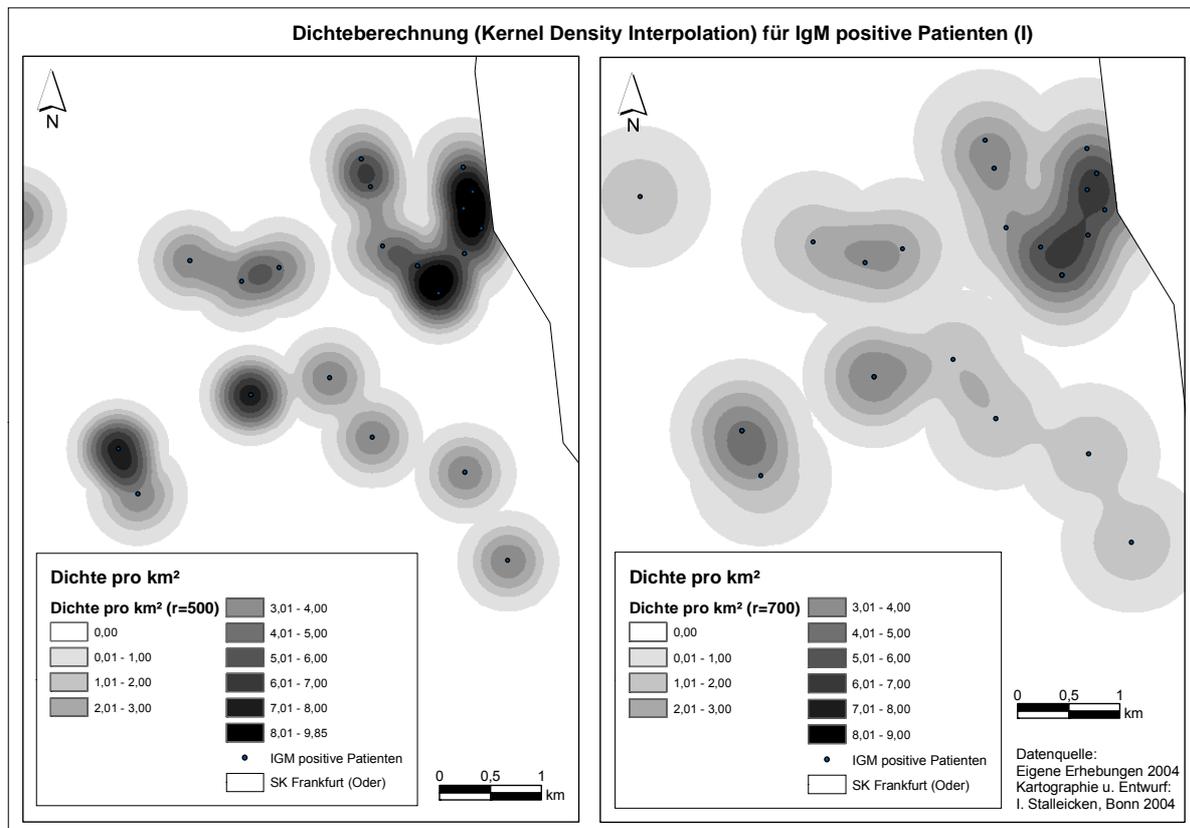


Abb. 43: Kernel-density-Interpolation für IgM positive Patienten (verschiedene Suchradien I)

Die Dichteberechnung erfolgte auch in diesem Beispiel für die Stadt Frankfurt (Oder). Die Anzahl IgM positiver Patienten ist deutlich geringer als die der IgG Positiven. Dementsprechend deutlich macht sich bei einer Dichteberechnung eine Veränderung des Suchradius bemerkbar (Abb. 43). In der linken Karte wurde ein Suchradius von 500 m verwendet. So lassen sich maximale Dichten von 9,85 pro km² ermitteln. Bei einem Radius von 700 m beträgt die maximale Dichte nur noch 6,71 pro km²¹⁰. Bei einem zusehends geringeren Suchradius erhöht sich die Dichte pro km², allerdings verstärkt sich auch der „Bullaugeneffekt“ (vgl. Abb. 44).

¹⁰ zur besseren Vergleichbarkeit wurde die Legende angepasst

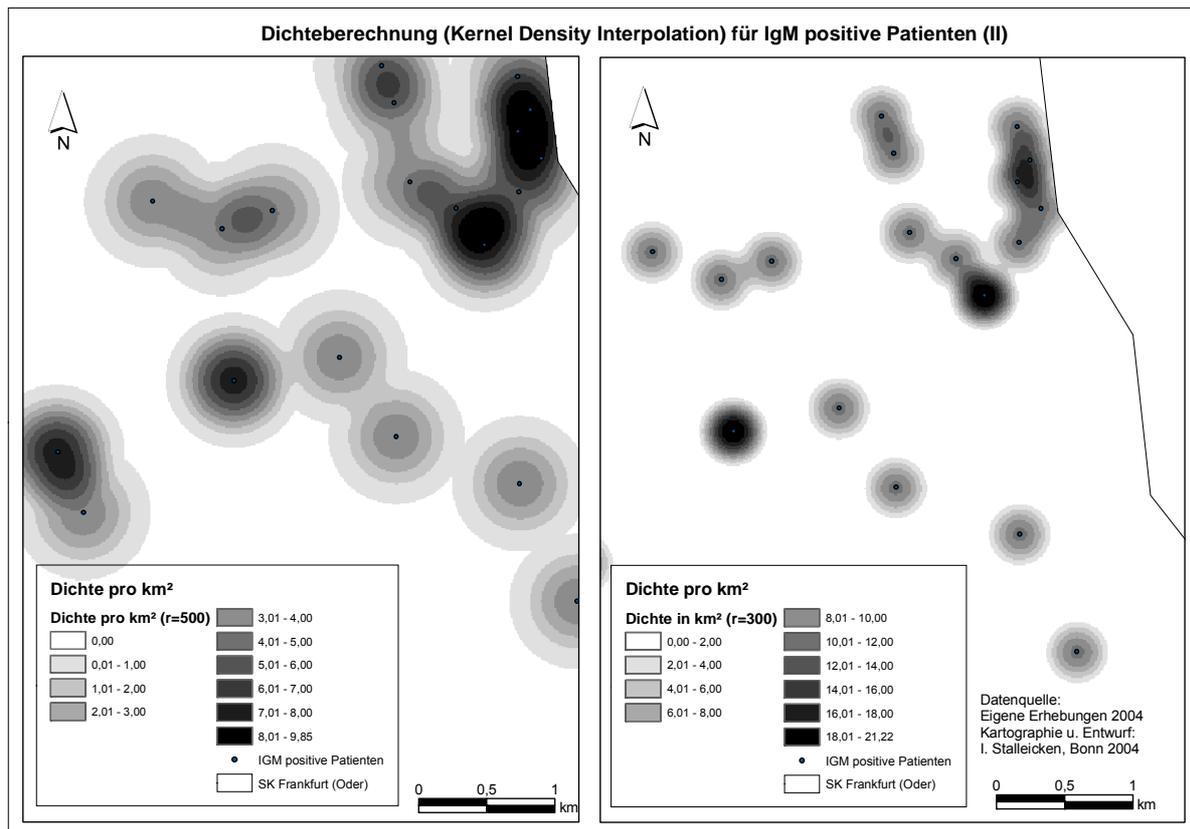


Abb. 44: Kernel-density-Interpolation IgM positiver Patienten (verschiedene Suchradien II)

Die Dichteberechnungen sowohl für IgG positive als auch für IgM positive Patienten gibt einen ersten Überblick darüber, welche Regionen im Einzugsgebiet des Klinikums eine besonders hohe Dichte an Legionelleninfektionen aufweisen. Aufgrund der starken Konzentration der Patienten im Raum Frankfurt (Oder) ist diese Methode problematisch. Zur Identifikation von Krankheitshäufungen wird daher im Folgenden auf hierarchische Clustermethoden zurückgegriffen.

5.3.2.3 Identifikation von Clustern serologisch positiver Patienten

Die Identifikation von Clustern erfolgte, wie in Kap. 4.3.3.2 ausführlich beschrieben, anhand der Nearest neighbour hierarchical clustering Methode und deren Modifikationen sowohl für IgG positive als auch für IgM positive Patienten.

Die erste Berechnung erfolgte anhand der Nearest neighbour hierarchical clustering Methode unter den in Abb. 45 dargestellten Standardeinstellungen (min= 10 Punkte, Standardabweichung 1, Wahrscheinlichkeit 50%). Die Größe des Untersuchungsgebietes entspricht erneut dem Bundesland Brandenburg. Das Ergebnis der Berechnung für IgG positive Patienten wird in CrimeStat 2.0 wie folgt dargestellt:

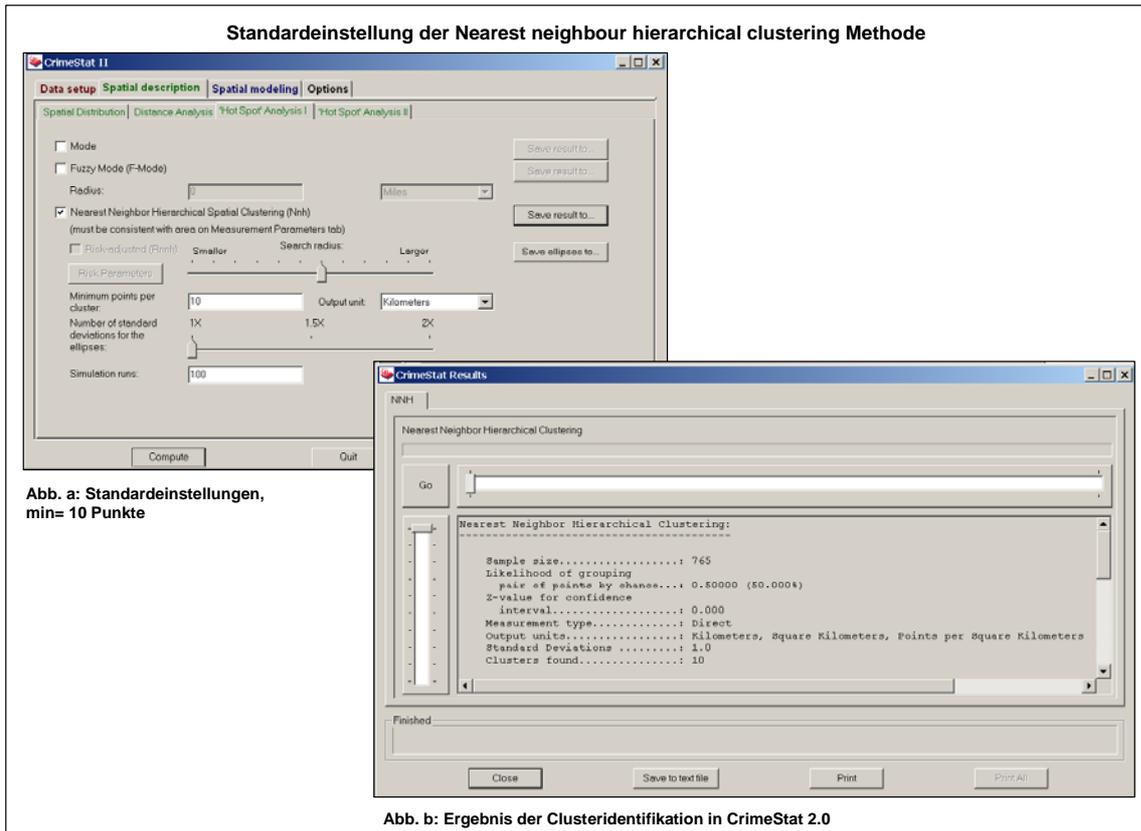


Abb. a: Standardeinstellungen, min= 10 Punkte

Abb. b: Ergebnis der Clusteridentifikation in CrimeStat 2.0

Abb. 45: Standardeinstellungen der NNH-Methode und deren Ergebnis

In Abb. 45 ist ersichtlich, dass unter Verwendung der Standardeinstellungen zehn Cluster identifiziert wurden. Die Ellipsen der primären und sekundären Cluster sind in Abb. 46 dargestellt:

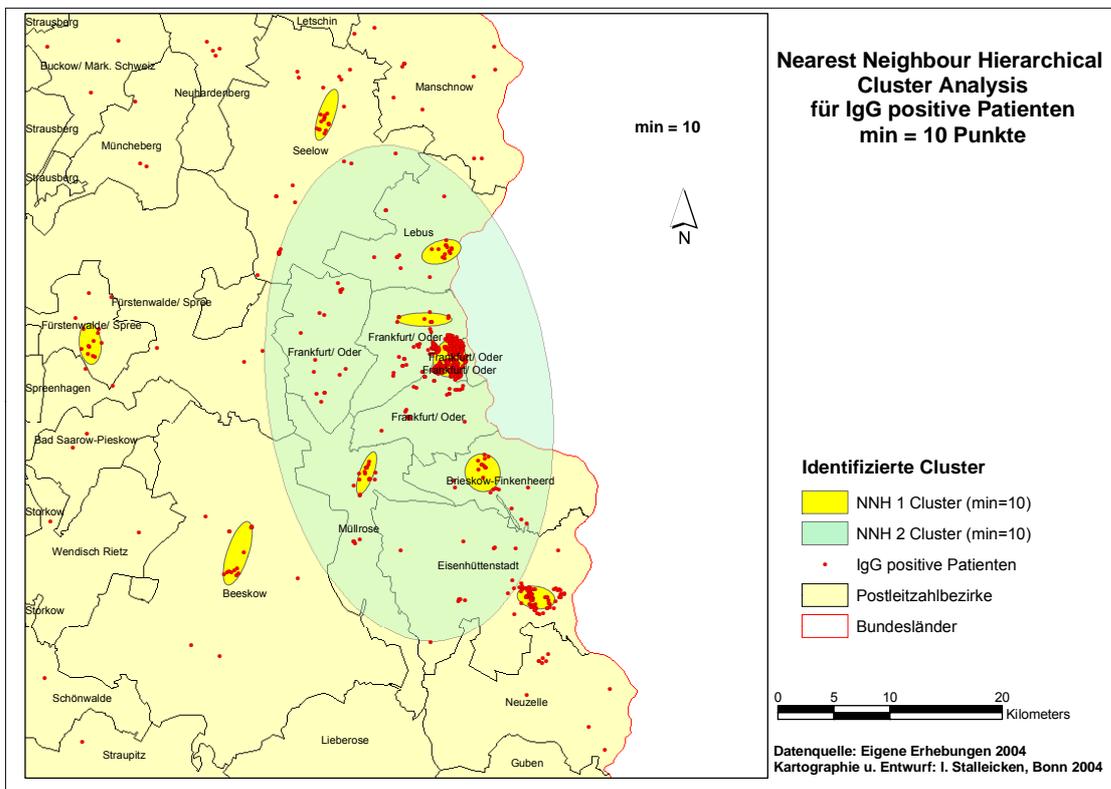


Abb. 46: NNH-Clustering, Minimum = 10 Punkte für IgG positive Patienten

Das Ergebnis der Clusterberechnung ist in Tab. 7 dargestellt.

Tab. 7: Ergebnis der Clusterberechnung IgG positiver Patienten (min=10)

ORDER	CLUSTER	MEAN X	MEAN Y	ROTATION	X-AXIS	Y-AXIS	AREA	POINTS
1	1	4672821,06	5803611,55	17,72	1,41	1,66	7,38	379
1	2	4670611,82	5807119,82	89,83	0,60	2,43	4,55	11
1	3	4680541,48	5782160,89	9,10	1,69	0,99	5,24	79
1	4	4665442,47	5793329,06	19,98	0,62	2,02	3,91	17
1	5	4661892,41	5825510,71	18,27	0,73	2,41	5,51	17
1	6	4675779,76	5793368,47	77,31	1,70	1,56	8,35	17
1	7	4653969,07	5786130,67	18,01	0,95	2,98	8,86	15
1	8	4672107,00	5813204,53	71,57	1,00	1,80	5,67	15
1	9	4640830,62	5804945,92	85,02	1,85	0,97	5,64	13
2	1	4669145,63	5800554,46	82,41	22,39	12,61	886,86	8

Auch mit dieser Methode ist ein großes Cluster im Bereich Frankfurt (Oder) erkennbar, welches insgesamt 379 IgG positive Patienten umfasst. Darüber hinaus sind vereinzelte Cluster in weiteren Teilen des Bundeslandes Brandenburg erkennbar, darunter in Lebus, Brieskow-Finkenheerd, Eisenhüttenstadt u.a. Bei der Evaluation der Ergebnisse anhand der Monte-Carlo Simulation konnte kein Cluster auf der Basis einer Zufallsverteilung gefunden werden, d.h. die Cluster aus der Nnh-Routine unterliegen nicht dem Zufall. Inwieweit diese Clusteridentifikation mit einem Minimum von zehn Patienten je Cluster mit der Kernel density Interpolation übereinstimmt, ist in Abb. 47 erkennbar:

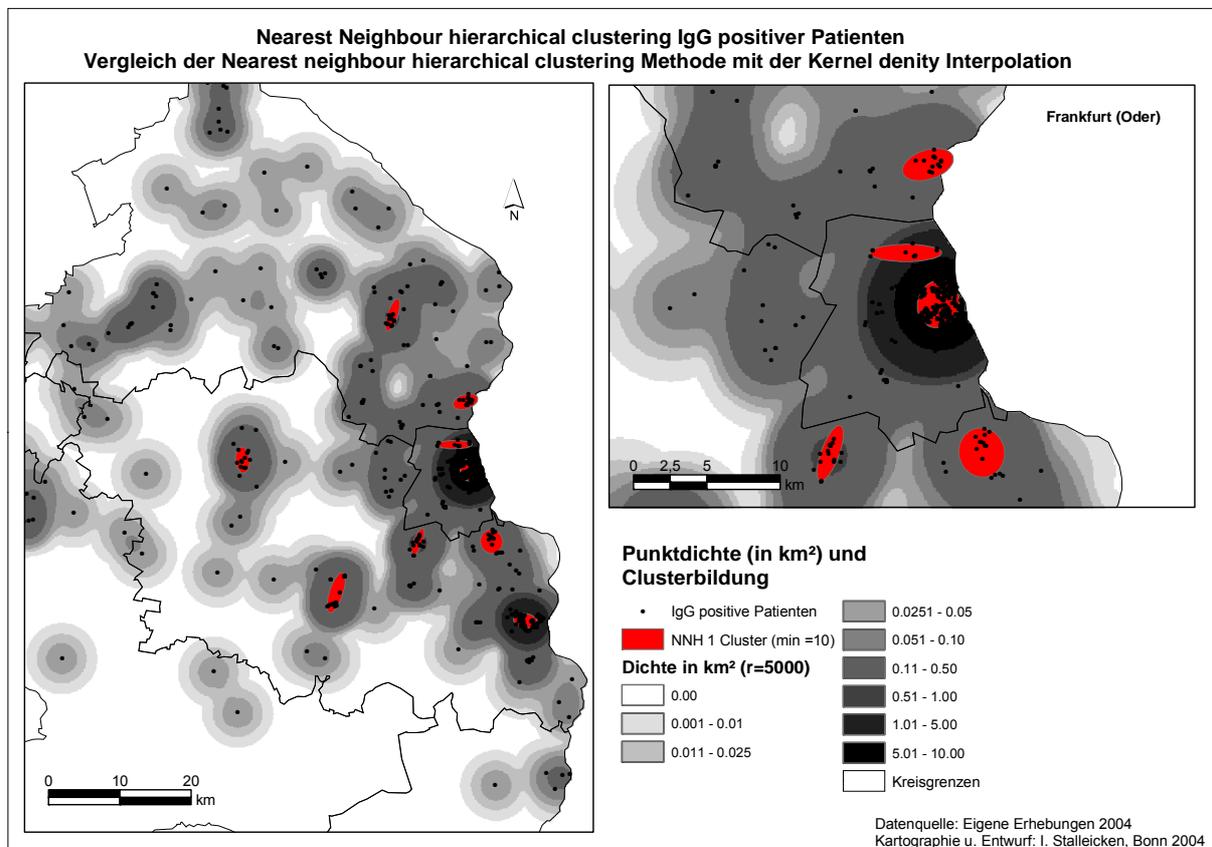


Abb. 47: Vergleich der NNh-Methode mit der Kernel-density-Interpolation für IgG positive Patienten

In Abb. 47 ist erkennbar, dass wesentliche primäre Cluster der Nnh-Methode mit den Punktdichteberechnungen übereinstimmen. Jedoch ist die Nnh-Methode bei einem Minimum von

zehn Patienten pro Cluster insbesondere im Bereich von Frankfurt (Oder) etwas genauer und differenziert dort einzelne Cluster.

Reduziert man nun die Mindestanzahl Patienten pro Cluster auf fünf, werden 15 primäre Cluster und ein sekundäres Cluster identifiziert (Abb. 48)

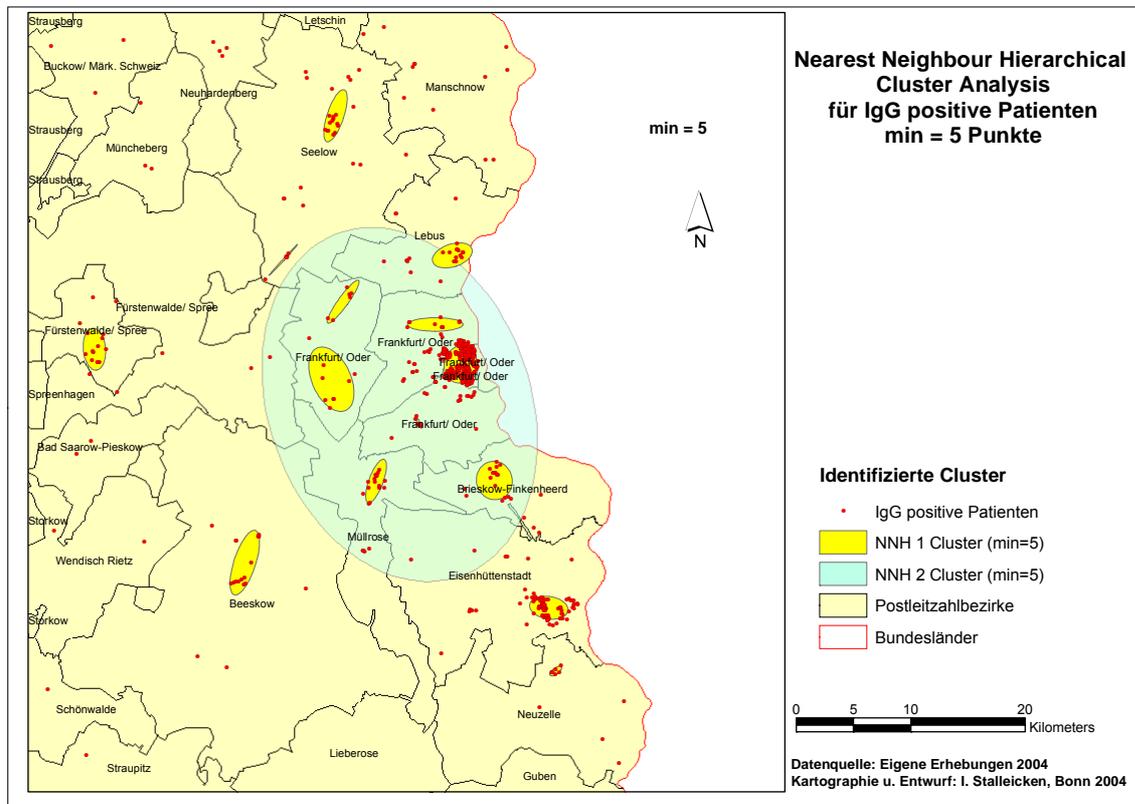


Abb. 48: NNH-Clustering, Minimum = 5 Punkte für IgG positive Patienten

In Abb. 48 erhöht sich die Anzahl der Cluster, insbesondere in einzelnen Bezirken von Frankfurt (Oder), aber auch in Bezirken, wie z.B. Neuzelle. Die ausführlichen Resultate der Clusterberechnung sind in Tab. 8 dargestellt:

Tab. 8: Ergebnis der Clusterberechnung IgG positiver Patienten (min=5)

ORDER	CLUSTER	MEAN X	MEAN Y	ROTATION	X-AXIS	Y-AXIS	AREA	POINTS
1	1	4672821,06	5803611,55	17,72	1,41	1,66	7,38	379
1	2	4670611,82	5807119,82	89,83	0,60	2,43	4,55	11
1	3	4680541,48	5782160,89	9,10	1,69	0,99	5,24	79
1	4	4665442,47	5793329,06	19,98	0,62	2,02	3,91	17
1	5	4661892,41	5825510,71	18,27	0,73	2,41	5,51	17
1	6	4675779,76	5793368,47	77,31	1,70	1,56	8,35	17
1	7	4653969,07	5786130,67	18,01	0,95	2,98	8,86	15
1	8	4672107,00	5813204,53	71,57	1,00	1,80	5,67	15
1	9	4640830,62	5804945,92	85,02	1,85	0,97	5,64	13
1	10	4669172,25	5798469,63	63,48	0,52	0,11	0,18	8
1	11	4661540,00	5802292,00	65,73	3,02	1,69	16,09	8
1	12	4662563,83	5809128,83	35,56	0,44	2,31	3,18	6
1	13	4657346,50	5812819,50	42,41	0,07	1,84	0,39	6
1	14	4681142,17	5776581,50	53,41	0,27	0,61	0,51	6
1	15	4610614,50	5796837,00	23,29	0,81	2,29	5,84	6
2	1	4667445,02	5800148,63	70,77	16,01	11,45	575,78	11

Die Monte Carlo Simulation dieser Clusterberechnung ergab maximal 5 Cluster (95% Perzentil = 2 Cluster), die auf der Basis einer zufälligen Verteilung identifiziert werden (Abb. 49)

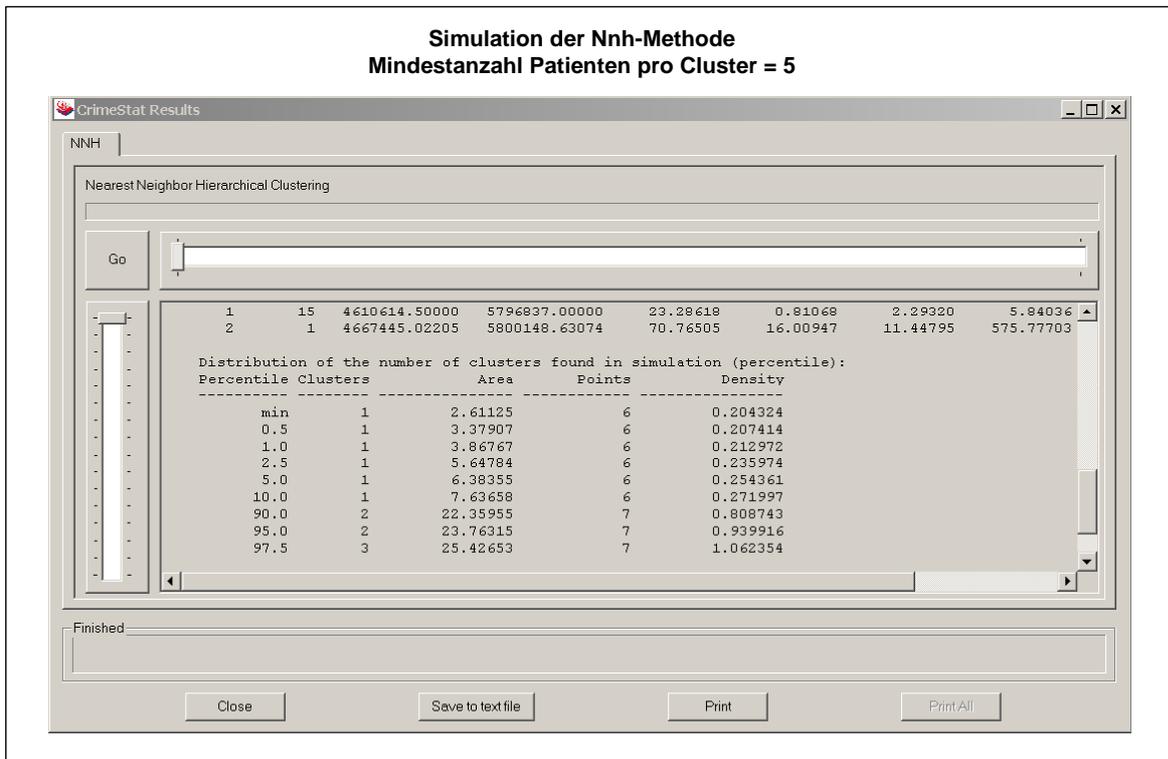


Abb. 49: Simulation der NNH-Methode (Mindestanzahl Patienten pro Cluster = 5)

Erhöht man nun die Mindestanzahl der Patienten je Cluster auf 15, wird folgendes Resultat erzielt:

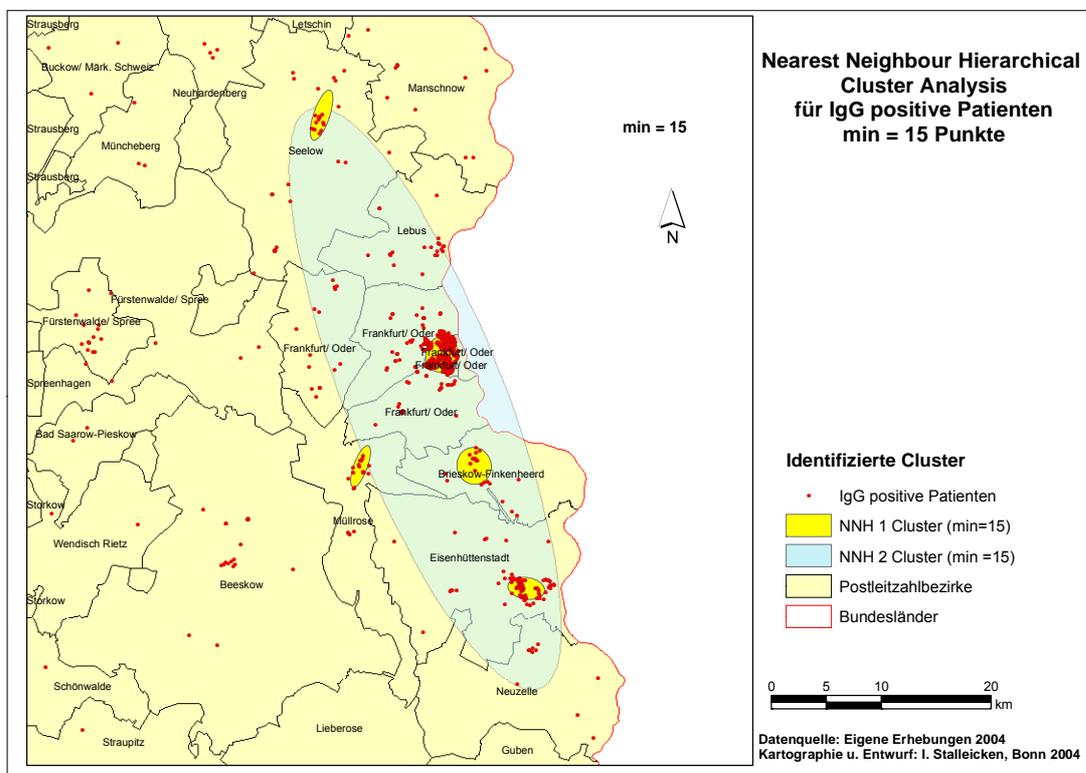


Abb. 50: NNH-Clusterberechnung, Minimum = 15 Punkte für IgG positive Patienten

Es verbleiben bei einer Mindestanzahl von 15 Patienten noch sechs Cluster, und zwar weiterhin eines im Raum Frankfurt (Oder), weitere in Eisenhüttenstadt, Brieskow-Finkenheerd, Seelow und Müllrose. Das detaillierte Clusterergebnis ist in folgender Tabelle dargestellt:

Tab. 9: Ergebnis der Clusterberechnung IgG positiver Patienten (min=15)

ORDER	CLUSTER	MEAN X	MEAN Y	ROTATION	X AXIS	Y AXIS	AREA	POINTS
1	1	4672821,06	5803611,55	17,72	1,41	1,66	7,38	379
1	2	4680541,48	5782160,89	9,10	1,69	0,99	5,24	79
1	3	4665442,47	5793329,06	19,98	0,62	2,02	3,91	17
1	4	4661892,41	5825510,71	18,27	0,73	2,41	5,51	17
1	5	4675779,76	5793368,47	77,31	1,70	1,56	8,35	17
2	1	4671295,44	5799596,13	68,55	28,55	7,18	644,11	5

Erhöht man weiter die Mindestanzahl Patienten pro Cluster auf 20, verbleiben lediglich zwei Cluster, eines im Stadtgebiet Frankfurt (Oder) und ein weiteres in Eisenhüttenstadt (Abb. 51)

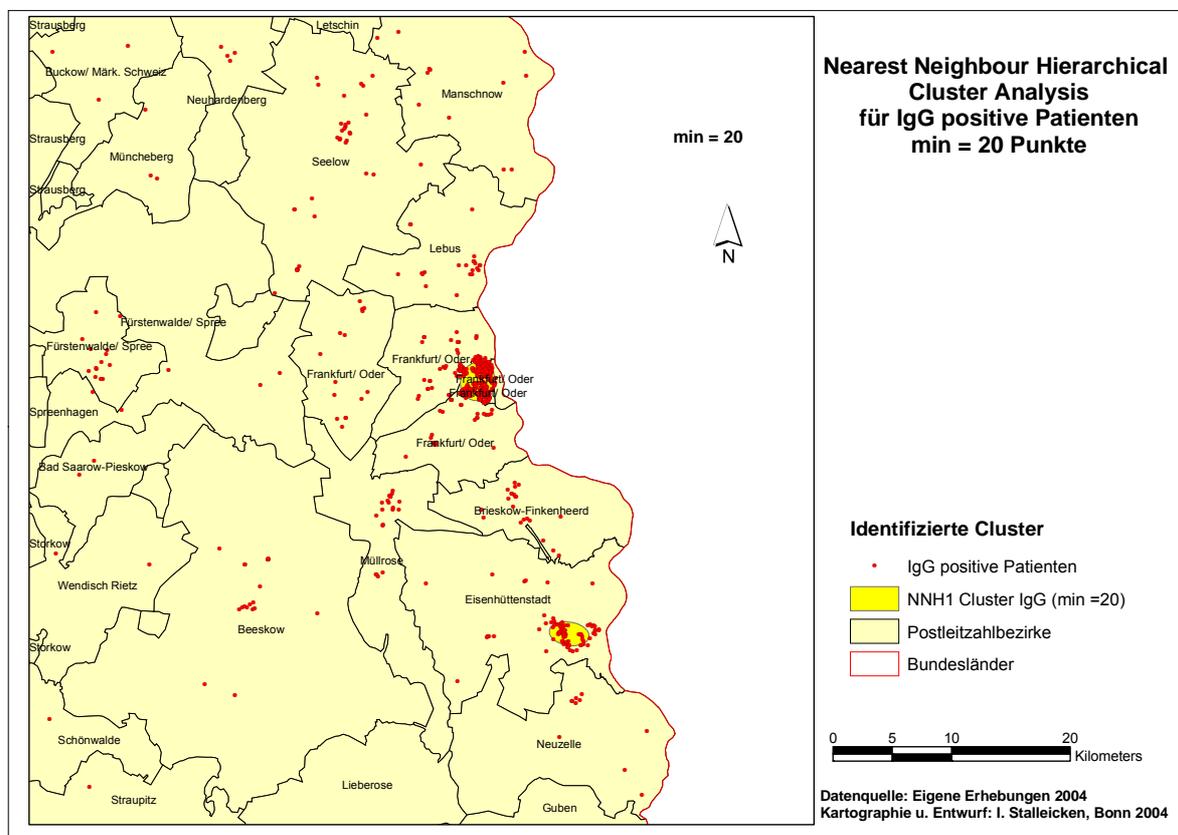


Abb. 51: NNH-Clusterberechnung, Minimum = 20 Punkte für IgG positive Patienten

Neben der Mindestanzahl Patienten pro Cluster ist es möglich, den Wahrscheinlichkeitsgrad der Auswahl zweier Punkte auf der Basis einer Zufallsverteilung anhand der „*threshold scale bar*“, zu variieren. Standardeinstellung ist das 50% Konfidenzintervall. Erhöht man dieses auf 99,9%, erhöht sich die Clustergröße um eins auf 17 Cluster (Abb. 52)

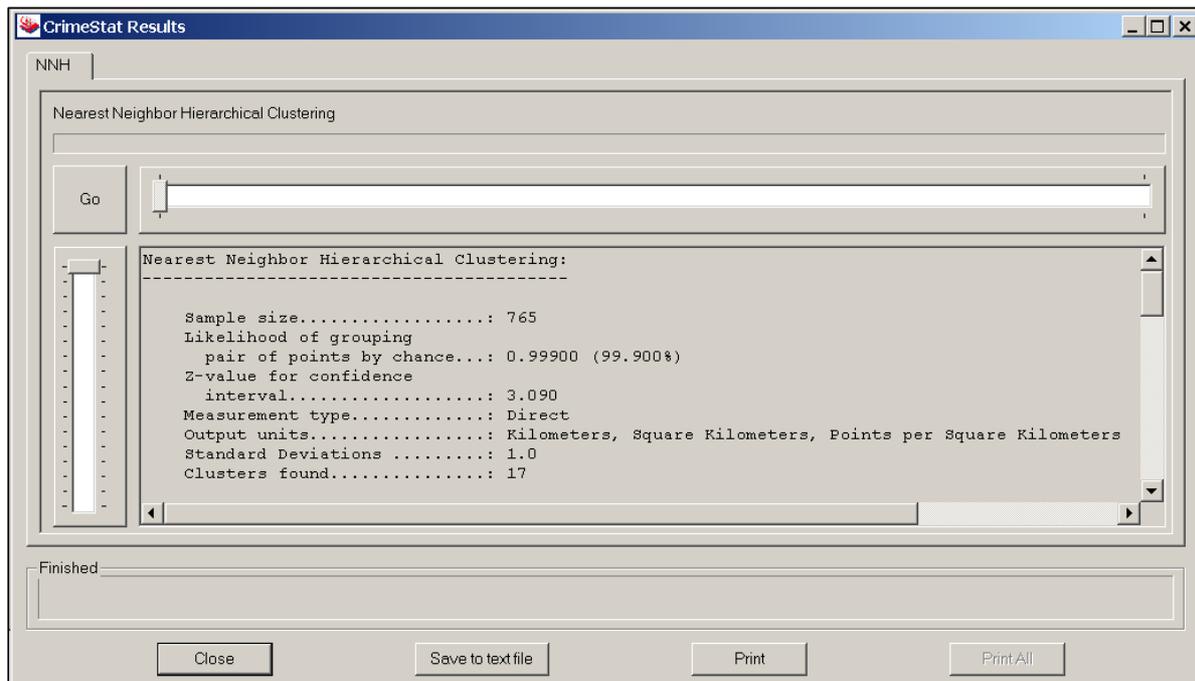


Abb. 52: Variation des Wahrscheinlichkeitsgrades auf 99,9%

Die Berechnung für IgM positive Patienten erfolgt zunächst für mindestens zehn Patienten pro Cluster. In diesem Fall wird lediglich ein Cluster identifiziert (Abb. 53).¹¹

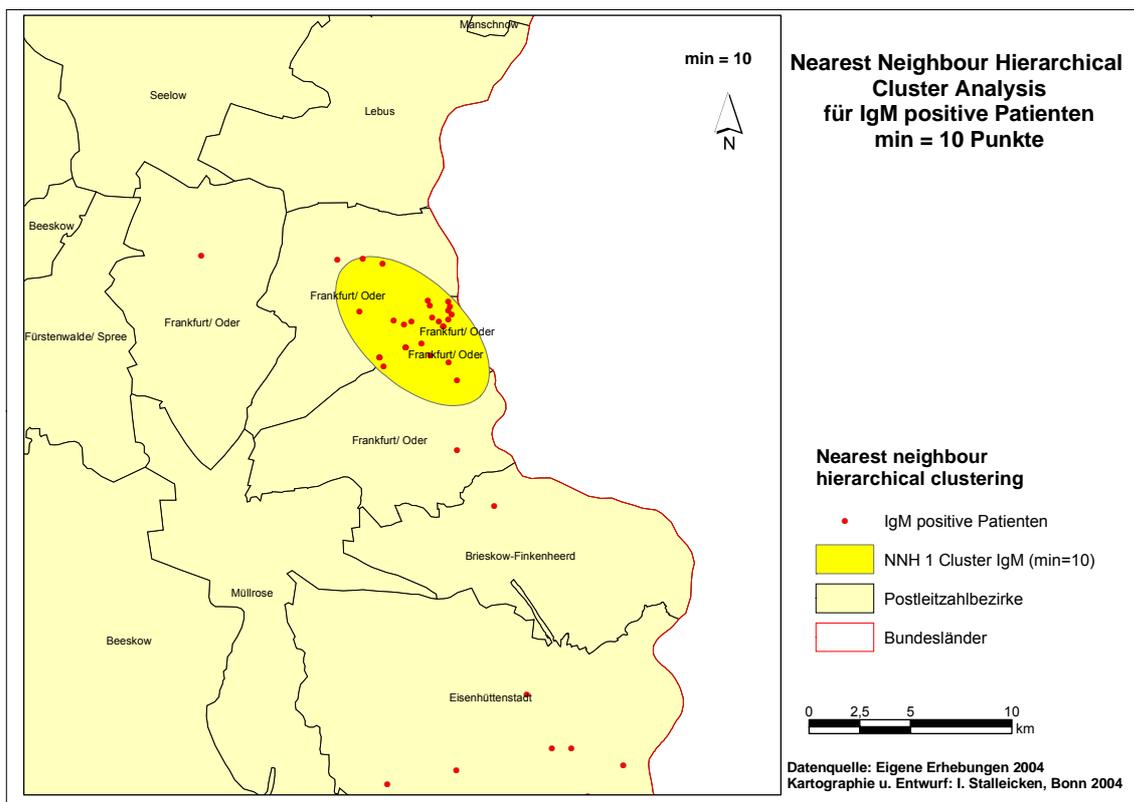


Abb. 53: NNH-Clusterberechnung, Minimum = 10 Punkte für IgM positive Patienten

¹¹ alle weiteren Berechnungen werden mit dem 50%-Wahrscheinlichkeitsgrad durchgeführt

Das einzig identifizierbare Cluster befindet sich erwartungsgemäß in Frankfurt (Oder). Verringert man die Mindestanzahl der Patienten auf 5, werden zwei Cluster erkannt (Abb. 54).

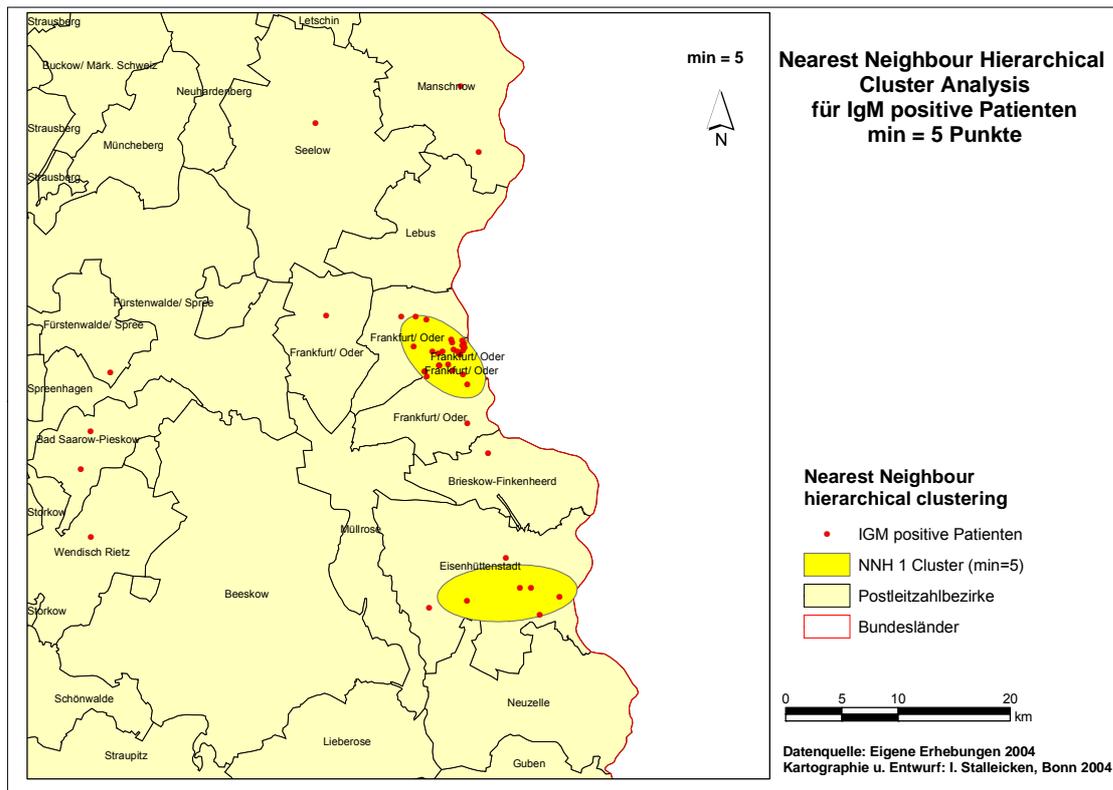


Abb. 54: NNH-Clusterberechnung, Minimum = 5 Punkte für IgM positive Patienten

Neben einem Cluster im Bereich Frankfurt (Oder) wird ein weiteres im Raum Eisenhüttenstadt identifiziert. Auch im Falle der IgM positiven Patienten wurden die Wahrscheinlichkeitsgrade variiert, was jedoch zu keinen Veränderungen der Anzahl Cluster führte.

Für IgG positive Patienten, die angaben, sich regelmäßig in Schwimmbädern oder sonstigen Sportstätten aufzuhalten, wurde ebenfalls eine Clusteranalyse auf Basis des *Nearest neighbour hierarchical clusterings* durchgeführt. Bei einer Anzahl von 95 Patienten, die den regelmäßigen Aufenthalt in Schwimmbädern und sonstigen Sportstätten bejahten und einer Mindestanzahl von fünf Patienten pro Cluster konnten zwei Cluster identifiziert werden (Abb. 55). Eines dieser beiden Cluster befindet sich in drei verschiedenen Bezirken der Stadt Frankfurt (Oder), ein weiteres in Eisenhüttenstadt. Erhöht man die Mindestanzahl Patienten je Cluster auf zehn, verbleiben ebenfalls zwei Cluster.

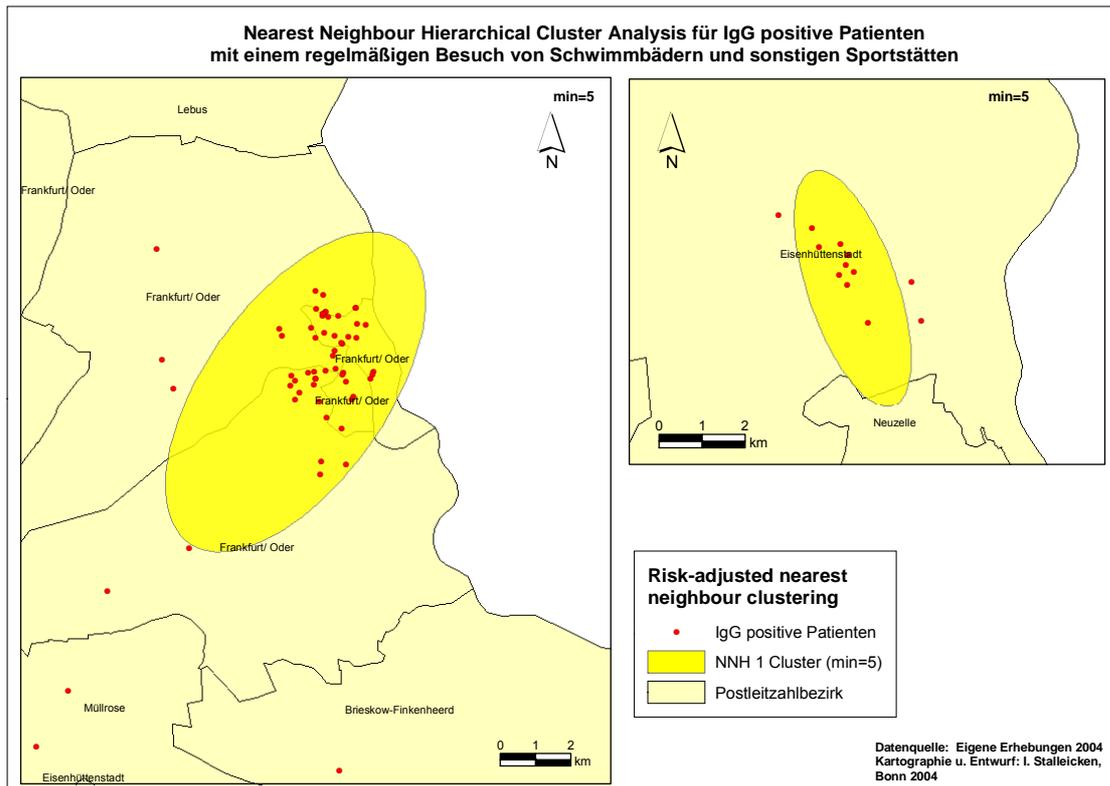


Abb. 55: NNH-Clusterberechnung, Minimum = 10 Punkte für IgG positive Patienten mit einem regelmäßigen Besuch von Schwimmbädern und sonstigen Sportstätten

Zusammenfassend befindet sich sowohl bei IgG positiven als auch bei IgM positiven Patienten der Schwerpunkt der Cluster im Raum Frankfurt (Oder) sowie im Raum Eisenhüttenstadt. Jedoch besteht auch bei der Methode der hierarchischen Clusteranalyse das Problem, dass die Wahrscheinlichkeit der Clusterbildung besonders hoch in Gebieten mit einer großen absoluten Patientenzahl ist. Aus diesem Grunde wurde bei den folgenden Berechnungen die absolute Patientenzahl als eine Basisvariable hinzugezogen. Dies geschieht anhand der sogenannten *Risk-Adjusted Nearest Neighbour Hierarchical Cluster Analysis*, einer Abwandlung der bisher verwendeten *Nearest neighbour hierarchical clustering Methode*, mit Hilfe der Software CrimeStat 2.0 (Kap. 4.3.3.2). Diese Methode ermöglicht die Ermittlung von Clustern im Verhältnis zu deren Grundgesamtheit, d.h. Regionen mit hoher oder niedriger Patientenzahl werden entsprechend bei der Identifikation von Clustern berücksichtigt (Levine 2002). Sie wurde zunächst für IgG positive Patienten angewendet. Auch hier konnte eine Vielzahl von Parametern angenommen werden und die Ergebnisse der Clusterberechnung unterscheiden sich mehr oder weniger deutlich. Zunächst wurden die Standardeinstellung soweit wie möglich beibehalten. Folgende Parameter wurden für die miteinander kombinierten Methoden (nearest neighbour hierarchical clustering und kernel density) angegeben (Tab. 10):

Tab. 10: Standardeinstellungen zur RNNH-Clustering Methode

Nearest Neighbour	Kernel density interpolation
Mindestanzahl Patienten pro Cluster 10	Methode der Interpolation: normal
Wahrscheinlichkeitsgrad: 50%	Suchradius: angepasst
Anzahl Standardabweichung der Ellipse: 1x	Mindestanzahl Punkte: 100
Untersuchungsgebiet: 30611 km ²	Angabe eines Rasters mit entsprechenden x- und y-Koordinaten

Folgendes Ergebnis wurde mit diesen Einstellungen erzielt (Abb. 56):

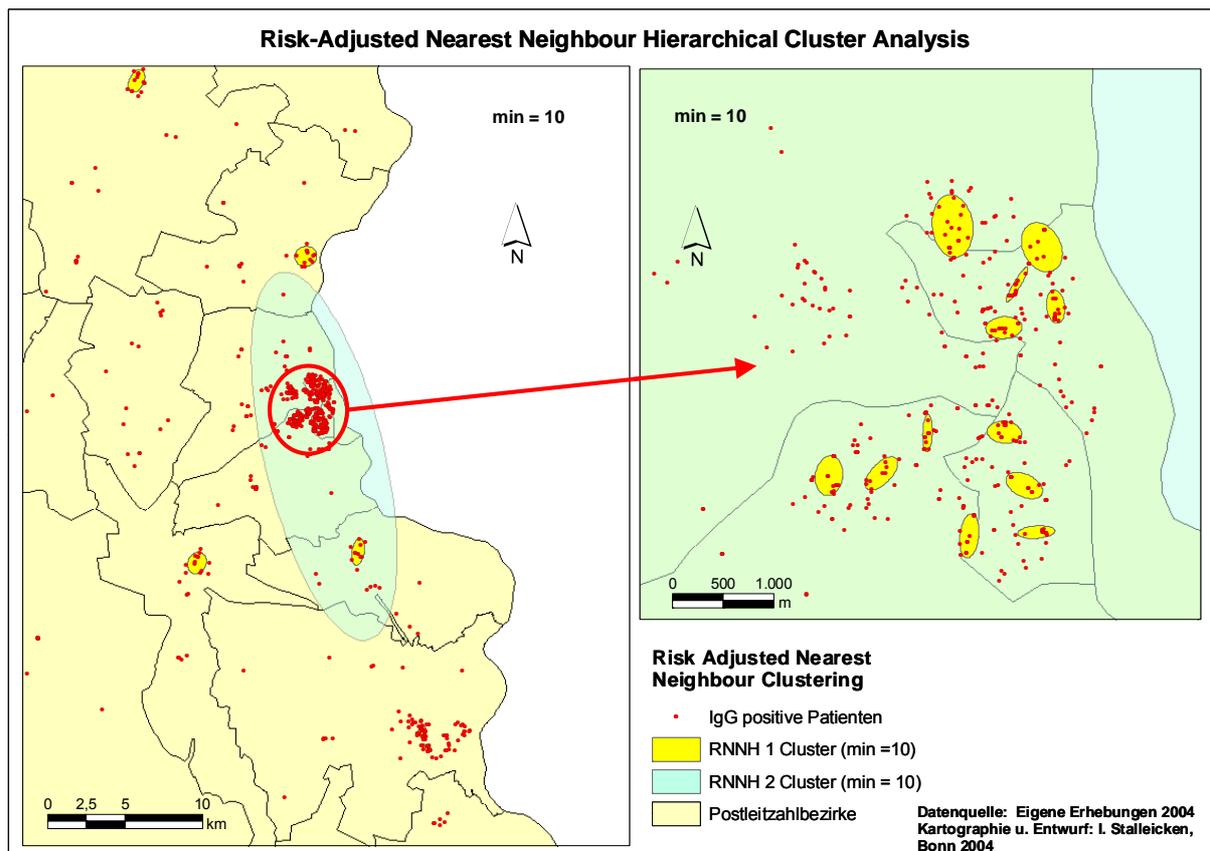


Abb. 56: RNNH-Clusterberechnung unter Berücksichtigung der absoluten Patientenzahl für IgG mit Standardeinstellungen (min=10)

Es wurden mit den genannten Parameterangaben 18 primäre Cluster und 1 sekundäres Cluster identifiziert, wobei insbesondere Konzentrationen im engeren Stadtgebiet von Frankfurt (Oder) auffällig sind. Dabei handelt es sich um die Postleitzahlbezirke 15230, 15232, 15234 und 15236 der Stadt Frankfurt (Oder). Weiter nördlich sind zwei weitere Cluster (15326 Lebus, 15306 Seelow) erkennbar sowie weitere zwei kleinere in der Region Eisenhüttenstadt. Die Angaben zu den Clustern sind aus Tab. 11 ersichtlich:

Tab. 11: Ergebnis der Rnnh-Clusterberechnung IgG positiver Patienten (min=10)

ORDER	CLUSTER	MEAN X	MEAN Y	ROTATION	X-AXIS	Y-AXIS	AREA	POINTS
1	1	4672894,28	5805277,03	83,71	0,31	0,20	0,20	29
1	2	4679975,45	5782449,68	14,73	0,24	0,35	0,26	22
1	3	4673531,88	5804697,56	30,33	0,04	0,20	0,02	16
1	4	4673774,83	5805066,33	63,66	0,26	0,19	0,15	12
1	5	4671677,22	5802791,67	7,73	0,13	0,20	0,08	18
1	6	4673400,56	5804265,94	86,18	0,11	0,18	0,06	18
1	7	4672648,29	5803216,76	1,60	0,05	0,19	0,03	17
1	8	4673911,00	5804475,06	82,75	0,17	0,09	0,05	17
1	9	4680311,85	5781350,08	41,57	0,32	0,18	0,18	13
1	10	4673405,13	5803224,80	5,60	0,17	0,11	0,06	15
1	11	4672474,31	5813399,08	77,53	0,65	0,72	1,46	13
1	12	4673720,08	5802231,15	85,87	0,06	0,18	0,04	13
1	13	4665440,18	5793399,09	26,32	0,59	0,72	1,33	11
1	14	4675844,82	5794117,64	8,91	0,39	0,88	1,07	11
1	15	4673059,36	5802192,82	6,57	0,09	0,22	0,06	11
1	16	4673602,18	5802695,18	26,48	0,19	0,11	0,07	11
1	17	4661542,90	5824833,60	25,33	0,47	0,77	1,14	10
1	18	4672194,55	5802815,45	43,66	0,10	0,21	0,07	11
2	1	4673639,17	5800450,90	76,69	12,28	3,87	149,20	17

Die Simulation auf Basis einer Zufallsverteilung identifiziert maximal fünf Cluster (95% Konfidenzintervall = drei Cluster) (Abb. 57)

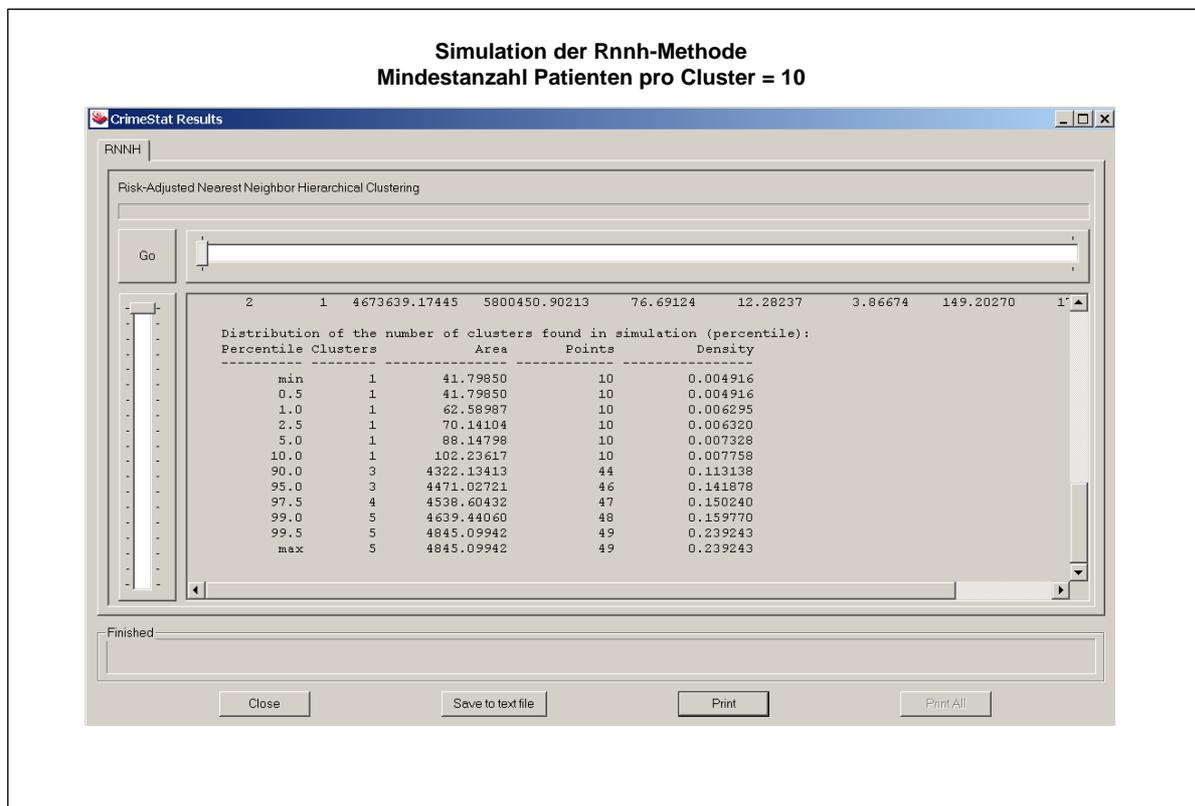


Abb. 57: Simulation der RNNH-Methode (Mindestanzahl Patienten pro Cluster =10)

Man kann also auch bei der RNNH-Methode davon ausgehen, dass es sich nicht um zufällige Cluster handelt, sondern dass die Patientenverteilung Hotspots aufzeigt.

Verändert man nun nur die Mindestanzahl Patienten pro Cluster auf 15 (alle übrigen Parameter werden beibehalten), reduziert sich die Clusteranzahl auf neun Cluster (Abb. 58).

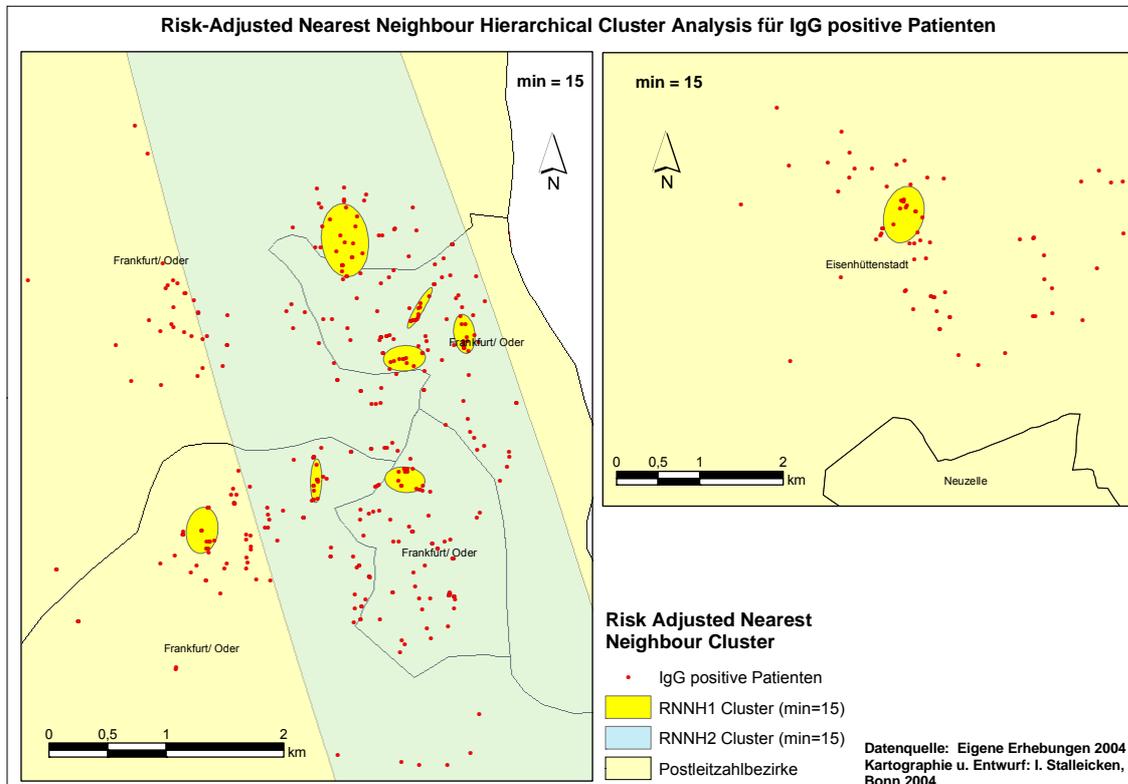


Abb. 58: RNNH-Clusterberechnung unter Berücksichtigung der absoluten Patientenzahl für IgG mit Standardeinstellungen (Min =15)

Betrachtet man diese acht primären Cluster bei mindestens 15 Patienten eingehender, so kann man einzelne Straßenzüge mit einer Serie von IgG-positiven Patienten identifizieren (Abb. 59 rechts). Es können rund 10 Patienten aus nur einer Straße mit IgG positiver Serologie identifiziert werden.

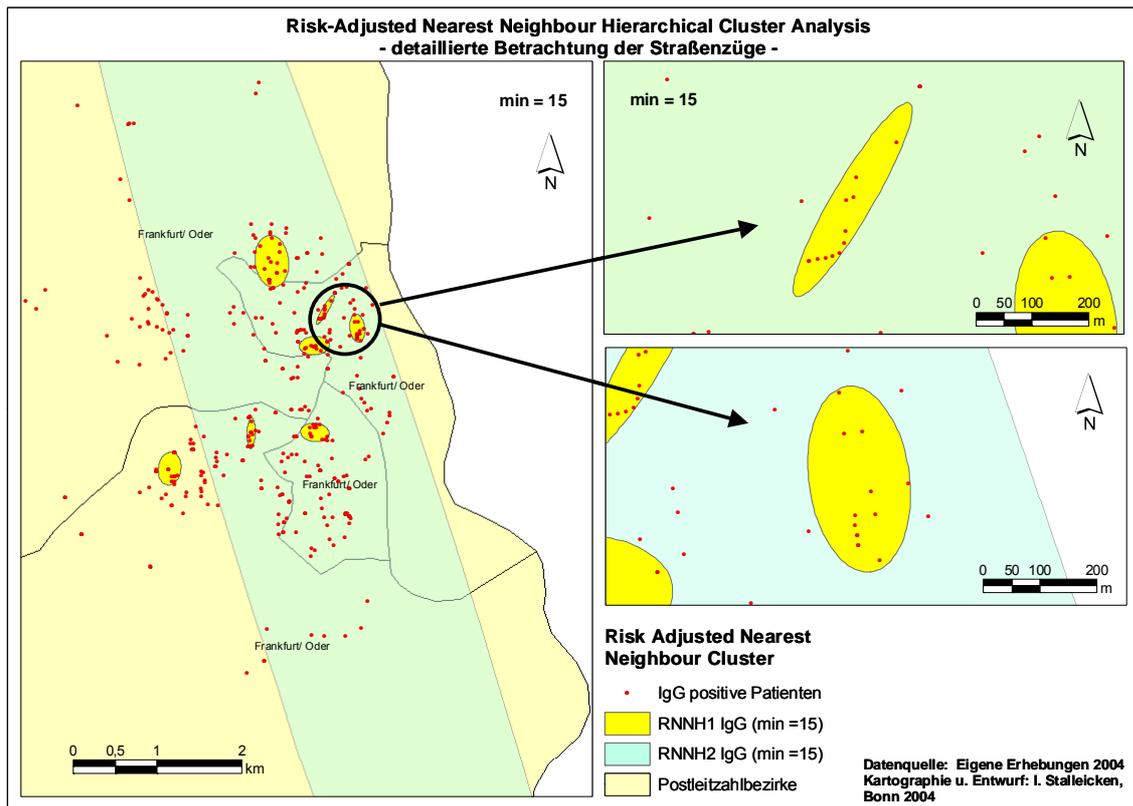


Abb. 59: RNNH-Clusterberechnung – detaillierte Betrachtung der Straßenzüge –

Die Ergebnisse der Clusteranalyse sind für jedes Cluster in Tab. 12 dargestellt:

Tab. 12: Ergebnis der RNNH-Clusterberechnung IgG positiver Patienten (min=15)

ORDER	CLUSTER	MEAN X	MEAN Y	ROTATION	X-AXIS	Y-AXIS	AREA	POINTS
1	1	4672894,28	5805277,03	83,71	0,31	0,20	0,20	29
1	2	4679975,45	5782449,68	14,73	0,24	0,35	0,26	22
1	3	4673531,88	5804697,56	30,33	0,04	0,20	0,02	16
1	4	4671677,22	5802791,67	7,73	0,13	0,20	0,08	18
1	5	4673400,56	5804265,94	86,18	0,11	0,18	0,06	18
1	6	4672648,29	5803216,76	1,60	0,05	0,19	0,03	17
1	7	4673911,00	5804475,06	82,75	0,17	0,09	0,05	17
1	8	4673405,13	5803224,80	5,60	0,17	0,11	0,06	15
2	1	4673930,48	5801299,81	72,53	12,27	1,27	48,77	8

Die Simulation (Anzahl =100) der Clusterberechnung mit einem Minimum von 15 Patienten pro Cluster auf Basis einer Zufallsverteilung ergibt das in Abb. 60 dargestellte Ergebnis. Unter der Annahme einer räumlichen Zufallsverteilung würden gemäß den vorgenommenen Einstellungen maximal zwei Cluster identifiziert werden. Auch in diesem Fall muss man davon ausgehen, dass es sich bei der Verteilung von IgG positiven Patienten um nicht zufällige Hotspots handelt.

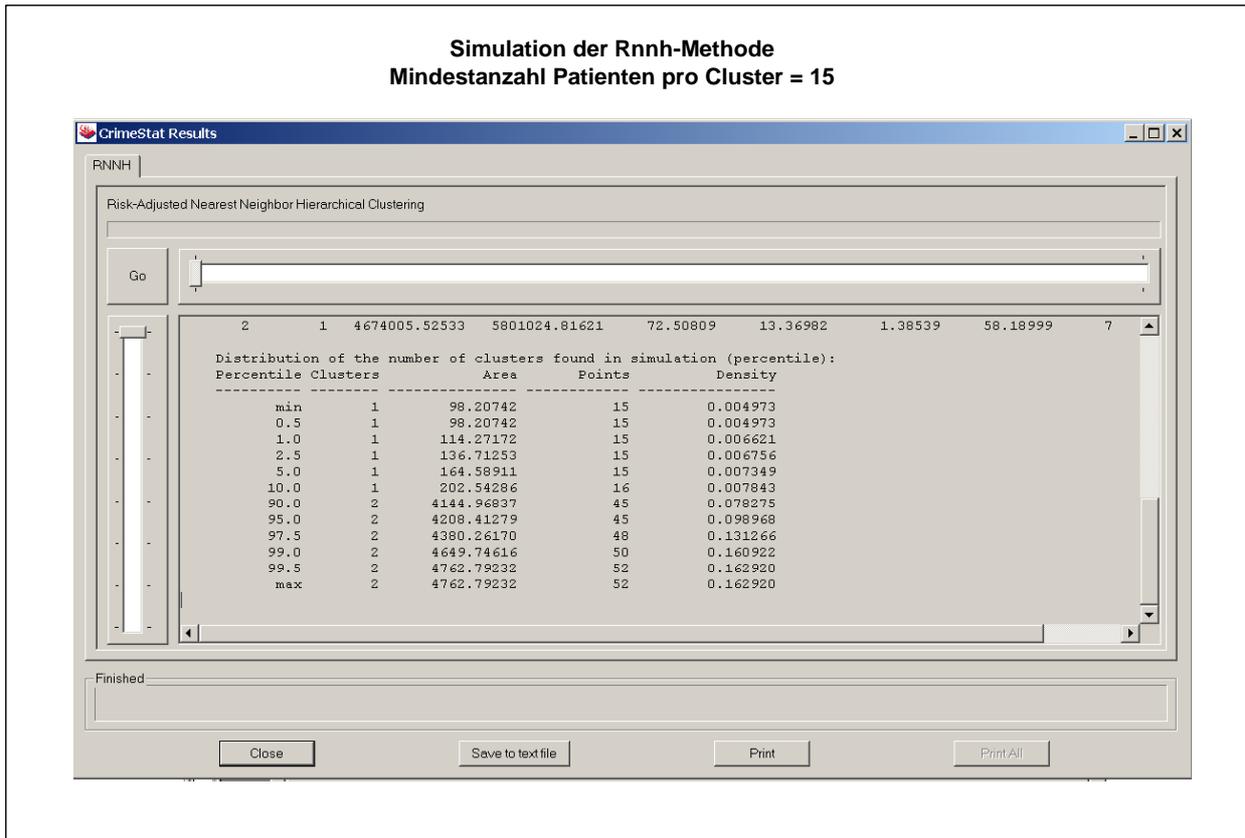


Abb. 60: Simulation der RNNH-Methode (Mindestanzahl Patienten pro Cluster =15)

In Abb. 61 sind die einzelnen Cluster gemeinsam mit der absoluten Anzahl der Patienten dargestellt.

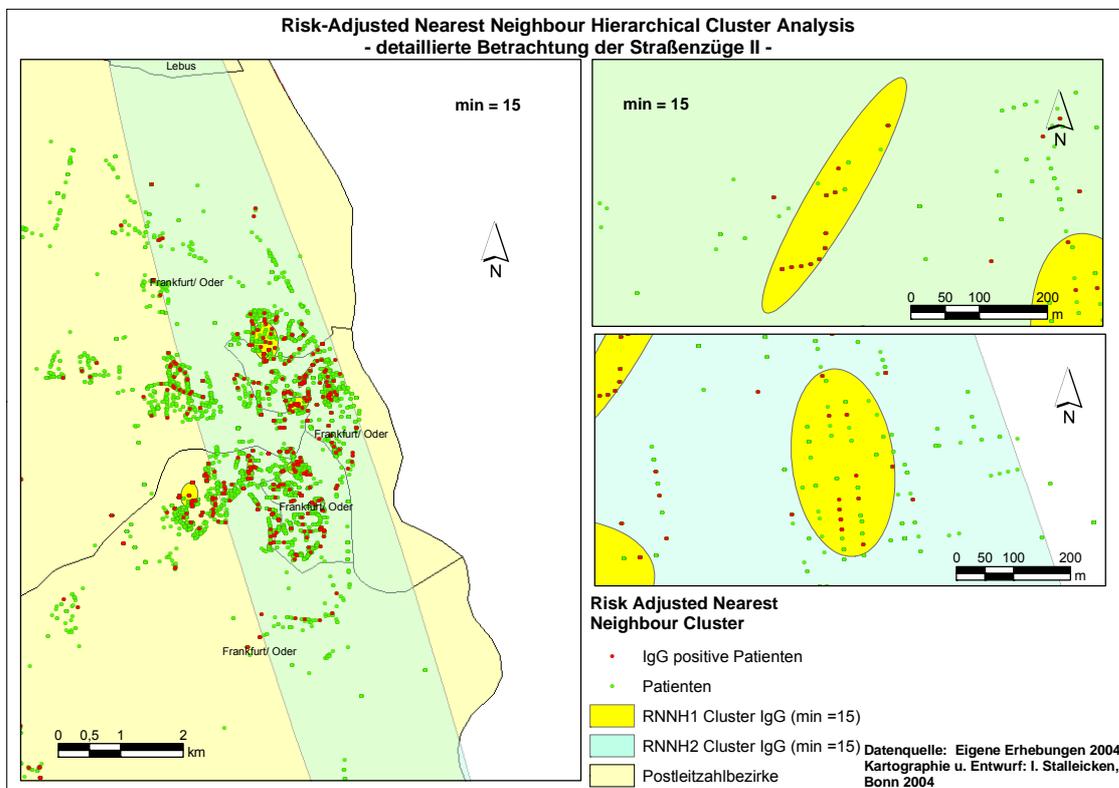


Abb. 61: RNNH-Clusterberechnung – detaillierte Betrachtung der Straßenzüge II –

Auf diese Weise kann für jedes Cluster die Verteilung serologisch positiver Patienten unter Berücksichtigung der absoluten Patientenzahlen nachvollzogen werden. Es ist deutlich erkennbar, dass einzelne Straßenzüge über eine große Anzahl IgG positiver Patienten verfügen (siehe Abb. 61 rechts).

Erhöht man die Mindestanzahl Patienten pro Cluster auf 20, verbleiben lediglich zwei Cluster: Eines in Frankfurt (Oder) und eines in Eisenhüttenstadt (Abb. 62). Auffällig ist insbesondere beim Cluster IgG positiver Patienten in Eisenhüttenstadt die hohe Konzentration der Patienten.

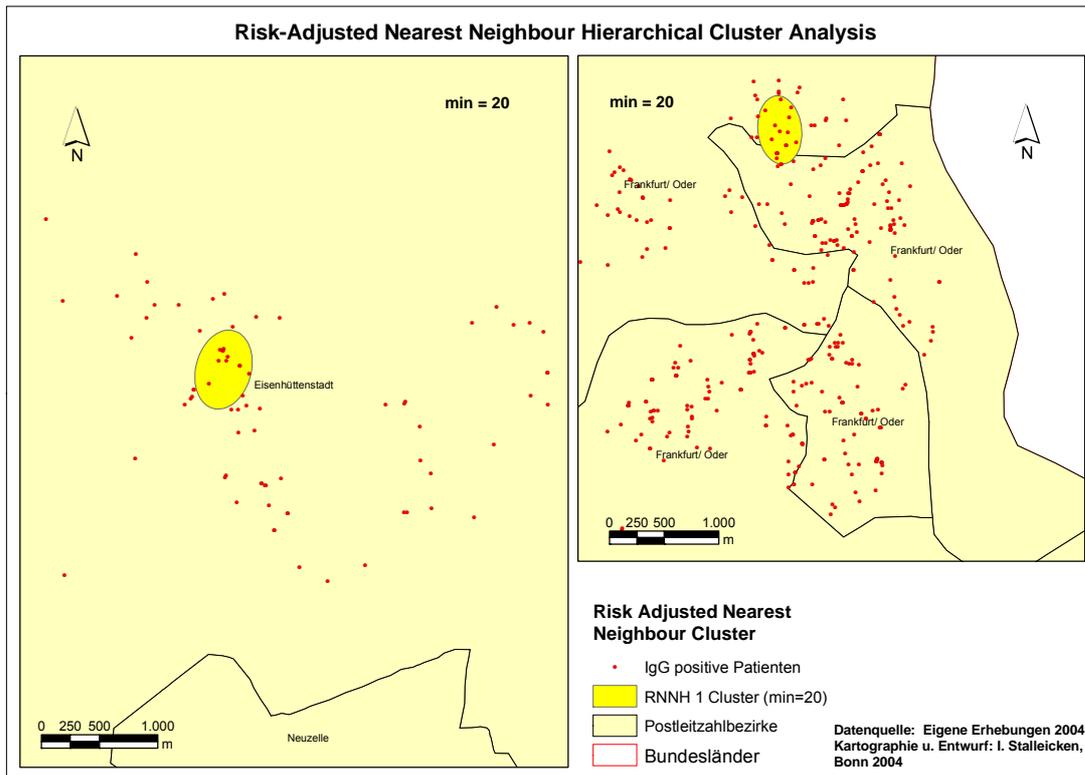


Abb. 62: RNNH-Clusterberechnung für IgG positive Patienten (min=20 Patienten)

Auch in diesem Beispiel ergibt die gemeinsame Darstellung mit der absoluten Anzahl der Patienten ein Bild über die Verteilung serologisch positiver und negativer Patienten innerhalb der einzelnen Cluster (Abb. 63)

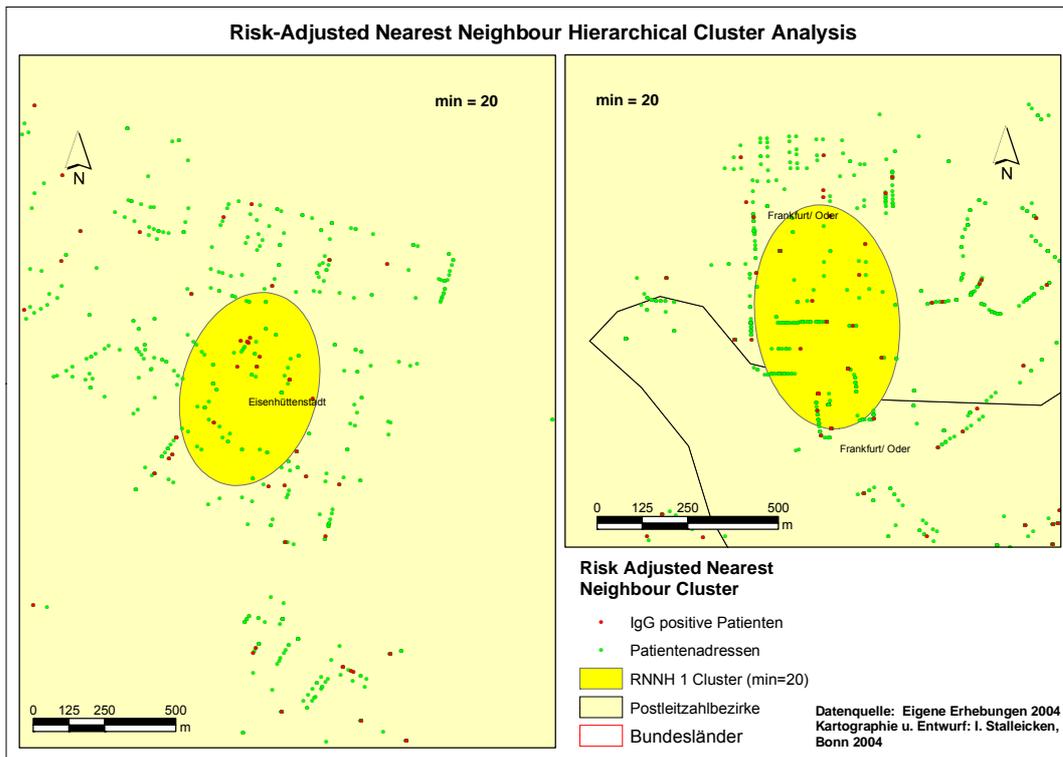


Abb. 63: RNNH-Clusterberechnung für IgG positive Patienten (min=20) II

Legt man eine Mindestanzahl von 25 Patienten pro Cluster zugrunde, tritt nur noch ein einziges Cluster auf (Abb. 64)

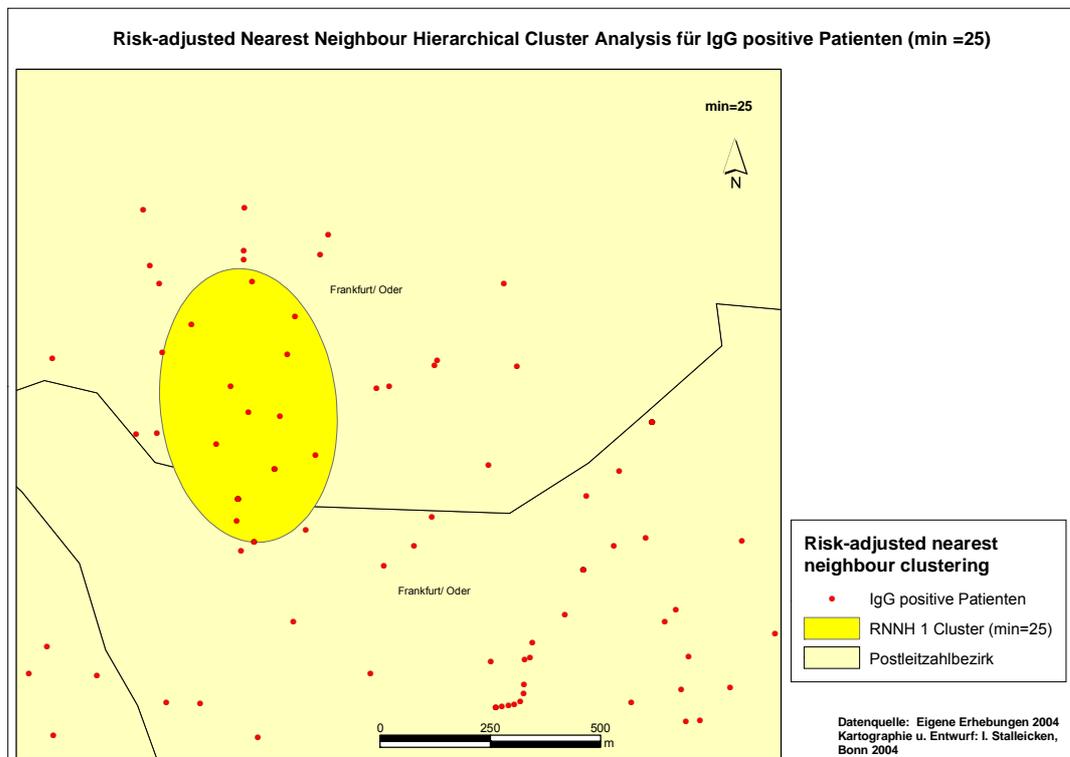


Abb. 64: RNNH-Clusterberechnung für IgG positive Patienten (min=25)

Veränderung der Standardeinstellungen

Neben den genannten Standardeinstellungen kann die Identifikation verändert werden. In Tab. 13 sind die

er-

Tab. 13: Veränderung der Standardabweichung der Ellipse

Nearest Neighbour	
Mindestanzahl Patienten pro Cluster 15	M
Wahrscheinlichkeitsgrad: 50%	S
Anzahl Standardabweichung der Ellipse: 2x	M
Untersuchungsgebiet: 30611 km ²	A K

y-

Es wurde lediglich die Standardabweichung der Ellipse für eine Mindestanzahl von 15 Patienten

ser

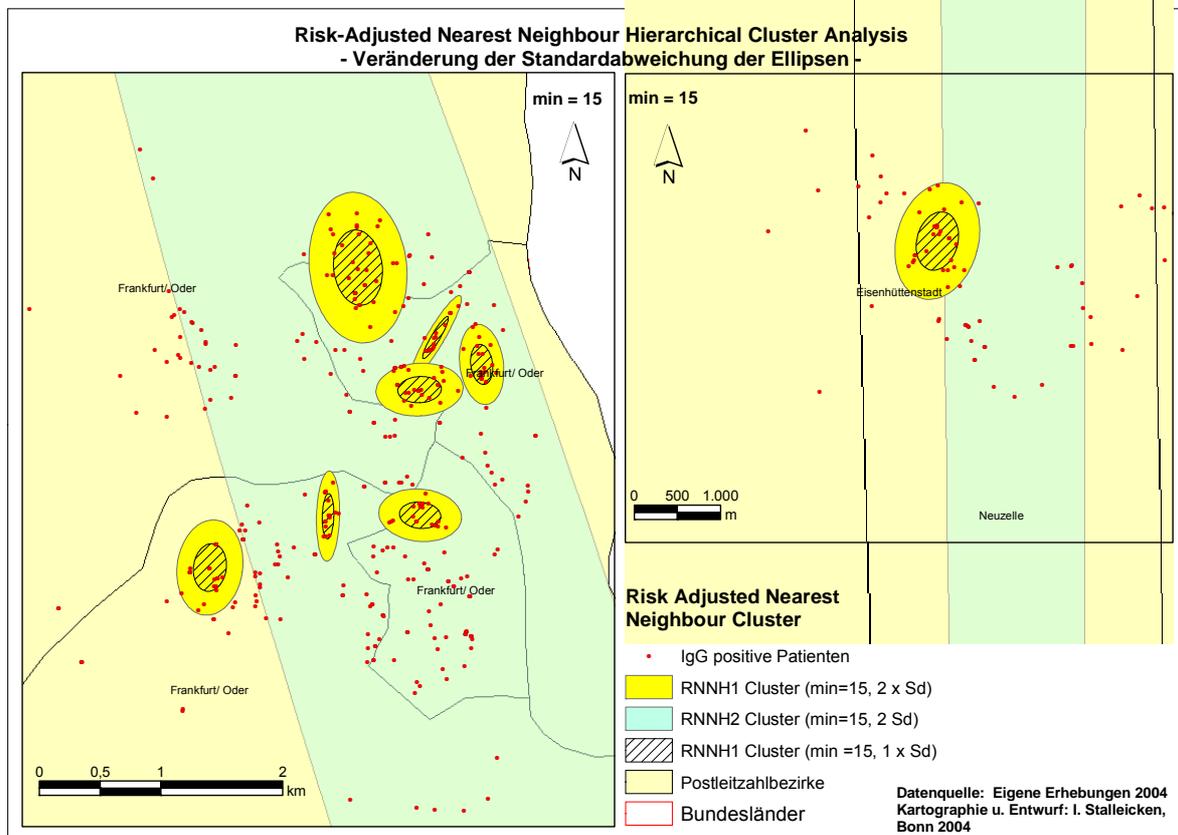


Abb. 65: RNNH-Clusterberechnung mit einer Standardabweichung der Ellipse von 2

In Abb. 65 ist erkennbar, dass die Ellipsen mit zweifacher Standardabweichung deutlich größer sind, als diejenigen mit nur einfacher Standardabweichung. In Tab. 14 sind die Angaben zu den einzelnen Clustern dargestellt.

Tab. 14: Ergebnis der RNNH-Clusterberechnung IgG positiver Patienten (min=15) mit veränderter Standardabweichung

ORDER	CLUSTER	MEAN X	MEAN Y	ROTATION	X-AXIS	Y-AXIS	AREA	POINTS
1	1	4672894,28	5805277,03	83,71	0,63	0,40	0,79	29
1	2	4679975,45	5782449,68	14,73	0,48	0,69	1,03	22
1	3	4673531,88	5804697,56	30,33	0,07	0,40	0,09	16
1	4	4671677,22	5802791,67	7,73	0,27	0,40	0,34	18
1	5	4673400,56	5804265,94	86,18	0,22	0,36	0,25	18
1	6	4672648,29	5803216,76	1,60	0,09	0,37	0,11	17
1	7	4673911,00	5804475,06	82,75	0,33	0,18	0,19	17
1	8	4673405,13	5803224,80	5,60	0,34	0,22	0,23	15
2	1	4673930,48	5801299,81	72,53	24,53	2,53	195,09	8

Die primären Cluster umfassen mindestens 15 bis maximal 29 serologisch positive Patienten. Neben der auffällig größeren Fläche jedes einzelnen primären Clusters ist auch die Fläche des sekundären Clusters deutlich vergrößert.

Eine zweite Möglichkeit der Parametermodifikation erfolgte bei der Kernel-density-Berechnung unter Verwendung von einer Mindestanzahl von 50 Punkten innerhalb des Suchradiuses (Tab. 15)

Tab. 15: Veränderung des Suchradius bei der RNNH-Methode IgG positiver Patienten

Nearest Neighbour	Kernel density interpolation
Mindestanzahl Patienten pro Cluster 15	Methode der Interpolation: normal
Wahrscheinlichkeitsgrad: 50%	Suchradius: angepasst
Anzahl Standardabweichung der Ellipse: 1x	Mindestanzahl Punkte: 50
Untersuchungsgebiet: 30611 m ²	Angabe eines Rasters mit entsprechenden x- und y-Koordinaten

Das Ergebnis der Berechnung ist in Tab. 16 und Abb. 66 dargestellt.

Tab. 16: Ergebnis der RNNH-Methode mit einem variablen Suchradius (n=50)

ORDER	CLUSTER	MEAN_X	MEAN_Y	ROTATION	X-AXIS	Y-AXIS	AREA	POINTS
1	1	4672896,26	5805227,11	83,90	0,29	0,19	0,17	27
1	2	4680015,05	5782339,86	69,65	0,35	0,22	0,24	21
1	3	4673531,88	5804697,56	30,33	0,04	0,20	0,02	16
1	4	4672648,29	5803216,76	1,60	0,05	0,19	0,03	17
1	5	4673279,06	5804357,19	53,96	0,16	0,07	0,04	16
1	6	4673903,31	5804479,19	86,31	0,17	0,08	0,04	16
2	1	4674378,98	5800719,61	73,05	14,92	0,84	39,29	6

Insgesamt wurden sechs primäre und ein sekundäres Cluster identifiziert. Dieses Ergebnis unterscheidet sich nur wenig von der Berechnung mit einem angepassten Suchradius von 100 Punkten (Abb. 66)

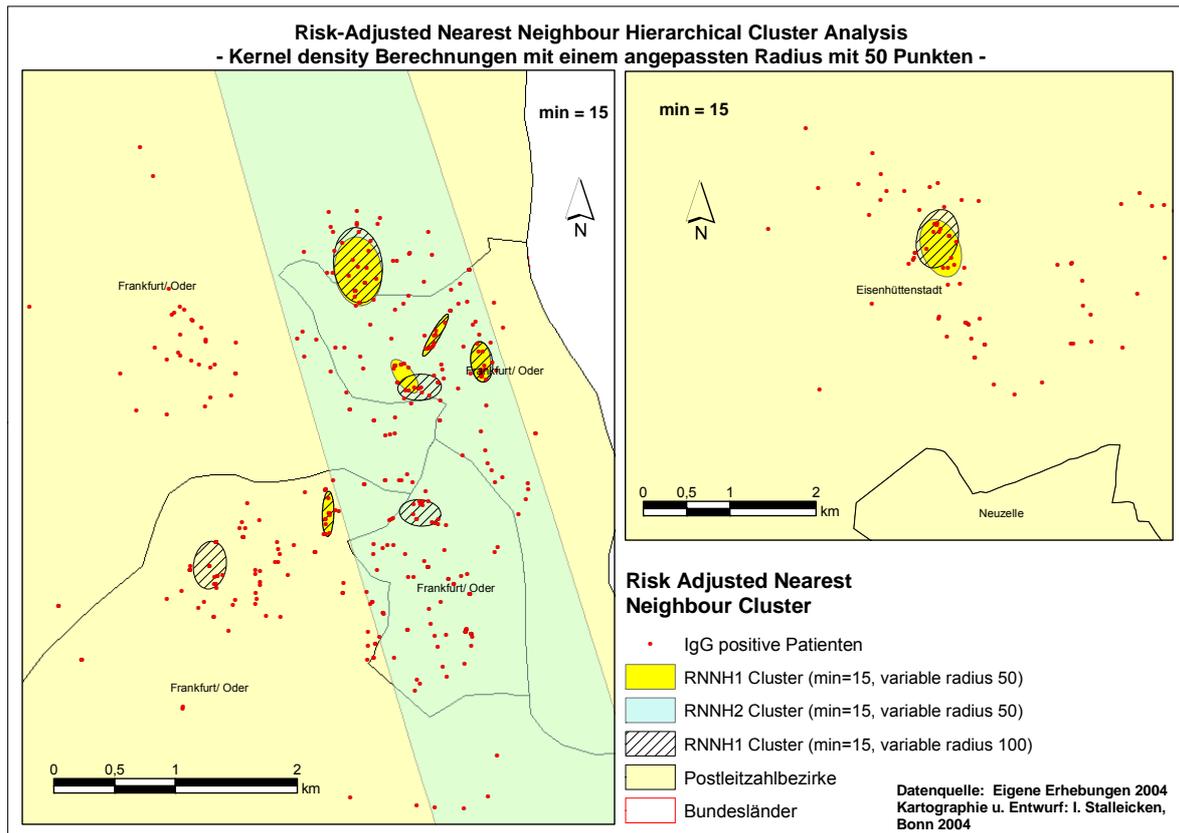


Abb. 66: RNNH-Clusterberechnung für IgG positive Patienten mit einem angepassten Suchradius von 50 Punkten

In Abb. 66 ist ersichtlich, dass sich nicht die Größe, sondern eher die Ausrichtung der Cluster unterscheiden. Darüber hinaus ist ein Cluster im Raum Frankfurt (Oder) weggefallen.

Eine dritte Möglichkeit der Parametermodifikation erfolgte ebenfalls bei der Kernel-density-Berechnung unter Verwendung eines festgesetzten Suchradius (Tab. 17)

Tab. 17: Modifikation der Standardeinstellungen für IgG positive Patienten: Festgesetzter Radius

Nearest Neighbour	Kernel density interpolation
Mindestanzahl Patienten pro Cluster 15	Methode der Interpolation: normal
Wahrscheinlichkeitsgrad: 50%	Suchradius: festgesetzt
Anzahl Standardabweichung der Ellipse: 1x	Intervall: 5000 m
Untersuchungsgebiet: 30611000000 m ²	Angabe eines Rasters mit entsprechenden x- und y-Koordinaten

Die Einstellungen für den festgesetzten Radius wurden den Kernel-density-Berechnungen in ArcGIS entnommen. Hier wurde ein Suchradius von 5000 m für das Untersuchungsgebiet Brandenburg angenommen. Das Ergebnis der Risk-adjusted nearest neighbour Analyse ist in Abb. 67 und in Tab. 18 dargestellt.

Tab. 18: Ergebnisse der RNNH-Berechnungen mit einem festgesetzten Suchradius von 5000m

ORDER	CLUSTER	MEAN X	MEAN Y	ROTATION	X-AXIS	Y-AXIS	AREA	POINTS
1	1	4673723,38	5804568,70	34,31	0,31	0,20	0,20	40
1	2	4673315,39	5804336,17	50,28	0,22	0,11	0,08	23
1	3	4680044,39	5782194,15	75,53	0,65	0,27	0,55	33
1	4	4671720,43	5802695,57	81,66	0,27	0,23	0,19	28
1	5	4672919,37	5805220,26	11,90	0,21	0,27	0,18	27
1	6	4673635,88	5802513,08	63,63	0,34	0,19	0,20	26
1	7	4673274,52	5803221,72	24,11	0,18	0,31	0,18	25
1	8	4672615,61	5803184,30	12,76	0,10	0,27	0,09	23
1	9	4673038,19	5802259,31	87,60	0,29	0,09	0,08	16
1	10	4671546,18	5804681,29	57,61	0,28	0,16	0,14	17
2	1	4673583,33	5801487,46	71,63	10,83	1,13	38,58	10

Insgesamt wurden 10 primäre Cluster und ein sekundäres Cluster identifiziert. Die Anzahl IgG positiver Patienten je Cluster (primär) erstreckt sich von mindestens 16 bis zu maximal 40 Patienten und verfügt damit über mehr Patienten pro Cluster als bei der Berechnung in Tab. 16 mit einem variablen Radius. Darüber hinaus werden zusätzlich vier Cluster identifiziert, die bei der Berechnung mit einem variablen Radius nicht auftraten (Abb. 67)

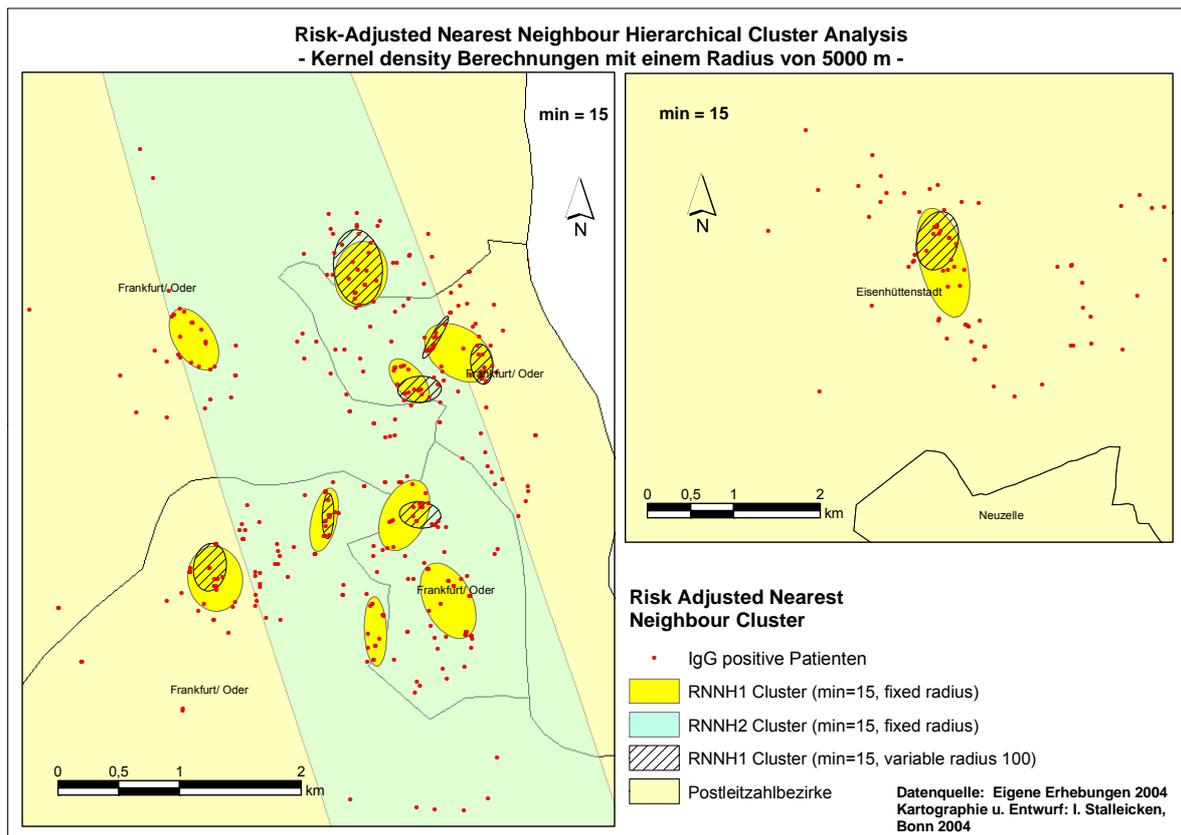


Abb. 67: RNNH-Clusterberechnung mit festgesetztem Radius der Kernel density Berechnungen

Die Berechnung des RNNH-Clustering mit einem festgesetztem Radius der Kernel-density-Interpolation ist jedoch problematisch. Die starke Konzentration der Patienten im Stadtgebiet Frankfurt (Oder) erschwert die Berechnung der Punktdichten, wenn das gesamte Einzugsgebiet des untersuchten Klinikums betrachtet werden soll. Dies wirkt sich auch auf die Clusterberechnung der RNNH-Methode aus.

Die Berechnung des RNNH-Clusterings für IgM positive Patienten erfolgte ebenfalls zunächst unter Verwendung der Standardparameter. Allerdings wurde aufgrund der sehr geringen Anzahl IgM positiver Patienten (n= 51) eine Mindestanzahl von 5 Patienten pro Cluster ausgewählt (Tab. 19)

Tab. 19: Modifikation der Standardeinstellungen für IgM positive Patienten: Variabler Radius

Nearest Neighbour	Kernel density interpolation
Mindestanzahl Patienten pro Cluster: 5	Methode der Interpolation: normal
Wahrscheinlichkeitsgrad: 50%	Suchradius: angepasst
Anzahl Standardabweichung der Ellipse: 1x	Interval: 100
Untersuchungsgebiet: 30611000000 m ²	Angabe eines Rasters mit entsprechenden x- und y-Koordinaten

Bei den gegebenen Einstellungen wird lediglich ein Cluster in Frankfurt (Oder) identifiziert (Abb. 68)

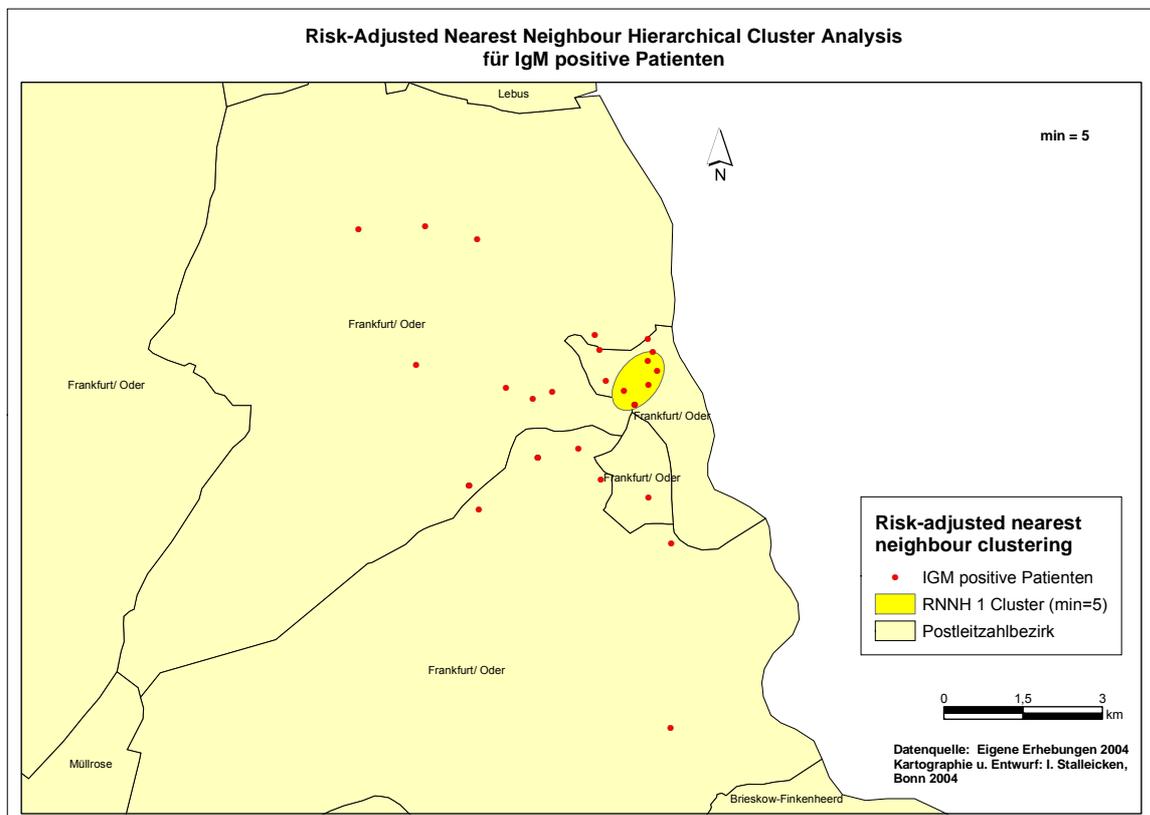


Abb. 68: RNNH-Clusterberechnung für IgM positive Patienten (min=5) mit einem angepassten Suchradius von 100

Erhöht man die Mindestanzahl von IgM positiven Patienten je Cluster auf sieben, erscheint dieses Cluster erneut. Bei einer Mindestanzahl Patienten von zehn Patienten wird kein Cluster mehr identifiziert.

Ebenso wie bei den IgG positiven Patienten können bei den IgM positiven Patienten weitere Parameter modifiziert werden, darunter Parameter der Kernel-density-Berechnung (Tab. 20):

Tab. 20: Modifikation der Standardeinstellungen für IgM positive Patienten: Angepasster Radius (n=50)

Nearest Neighbour	Kernel density interpolation
Mindestanzahl Patienten pro Cluster 5	Methode der Interpolation: normal
Wahrscheinlichkeitsgrad: 50%	Suchradius: angepasst
Anzahl Standardabweichung der Ellipse: 1x	Mindestanzahl Punkte: 50
Untersuchungsgebiet: 30611 km ²	Angabe eines Rasters mit entsprechenden x- und y-Koordinaten

Es werden bei einem angepassten Suchradius von 50 Punkten zwei Cluster identifiziert, die jedoch recht nah beieinander liegen (Abb. 69):

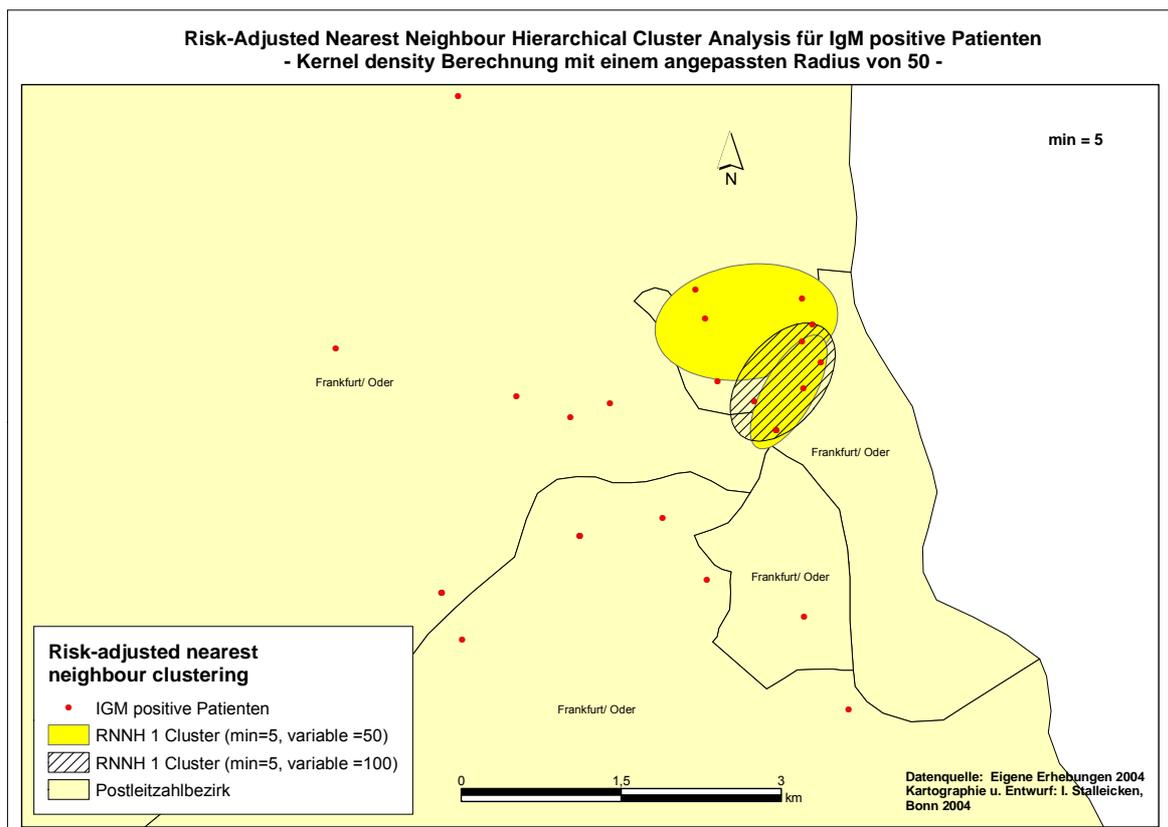


Abb. 69: RNNH-Clusterberechnung für IgM positive Patienten (min=5) und angepasster Suchradius von 50 Punkten

Verglichen mit der Berechnung anhand eines angepassten Suchradius von 100 Punkten werden (schraffiertes Cluster) wird ein zweites Cluster identifiziert, welches sich leicht mit dem ersten Cluster überschneidet. Reduziert man den angepassten Suchradius auf 20 Punkte, verbleiben jedoch weiterhin diese zwei Cluster.

Eine weitere Modifikation wurde anhand der Festsetzung des Suchradius auf 5000 m durchgeführt (Tab. 21).

Tab. 21: Modifikation der Standardeinstellungen für IgM positive Patienten: Festgesetzter Radius

Nearest Neighbour	Kernel density interpolation
Mindestanzahl Patienten pro Cluster 5	Methode der Interpolation: normal
Wahrscheinlichkeitsgrad: 50%	Suchradius: festgesetzt
Anzahl Standardabweichung der Ellipse: 1x	Interval: 5000 m
Untersuchungsgebiet: 30611 km ²	Angabe eines Rasters mit entsprechenden x- und y-Koordinaten

Bei dieser Variante wurden unter Verwendung von mindestens fünf Patienten pro Cluster zwei Cluster identifiziert (Abb. 70)

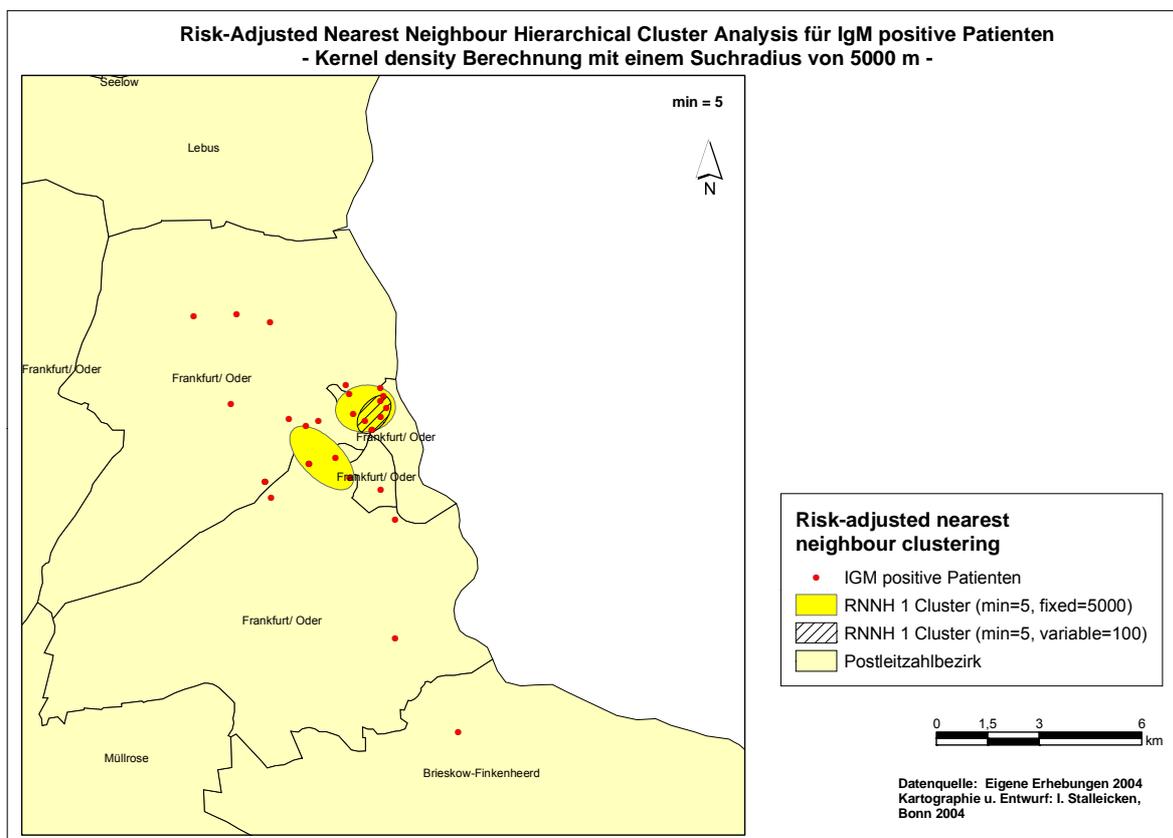


Abb. 70: RNNH-Clusterberechnung für IgM positive Patienten (min=5) unter Anwendung eines festgesetzten Suchradius von 5000 m

Wie auch bei den IgG positiven Patienten erzeugt ein festgesetzter Suchradius von 5000 m, bezogen auf das gesamte Einzugsgebiet des Klinikums ein zusätzliches Cluster. Auch in diesem Fall ist das Ergebnis aufgrund der großen Dichteunterschiede der absoluten Patientenzahlen im Einzugsgebiet und der hohen Patientendichte im Raum Frankfurt (Oder) mit Vorsicht zu genießen.

Über die Clusteridentifikation für IgG und IgM positive Patienten hinaus, wurde für IgG positive Patienten, die angaben, sich regelmäßig in Schwimmbädern oder sonstigen Sportstätten aufzuhalten, eine räumliche Clusteranalyse durchgeführt, da diese Patienten gemäß statistischer Risikoanalyse ein erhöhtes Erkrankungsrisiko besitzen (Kap. 4.3.2). Hierbei handelt es sich um 95 Patienten, die recht konzentriert auf die Stadt Frankfurt (Oder) verteilt sind.

Folgende Parameter wurden festgesetzt (Tab. 22):

Tab. 22: Modifikation der Standardeinstellungen für IgG positive Patienten mit einem regelmäßigen Besuch von Schwimmbädern und sonstigen Sportstätten (variabler Radius)

Nearest Neighbour	Kernel density interpolation
Mindestanzahl Patienten pro Cluster 5	Methode der Interpolation: normal
Wahrscheinlichkeitsgrad: 50%	Suchradius: angepasst
Anzahl Standardabweichung der Ellipse: 1x	Intervall: 100
Untersuchungsgebiet: 30611000000 m ²	Angabe eines Rasters mit entsprechenden x- und y-Koordinaten

Aufgrund der recht geringen Anzahl Patienten (n=95) wurde zunächst eine Mindestanzahl von 5 Patienten pro Cluster angenommen. Insgesamt wurden 5 primäre und ein sekundäres Cluster identifiziert (Abb. 71):

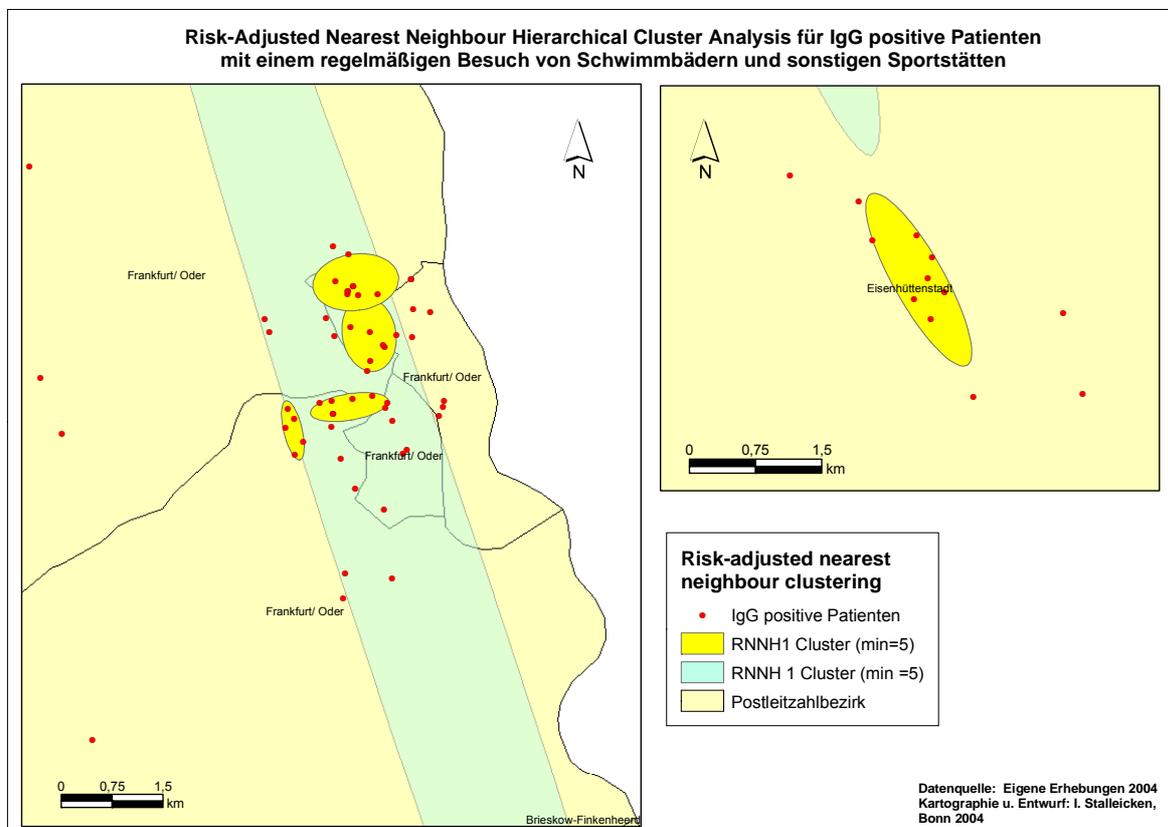


Abb. 71: RNNH-Clusterberechnung für IgG positive Patienten (min=5) mit regelmäßigem Besuch von Schwimmbädern und sonstigen Sportstätten

Vier primäre Cluster befinden sich in der Stadt Frankfurt (Oder), eines in Eisenhüttenstadt. Nimmt man nun mindestens zehn Patienten pro Cluster an, verbleiben lediglich zwei Cluster, die recht eng beieinander liegen (Abb. 72).

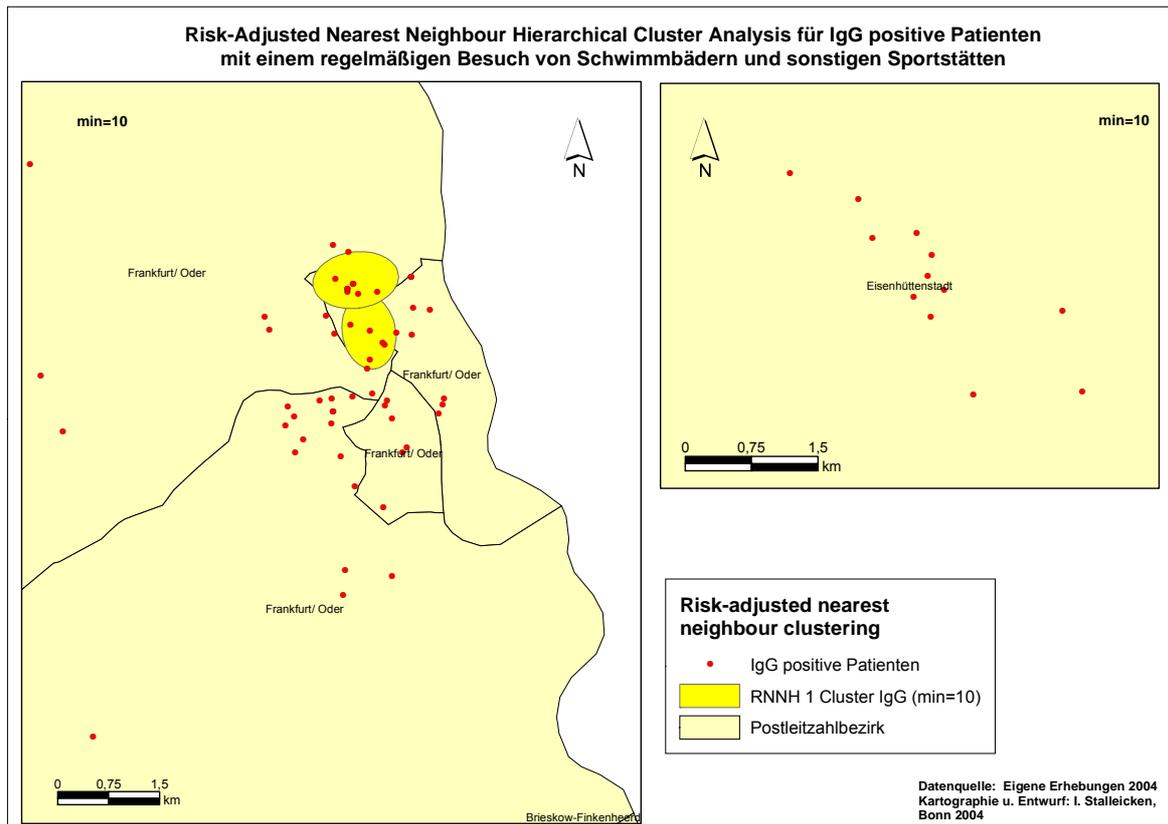


Abb. 72: RNNH-Clusterberechnung für IgG positive Patienten (min=10) mit regelmäßigen Besuch von Schwimmbädern und sonstigen Sportstätten

Die Cluster befinden sich zum überwiegenden Anteil in einem Postleitzahlbezirk in Frankfurt (Oder). Das Cluster in Eisenhüttenstadt ist nicht mehr vorhanden. Dennoch sollten Teile Frankfurt (Oder)s und Eisenhüttenstadt für weitere Beobachtungen (vor Ort) herangezogen werden.

Zusammenfassend kann festgestellt werden, dass die Berücksichtigung der absoluten Patientenzahl bei der Clusteranalyse zu deutlich veränderten Ergebnissen bei der Clusteridentifikation führt. Während die NNH-Methode in der Stadt Frankfurt (Oder) nur ein einziges großes Cluster ermittelt, resultieren aus der RNNH-Methode verschiedene kleinere Cluster innerhalb des Stadtgebietes mit einer hohen Anzahl dort wohnhafter Patienten. Betrachtet man sich diese Cluster eingehender, fallen Straßenzüge auf, die über eine ganze Serie von IgG positiven Patienten verfügen. Die Validierung der Ergebnisse aus der RNNH-Methode erfolgte anhand der Monte-Carlo Simulation. Hierbei lag die Anzahl Cluster, die auf der Basis einer räumlichen Zufallsverteilung identifiziert werden, weit unter der Anzahl Cluster, die mit der RNNH-Methode bei vorliegender Punkteverteilung ermittelt wurden. Die Modifikation der Anzahl Patienten pro Cluster sowie weiteren Parametern, führen zu Veränderungen in der Anzahl, Größe und Ausrichtung der Cluster.

Im folgenden Kapitel werden die Ergebnisse der verschiedenen hier angewendeten Methoden diskutiert sowie Vor- und Nachteile herausgearbeitet.

6. Diskussion

Insgesamt wurden im Untersuchungszeitraum von August 2003 bis Juli 2004 14.389 Patienten in der Studie erfasst, im monatlichen Mittel etwa 1.200. Es handelt sich dabei sicherlich um eine der größten, nach Literatursichtung möglicherweise die größte bisher durchgeführte Kohortenstudie zur Seroprävalenz von Legionella species-Antikörpern weltweit. Bei einer entsprechenden Untersuchung unter isländischen Kindern wurden über den Zeitraum eines Jahres 424 Kinder erfasst (Haraldsson et al. 1990). Nach einem großen Legionellose-Ausbruch im Zusammenhang mit einer Blumenausstellung im Jahr 1999 in den Niederlanden wurden alle Aussteller (n=1616) in eine Kohortenstudie einbezogen (den Boer et al. 2002). Bekannt ist derzeit lediglich eine Studie von Cockings et al. (1997) zu Legionellenerkrankungen in der Stadt Glasgow, die ebenfalls räumlich-statistische Analysemethoden verwendet. Mit Hilfe von GIS und Point Pattern Analysis Methoden wurde untersucht, ob Menschen, die in der Nähe von Kühltürmen leben, einem erhöhten Infektionsrisiko ausgesetzt sind. In Deutschland ist eine derartige auf Initiative des Krankenhauses veranlasste Studie zu Legionelleninfektionen neu, die Untersuchung von mehr als 14.000 Patienten auf Legionelleninfektionen mit anschließender räumlich-statistischer Auswertung hat auch weltweit Neuheitscharakter. Im Folgenden sollen die Schlussfolgerungen aus den Ergebnissen sowohl der deskriptiven Datenauswertung, der Risikoanalyse und insbesondere der räumlich-statistischen Analyse erläutert werden.

6.1 Erste Hinweise der deskriptiven Datenauswertung

Der Anteil des weiblichen Geschlechts an der Studienpopulation lag 7,5% über dem des männlichen Geschlechts. Einen deutlichen Überhang haben Frauen im Zusammenhang mit geburtshilflichen Klinikumsaufnahmen in den Altersklassen von 16 bis 35 Jahren. Jenseits des 75. Lebensjahres steigt der Anteil weiblicher Patienten, bedingt durch die höhere Lebenserwartung der Frauen (im Land Brandenburg mehr als sieben Jahre höher als bei Männern; Gans et al. 2001) erwartungsgemäß auf zwei Drittel an.

Die IgM-Seroprävalenz der gesamten Untersuchungsgruppe betrug 3,4 pro 1000. Bei nur 13 Patienten wurde eine hohe oder sehr hohe IgM-Konzentration im Sinne einer kürzlich erfahrenen Infektion festgestellt. Dieses Resultat korrespondiert gut mit dem lediglich einmaligen Antigen-Nachweis bei einem Patienten im Untersuchungszeitraum. Über 5,0 und damit deutlich über dem Jahresdurchschnitt lag die IgM-Seroprävalenz in den Herbstmonaten (September – November 2003), während die Prävalenz im Frühjahr (März – Juni 2004) besonders niedrig war. Es wurde also eine deutliche Saisonalität der IgM-Seroprävalenz festgestellt, die auf eine höhere Infektionsrate im Herbst hinweist. Die Beobachtungen bestätigen entsprechende Angaben über die Saisonalität der Legionellose in Deutschland (RKI 1999).

In der IgG-Seroprävalenz bildet sich diese Saisonalität naturgemäß verwaschener bzw. mit einer zeitlichen Verzögerung ab, da die IgG-Antikörper wesentlich länger nachweisbar bleiben. Die mittlere IgG-Seroprävalenz betrug 54 pro 1000 Untersuchte. Die höchsten Werte traten im späteren Winter auf (Januar – März 2004).

Die altersdifferenzierte Betrachtung der Seroprävalenzen ergibt unterschiedliche Bilder für IgM und IgG. Die IgM-Prävalenz ist zweigipflig mit einem ersten Maximum im Alter 11-35 und einem zweiten bei den über 85jährigen. Hier bilden sich offensichtlich zwei Gruppen von Risikofaktoren ab, die zahlreich benannt werden (Exner 1991, 1997b, Pleischl et al. 2001; RKI 1999, 2003): einerseits risikobehaftete Aktivitäten, die in jüngerem Alter häufiger ausgeübt werden (Reisen mit Hotelaufenthalten, Besuch von Badeeinrichtungen einschließlich Whirlpools etc. sowie anderer öffentlicher Einrichtungen, in denen geduscht wird); andererseits fortgeschrittenes Lebensalter, chronische Grundkrankheiten und reduzierte Immunkompetenz.

Der Altersverlauf der IgG-Prävalenz zeigt das typische Bild eines kumulativen Prozesses: Bis zur Altersklasse 26-35 Jahre steigt die Prävalenz steil an, danach ist sie recht gleichmäßig und fällt erst für die älteste Gruppe (älter als 85 Jahre) wieder ab. Die langfristig erhaltenen immunologischen IgG-Reaktionen auf Erregerkontakte in Kindheit, Jugend und jüngerem Erwachsenenalter summieren sich und erreichen erst bei Erwachsenen ein stabileres Plateau, wenn Wiederkontakte immer häufiger werden und selten neue Immunreaktionen in der Kohorte auslösen. Entsprechend diesem Bild liefert die analytische Studie keine signifikanten Resultate für den Risikofaktor „höheres Lebensalter“ (65+). Vor dem Hintergrund dieses Resultates einer großen Kohortenstudie ist die Frage zu stellen, ob höheres Lebensalter zu Recht als Risikofaktor für *Legionella-Infektionen* angesehen wird. In Deutschland werden – wie in anderen Ländern, z.B. in Frankreich (Campese und Decludt 2002) – Erkrankungsfälle hauptsächlich bei Erwachsenen gemeldet, während Kinder und Jugendliche in der Regel kaum betroffen sind. So war etwa im Jahr 2002 das durchschnittliche Erkrankungsalter 56,3 Jahre (RKI 2003; Frankreich 2001: 59 Jahre; Campese und Decludt 2002). Möglicherweise handelt es sich hierbei jedoch um ein meldesystembedingtes Artefakt. Nach Ansicht des RKI repräsentieren die übermittelten Legionellose-Fälle nach allen Erkenntnissen nur einen Bruchteil der tatsächlich aufgetretenen Erkrankungen.

Die Untersuchung des Geschlechts als eventueller Risikofaktor ergab überraschenderweise für die Studienpopulation, dass weibliches Geschlecht einen signifikanten Risikofaktor der Seroprävalenz (IgG sowie IgM+IgG) darstellte (RR=1,26/1,27, CI 1,1-1,5, p=0.001). In der Literatur wird regelmäßig männliches Geschlecht als Risikofaktor angegeben (Exner 1997, Pleischl et al. 2001). Auch die nationalen Meldestatistiken weisen Entsprechendes aus. In Deutschland war das Geschlechtsverhältnis m:w im Jahr 2002 1,75:1 (RKI 2003), in Frankreich im Jahr 2001 sogar 3:1 (Campese und Decludt 2002). Bei der Interpretation des Studienergebnisses ist allerdings zu berücksichtigen, dass die dem Bericht zugrunde liegenden Daten nicht altersstandardisiert wurden. Gerade in den beiden Altersbereichen, für die die IgM-Seroprävalenz überdurchschnittlich hoch war (11-35; 85+) ist der Frauenanteil besonders hoch. Es ist deshalb zu vermuten, dass die signifikanten Ergebnisse für den Risikofaktor „Geschlecht“ wesentlich durch die geschlechtsspezifischen Altersstruktur der Studienpopulation beeinflusst wurden. Möglich wäre ebenfalls, dass die Krankheit häufiger bei Männern ausbricht, während sie bei Frauen symptomlos bleibt.

Die deskriptive Auswertung der Untersuchungsergebnisse zur Häufigkeit von Klinikumsaufenthalten zeigte nicht, dass mit zunehmender Zahl von Aufnahmen in das Klinikum Frankfurt (Oder) die Legionella-Seroprävalenz ansteigt. Dies gilt sowohl für IgM als auch für IgG. Dies

spricht nicht für ein erhöhtes Risiko einer Legionella-Infektion bei häufigem Klinikumaufenthalt.

6.2 Risikoanalyse auf Grundlage des Fragebogens

Die Bedeutung von vier Faktoren der allgemeinen Umgebung, die als potentielle Risikofaktoren beschrieben sind (Exner 1991, 1997, Yu 1995, Pleischl et al 2001), wurde für die Legionella-Seroprävalenz überprüft. Dabei stellte sich heraus, dass nur der Besuch von Schwimmbädern und anderen Sportstätten und nur unter Einbeziehung der IgG-Serologie signifikant war (Tab. 4 und Tab. 5). Der Besuch derartiger Einrichtungen stellt das Risiko einer Exposition gegenüber Legionella-haltigen Aerosolen dar, v.a. beim Duschen, in Whirlpools und in der Nähe von Warmwasserfontänen u.ä. (Groothuis et al. 1985). Diese Assoziation wurde durch die Studie signifikant belegt. Es handelt sich um ein scharfes, gut abfragbares Kriterium. Das Ergebnis spiegelt sich sehr wahrscheinlich auch in der altersdifferenzierten IgM-Seroprävalenz wider (s.o.). Eine ähnlich atypische Altersverteilung konnte auch im Zusammenhang mit dem bereits erwähnten Legionelloseis-Ausbruch in den Niederlanden beobachtet werden (van Steenbergen et al., 1999).

Andere in der Literatur als typisch erachtete Risikofaktoren der allgemeinen Umwelt konnten hingegen nicht bestätigt werden. Die Verwendung von Luftbefeuchtern sowie die Wohnnähe zu Rückkühlwerken wurden allerdings von den Patienten insgesamt nur selten als Risikofaktor angegeben: Nur 6,5% benutzen zuhause einen Luftbefeuchter, und nur 0,6% gaben an, in der Nähe eines Rückkühlwerkes zu leben. Letztere Zahl ist unrealistisch niedrig; es muss davon ausgegangen werden, dass viele Patienten mit der richtigen Beantwortung dieser Frage überfordert waren. Es liegt die Vermutung nahe, dass viele Patienten keine genaue Vorstellung von solchen Rückkühlwerken haben, schon gar nicht, ob es sich um Rückkühlwerke mit Aerosolbildung handelt. Die niedrige Frequenz der Angabe von Rückkühlwerken verursacht ein breites Konfidenzintervall für diesen Faktor in der Kohortenstudie (Tab. 3 bis Tab. 5, jeweils Zeile 4). Der alleinige Vergleich der Prävalenzen führt deshalb in die Irre.

Die Frage nach Reisen in den letzten zwölf Tagen vor Aufnahme wäre insbesondere im Zusammenhang mit akuten Legionella-Infektionen bei der Aufnahme (d.h. bei Antigen-Nachweis) von großem Interesse gewesen, da Hotelaufenthalte ein Risikofaktor für derartige Infektionen sind (Exner et al., 2002). Es trat im Untersuchungszeitraum jedoch nur ein einziger Antigen-positiver Patient auf. Dass eine Assoziation zu IgM- und insbesondere zur IgG-Serologie nicht besteht, ist nicht verwunderlich, da die Immunreaktion sich erst einige Zeit nach stattgehabter Infektion entwickelt.

6.3 Interpretation der räumlich-statistischen Analyse

Die Interpretation der räumlich-statistischen Analyse erfolgt gemäß der Ergebnisdarstellung zunächst für die flächenhafte räumliche Analyse auf Basis der Postleitzahlbezirke. Daraufhin wird die punktbasierte räumliche Analyse erläutert, deren Voraussetzung die Geocodierung der Patientenadressen ist. Es sollen Vor- und Nachteile der angewandten Verfahren im

Rahmen dieser Studie diskutiert werden und deren Einsatzmöglichkeiten in der Praxis des Öffentlichen Gesundheitswesens aufgezeigt werden.

6.3.1 Flächenhafte räumliche Analyse

Die flächenhafte räumliche Analyse der absoluten Anzahl IgG positiver Patienten ließ erkennen, dass die Stadt Frankfurt (Oder) (422 IgG positive Patienten) sowie Eisenhüttenstadt (90 Patienten) die höchste Anzahl IgG positiver Patienten aufwiesen, d.h. 67% der IgG positiven Patienten entfallen auf diese beiden Städte. Diese Darstellungsform ist nun wenig aussagekräftig, da diese zwei Städte gleichzeitig über die höchsten absoluten Patientenzahlen verfügen. Daher wurde die Berechnung der IgG-Prävalenz hinzugezogen. Die IgG-Prävalenz ergab Unterschiede für die verschiedenen räumlichen Einheiten. Entsprechend ließen sich differenzierte Choroplethenkarten erstellen.¹² Allerdings ist bei dieser Analyse- und Darstellungsform das Problem der Verzerrung durch unterschiedlich große und teilweise sehr niedrige absolute Patientenzahlen zu beachten. So resultiert beispielsweise die hohe Prävalenz im Bezirk Strausberg daraus, dass von insgesamt nur 53 untersuchten Patienten 5 IgG-positiv waren. Je geringer die absolute Patientenzahl in einer Flächeneinheit ist, desto unzuverlässiger wird die Schätzung der Prävalenz. D.h., in einem Kartenbild werden Prävalenzschätzungen unterschiedlicher Güte zusammengefasst. Darüber hinaus besteht das Problem, dass der optische Eindruck von den größeren Flächen bestimmt wird, die aber häufig die weniger zuverlässigen Prävalenzschätzungen aufweisen, da die Einwohner- und damit auch die Patientendichte gerade in den flächengroßen ländlichen Kreisen oft niedrig ist (Kistemann et al. 2002, Dreesmann 2004).

Anhand des Moran's und Geary's Autokorrelationskoeffizienten wurde ermittelt, ob räumliche Beziehungen zwischen den Prävalenzen einzelner Postleitzahlbezirke bestehen. Beide Ergebnisse weisen auf eine schwach positive räumliche Autokorrelation hin. Das Ergebnis der Moran's Statistik ist statistisch signifikant, das der Geary's Statistik nicht. Da das Ergebnis der Moran's Statistik nur schwach signifikant ist ($2,44 > 2,33$), kann man schließen, dass es sich in diesem Fall um einen Grenzbereich hinsichtlich der Signifikanz der positiven räumlichen Autokorrelation handelt. Betrachtet man die Darstellung der Prävalenzen, so fallen hohe Prävalenzen im nördlichen Teil des Untersuchungsgebietes auf, vereinzelte hohe Prävalenzen im engeren Raum von Frankfurt (Oder) sowie geringe Prävalenzen in den ländlichen Regionen Brandenburgs. Auf Basis der Berechnungen der Moran's und Geary's Statistik wurde der Lokationsquotient berechnet. Wie auch schon die Prävalenz einzelner Postleitzahlbezirke zeigte, besitzen die Postleitzahlbezirke Strausberg, Buckow, Bad Freienwalde, Eisenhüttenstadt und einzelne Bezirke von Frankfurt (Oder) einen Lokationsquotienten > 1 , was auf eine überdurchschnittlich hohe Prävalenz hinweist. Auch in diesem Fall besteht das Problem der Überbetonung von Gebieten mit einer geringen Patientenzahl, wie dies insbesondere für Strausberg, Buckow und Bad Freienwalde der Fall ist, sowie eine visuelle Überbetonung von Regionen mit größeren Flächeneinheiten. Aus diesen Gründen besitzt die räumliche Analyse von Punktdaten die höhere Aussagekraft.

¹² Die Zahl IgM-positiver Patienten war für eine sinnvolle räumlich-statistische Analyse zu niedrig.

Verschiedene Studien haben die räumliche Variation der Legionellose belegt. Allerdings ist die Vergleichbarkeit von Prävalenzraten unterschiedlicher Studien und Statistiken oftmals durch Differenzen der zugrunde gelegten Diagnosekriterien, der Erfassungsmethode u.a.m. eingeschränkt, so dass mit Artefakten der räumlichen Variation zu rechnen ist (Bhopal 1993). Dieses Problem entfällt für die vorliegende, einheitlich durchgeführte, homogene Kohortenstudie.

6.3.2 Geocodierung der Patientenadressen

Voraussetzung der räumlichen Analyse war die Geocodierung der Patientendaten (n=14.452). Die Bewertung der Geocodierungsqualität ermöglichte eine manuelle Korrektur der Adressen noch während des Geocodierungsprozesses. Damit konnten fehlerhafte Adressen weitestgehend vermieden werden. Allerdings ist der Arbeitsaufwand schon bei einer Anzahl von mehr als 14.000 Adressen erheblich, zumal die Gebietsreform im Land Brandenburg zusätzliche Probleme aufwarf. So hat die Zahl der Gemeinden im Land Brandenburg seit 1992 stetig abgenommen. Laut Angaben der Bundesanstalt für Bauwesen und Raumordnung (BBR) wurde die Anzahl der Gemeinden in Brandenburg von 1.813 im Jahr 1992 auf 1.516 im Jahr 1997 reduziert (BBR 1999). Dies hatte zur Folge, dass die genannten Ortschaften häufig nicht mehr mit der Postleitzahl übereinstimmten oder die genannten Ortschaften nicht mehr in offiziellen Verzeichnissen genannt wurden. Der Kommentar in Abb. 73 illustriert sehr deutlich, inwieweit die Gebietsreform in der Praxis zu Problemen führt.

04.11.03, Gebietsreform in Brandenburg verwirrt Briefträger

Die Gebietsreform in Brandenburg hat die Deutsche Post anscheinend überraschend getroffen. Aufgrund der Zusammenlegung vieler ehemals eigenständiger Orte gibt es nun dieselben Straßennamen doppelt und dreifach, alle mit der gleichen Postleitzahl. Dies berichten die Potsdamer Neueste Nachrichten. Von den 1.479 Orten sind nach der Gebietsreform nur noch 422 übrig geblieben. Laut Gesetz soll nur noch der neue Ortsname verwendet werden. Eine Sprecherin der Deutschen Post empfiehlt aber bis auf weiteres die alten Ortsnamen zu verwenden, um die Zustellung nicht unnötig zu verlängern.

Gleiche Straßennamen sind vielerorts keine Besonderheit. Jedoch können diese durch unterschiedliche Postleitzahlen unterschieden werden. Dies gilt jedoch nicht für die Brandenburger Großgemeinden. Sie haben alle dieselbe Postleitzahl. (te)

<http://www.posttip.de/news/13998/archiv/Gebietsreform-in-Brandenburg-verwirrt-Brieftraeger.htm>

▶ weitere News

Abb. 73: Probleme der Gebietsreform im Land Brandenburg

Ähnliche Erfahrungen wurden im Rahmen der Geocodierung der Patientenadressen gemacht. Das Hauptproblem bestand in der Zuordnung von Straßen, deren Namen, z.B. Dorfstraße, nicht eindeutig für eine Ortschaft zuzuweisen war. Dies erforderte erheblichen zusätzlichen Arbeitsaufwand, da insbesondere ältere Menschen bei der Angabe ihrer Adresse die alte Ortsbezeichnung verwendeten und eine Neuzuweisung problematisch war. Als ein wesentlicher Vorteil des Geocodierungsprozesses ist deshalb umso mehr die Qualitätsprüfung zu nennen, die auf falsche oder widersprüchliche Angaben hinweist und eine manuelle Korrektur ermöglicht. Nachteilig ist jedoch, dass der „Geocoder“ ebenso automatisch Adressen korrigiert, die ihm nicht bekannt sind, bei denen er aber einen Bezug zu einer anderen

Adresse herstellen kann. Daraus folgt, dass Straßen oder Gemeinden beispielsweise aufgrund von Namensverwandtschaften falsch zugewiesen werden. Es ist somit notwendig, die korrigierten Adressen nochmals zu überprüfen. Bei einer Anzahl von mehr als 14.000 Adressen ist dies mit einem erheblichen Arbeitsaufwand verbunden. In seltenen Fällen wurden Adressen zwar nach der Bezeichnung richtig geocodiert, die geographische Lage war jedoch fehlerhaft. Daher empfiehlt es sich in jedem Falle Stichproben zur Kontrolle der Geocodierung durchzuführen. Insgesamt sollte der Arbeitsaufwand in der Praxis nicht unterschätzt werden.

6.3.3 Punktbasierte räumliche Analyse

Auf der Basis der Geocodierung der Patientenadressen konnte eine punktbasierte räumliche Analyse durchgeführt werden. Die reine Visualisierung der gesamten Patientenadressen ($n=14.452$) zeigt, dass das Einzugsgebiet des Krankenhauses im Wesentlichen durch die Bundesländer Brandenburg und Berlin charakterisiert werden konnte (Abb. 34). Dieses Gebiet von etwa 30611 km² wurde als weitere Berechnungsgrundlage herangezogen. Die reine Visualisierung der Punktdaten erzeugt zudem einen ersten Eindruck über die räumliche Verteilung von serologisch positiven Patienten im Einzugsgebiet. Nur eine sehr geringe Anzahl dieser Patienten liegt außerhalb der Bundesländer Brandenburg und Berlin. Neben der räumlichen Darstellung der serologisch positiven Patienten, konnten diejenigen selektiert und visualisiert werden, die über ein bestimmtes Risikopotenzial, z.B. der Besuch von Schwimmbädern oder sonstigen Sportstätten, verfügen (Abb. 37). Auch in diesem Fall sind erste räumliche Verteilungsmuster erkennbar.

Erste Erkenntnisse der räumlichen Punkteverteilung serologisch positiver Patienten im Untersuchungsgebiet erbrachte die Nächste Nachbar Statistik und die Ripley Statistik. Anhand der Nächste Nachbar Statistik konnte ermittelt werden, dass es sich bei IgG positiven Patienten um eine eher konzentrierte als zufällige räumliche Verteilung handelt. Dies verändert sich leicht in Anbetracht des zugrunde gelegten Untersuchungsgebietes, welches gleichzeitig als ein wichtiger Nachteil dieser Methode gesehen werden muss. Bei gleichbleibender Punktmenge und größer werdendem Untersuchungsgebiet (maximale Punktausdehnung) sinkt der Wert der Nächste Nachbar Statistik R , d.h. es ergibt sich ein Trend zur Konzentration. Bei kleiner werdendem Untersuchungsgebiet (Brandenburg und Berlin) und gleicher Anzahl von Punkten steigt der Wert von R (Trend zur zufälligen bzw. gleichmäßigen Verteilung).

Die Ergebnisse der k -Funktion bestätigen die Aussagen der Nächste Nachbar Statistik. Auch in diesem Fall handelt es sich um eine eher konzentrierte Punkteverteilung. Bei den IgM positiven Patienten liegen die $L(t)$ Werte der K -Funktion bis zu einer Entfernung von 65 km oberhalb des Konfidenzbandes. Über diese Entfernung hinaus verlaufen die Werte ähnlich denen des maximalen Konfidenzintervalls, welches auch in diesem Fall durch die zugrunde gelegte Größe des Untersuchungsgebietes verursacht wird. Die Ergebnisse beider Statistiken ergeben ein erstes Verständnis für die vorliegende Punktverteilung. Sie machen jedoch keine Aussage dazu, ob und wo sich im Untersuchungsgebiet räumliche Cluster befinden.

Neben der räumlichen Verteilung serologisch positiver Patienten im Untersuchungsgebiet wurde die räumliche Dichte der Patienten im Einzugsgebiet des Krankenhauses mittels Ker-

nel-density-Interpolation untersucht. Aufgrund der starken Konzentration der serologisch positiven Patienten im Raum Frankfurt (Oder) und der sonst sehr geringen Dichte im Raum Brandenburg war eine differenzierte Betrachtung des Raumes Frankfurt (Oder) schwierig. So wurde bei der Betrachtung des Raumes Frankfurt (Oder) ein wesentlich geringerer Suchradius angenommen, der ein differenziertes Bild der Punktdichte erlaubte (Abb. 42). So konnten im Raum Frankfurt (Oder) maximal 104 Patienten pro km² ermittelt werden. Legt man jedoch den Raum Brandenburg bei der Analyse zugrunde mit einem entsprechend erhöhten Suchradius von 5.000 Metern, werden geringere Dichten errechnet. Die Kernel-density-Interpolation eignet sich in beiden Beispielen dazu, einen ersten Überblick über die Patientendichte zu erlangen, jedoch ist eine differenzierte Betrachtung, insbesondere bei einem großen Untersuchungsgebiet, mit erheblichen Unterschieden der Punktdichten problematisch.

Die Identifikation räumlicher Cluster serologisch positiver Patienten erfolgte daher nicht ausschließlich mit der Kernel-density-Methode, sondern anhand der Nearest Neighbour Methode und der Kombination von Kernel density und Nearest neighbour hierarchical clustering Ansatz (Risk-adjusted nearest neighbour clustering).

Die Ergebnisse der Nearest neighbour hierarchical clustering Methode zeigen ein großes Cluster im Stadtgebiet Frankfurt (Oder) sowie eines in Eisenhüttenstadt und weitere in den ländlichen Bereichen des Bundeslandes Brandenburg. Es sind große Übereinstimmungen mit den Ergebnissen der Kernel-density-Interpolation erkennbar. Die Anzahl der Cluster variiert entsprechend der vorher festgelegten Mindestanzahl Patienten pro Cluster. Bei einer Mindestanzahl M von fünf Patienten pro Cluster sind es noch 16 identifizierte Cluster, M=10 noch zehn Cluster, bei M=15 noch sechs Cluster, bei M=20 Patienten nur noch zwei Cluster (Abb. 74).

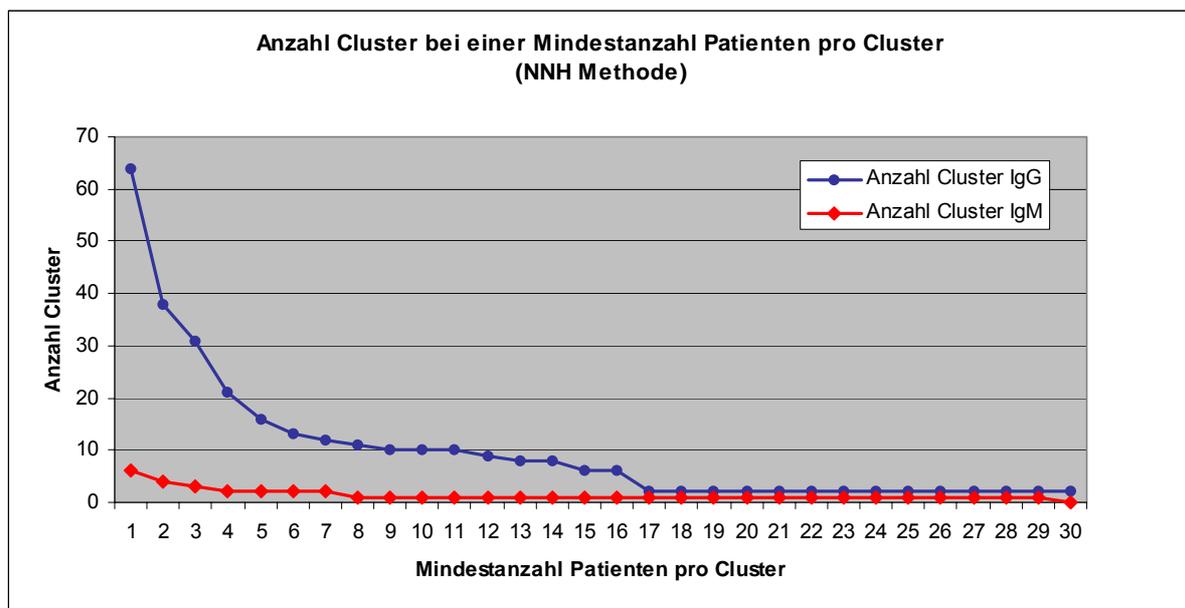


Abb. 74: Anzahl Cluster bei festgelegter Mindestanzahl Patienten pro Cluster (NNH Methode)

In der vorliegenden Studie wurde M = 10, 15 und 20 gewählt, um eine realistische Mindest-Clustergröße zu erhalten und den Effekt des sich ändernden M zu untersuchen. Die zwei verbleibenden Cluster bei M=20 befinden sich im Stadtgebiet Frankfurt (Oder) sowie in Ei-

senhüttenstadt. Dies ist auch für IgM positive Patienten zu beobachten. Die Wahl der Größe M , d.h. die Anzahl Patienten pro Cluster ist nicht ganz unproblematisch. Es bleibt somit dem Untersuchenden überlassen, welche Mindestanzahl Patienten pro Cluster die geeignetste ist. Das Ergebnis einer zu geringen Anzahl Patienten ist eine Vielzahl von Clustern, die nur mühsam überprüfbar sind. In der vorliegenden Studie sind 10 bis 15 Patienten pro Cluster eine sinnvolle Größe. Um ein differenzierteres Bild über Cluster im Stadtgebiet Frankfurt (Oder) und Eisenhüttenstadt zu erhalten, wird anhand der *Risk-adjusted nearest neighbour Methode* die absolute Patientenzahl in der Clusteridentifikation berücksichtigt. Das Ergebnis ist die Identifikation verschiedener Cluster im engeren Stadtgebiet Frankfurt (Oder) und Eisenhüttenstadt. Die adressen-scharfe Geocodierung der Patienten erlaubt ihre sehr genaue räumliche Zuordnung. So ist es möglich, Straßenabschnitte mit einer Häufung IgG-positiver Patienten zu identifizieren (Abb. 61), deren Ursache durch weitergehende Untersuchungen (u.a. Vor-Ort-Untersuchungen) geklärt werden könnten. Bei dem verwendeten Verfahren der *Risk-Adjusted nearest neighbour hierarchical cluster analysis* bleibt es ebenso dem Untersucher überlassen, die Mindestzahl M positiver Probanden für die Identifikation eines Clusters festzulegen. Naturgemäß sinkt die Zahl der identifizierten Cluster, wenn M größer gewählt wird. Die Zahl der Cluster sinkt von 18 über neun auf schließlich nur noch zwei (Abb. 75).

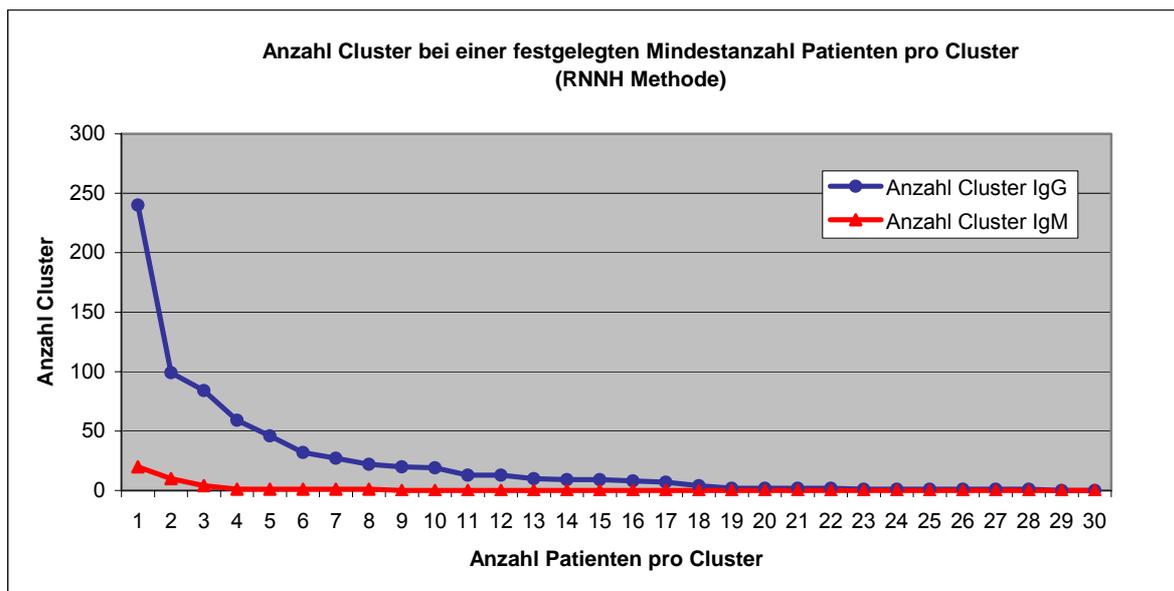


Abb. 75: Anzahl Cluster bei einer festgelegten Mindestanzahl Patienten pro Cluster

Bei den IgG positiven Patienten sinkt die Anzahl der identifizierten Cluster bei geringem M rapide ab, nur wenige Unterschiede ergeben sich in dem Bereich von $M=10-17$ Patienten. In diesem Bereich sollten auch eine eingehende Betrachtung der Cluster erfolgen. Die Modifikation der Parameter der Kernel density Interpolation verursacht vereinzelte Veränderungen in Hinblick auf die Clusteranzahl und Clustergröße. Aufgrund der Erkenntnisse aus der vorangegangenen Kernel-density-Interpolation (starke Differenzierung der Punktdichte) sollte ein variabler Suchradius verwendet werden.

Die Clusterberechnung mit Hilfe der *Risk-adjusted nearest neighbour hierarchical clustering Methode* liefert die geeignetsten Ergebnisse zur Identifikation von Clustern, da die absolute Patientenzahl bei der Clusterberechnung berücksichtigt wird. Sowohl die *Nearest neighbour hierarchical clustering Methode* als auch die Kernel-density-Interpolation sind hilfreiche Me-

thoden um erste Hotspots zu identifizieren. Jedoch unterliegen diese Methoden dem Einfluss der Basisvariable (z.B. absolute Bevölkerungszahl oder hier Patientenzahl). Eine nacheinander geschaltete Anwendung der Methoden ermöglicht ein Verständnis für die existierende Punktverteilung.

Die identifizierten Cluster bilden jedoch kein akutes, sondern ein kumulatives oder länger zurückliegendes Geschehen ab. Ein derartiger Ausbruch könnte sich in der Infektionsstatistik des Klinikums in der Form einer Häufung wohnbenachbarter Patienten mit akuter Legionellose mitgeteilt haben. Dies ließe sich auch retrospektiv durch die Zusammenstellung der Adressen von Legionellose-Patienten im Klinikum überprüfen.

7. Ausblick

Die Ursachen für die gefundenen Cluster IgG-positiver Patienten werden im Rahmen dieser Studie nicht näher untersucht. Man kann vermuten, dass hierfür Faktoren der technisch konditionierten Umwelt eine Rolle spielen. Für eine Bestätigung dieser Hypothese bedarf es jedoch gründlicher Ortsinspektionen und ggf. hygienisch-mikrobiologischer Untersuchungen sowie weitergehender geostatistischer Analysen unter Einbeziehung der gewonnenen umwelthygienischen Detailinformationen.

Ein weiterer wichtiger Anhaltspunkt in dieser Studie, eng verknüpft mit der räumlichen Verteilung der Patienten im Untersuchungsgebiet, ist die Frage, welche Krankenhäuser von den Patienten im Untersuchungsgebiet aufgesucht werden und zu welchen Anteilen. Denn das Land Brandenburg besitzt eine Eigenversorgungsquote von nur 89,7%, d.h. 10,3% der Patienten werden außerhalb des eigenen Bundeslandes versorgt. Allein rund 9% der Patienten gehen nach Berlin, um sich dort behandeln zu lassen. Es steht also eine räumliche Analyse der Krankenhausinfrastruktur der Region aus. Auch dies sollte mit Unterstützung eines GIS durchgeführt werden, da es Möglichkeiten der Visualisierung und des Einsatzes räumlicher Analysemethoden bietet.

Die in dieser Studie verwendeten Methoden haben dazu verholfen, Häufungen von Patienten unter Berücksichtigung der absoluten Patientenzahl zu erkennen. Dies wäre ohne die Anwendung dieser GIS-gestützten Methoden erheblich erschwert gewesen, zumal die Ergebnisse der statistischen Analyse (Risikoanalyse) keine deutlichen Hinweise auf umweltassoziierte Risiken erbrachten. Die Kombination von statistischen und räumlich-statistischen Methoden ist im Rahmen einer solchen Studie ein geeignetes Mittel zur Identifikation von Risikofaktoren und ist daher für die Praxis des Öffentlichen Gesundheitswesens zu empfehlen.

Trotzdem muss berücksichtigt werden, dass insbesondere im Öffentlichen Gesundheitswesen das Vorhandensein eines solchen umfangreichen Datenpools eher selten ist. Meist liegen Daten nur auf Kreisebene vor und bieten sich daher nur für einen Teil der angewendeten Methoden an (z.B. Ermittlung der räumlichen Autokorrelation). Die Erhebung von Patientendaten auf Ebene der Straßen oder Hausnummern mit Hilfe der Geocodierung ermöglicht die Anwendung weitergehende räumliche Analysemethoden und sollte in Zukunft, wenn möglich, entsprechend anonymisiert eingesetzt werden.

Der Einsatz von GIS in gesundheitlichen Einrichtungen soll nicht nur in Deutschland sondern auch in anderen Ländern in Zukunft eine herausragende Rolle spielen. Dies erfordert auch die Überwindung von Anwendungsbarrieren, wie sie in vielen Institutionen bestehen. *„Despite of the demonstrations of the usefulness of techniques, such as GIS and point pattern analysis in academic-health service collaborations, it is clear that there remain barriers to their widespread use...There remains though a need for greater levels of understanding of, andn confidence in, such techniques amongst health service professionals, coupled with a need for researchers to ensure, that they are developing techniques that are suitable for operational and policy use within the health sector”.* (Cockings et al. 2004: 180). Zu der Überwindung derartiger Anwendungsbarrieren zählt u.a. eine regelmäßige Schulung von Mitarbeitern, die in der Praxis des Öffentlichen Gesundheitswesens GIS einsetzen möchten.

Gegen einen Einsatz von GIS sprechen zudem häufig die Anschaffungskosten des jeweiligen Software-Produktes. Jedoch sind mittlerweile verschiedene kostenlose Produkte erhältlich (z.B. Crimestat 2.0), die über eine Vielzahl Analysefunktionen verfügen.

Somit sind GIS-Funktionalitäten, insbesondere räumliche-statistische Analysemethoden, für die Praxis des Öffentlichen Gesundheitswesens, anfänglich begleitet durch entsprechende Schulungsmaßnahmen, zu empfehlen. Sie leisten, wie auch diese Studie zeigen konnte, einen wichtigen Beitrag zur Lösung räumlich-statistischer Fragestellungen, insbesondere dann, wenn rein statistische Methoden zur Lösung des Problems nicht herangezogen werden können.

8. Zusammenfassung

Der Anwendung Geographischer Informationssysteme fällt in allen Bereichen der Gesundheitsforschung eine zentrale Rolle zu, unabhängig davon, ob es um die Beschreibung und Erklärung raumbezogener Variation von Krankheit oder um die Planung von Gesundheitseinrichtungen geht (Gatrell und Senior 1999). Die Möglichkeiten und Grenzen räumlich-statistischer Analysemethoden mit GIS soll am Beispiel einer Studie zur Verbreitung von Legionelleninfektionen im Einzugsgebiet eines Brandenburger Krankenhauses dargestellt und diskutiert werden.

Vor dem Hintergrund der ungeklärten Ursache des gehäuften Auftretens von Legionellose in einem Brandenburger Klinikum in den Jahren 2002 und 2003 wurde im Auftrag des Klinikums eine prospektive epidemiologische Studie durchgeführt, bei der der Legionellen-Serostatus der Klinikumspatienten zum Zeitpunkt ihrer Aufnahme sowie mögliche Risikofaktoren für Legionella-Infektionen außerhalb des Klinikums erfasst und analysiert wurden.

Informationen zu den Patienten, die für eine mögliche Infektion mit Legionellen relevant sein können, wurden mittels Fragebogen zum Zeitpunkt der Aufnahme in das Klinikum erfasst. Dabei handelte es sich um allgemeine Patientendaten (Adresse mit Angabe von Straße und Hausnummer, Alter, Geschlecht, Zahl der Klinikumsaufnahmen), sowie um Angaben zu möglicherweise relevanten Risikofaktoren vor Aufnahme in das Klinikum: Reisen, der Besuch von Schwimmbädern oder sonstigen Sportstätten, der Aufenthalt in der Nähe von Luftbefeuchtern oder Rückkühlwerken. Außerdem wurde der Serostatus der Patienten bei Aufnahme erfasst. Die Daten wurden schwerpunktmäßig räumlich-statistisch (flächenhafte und punktbasierte Analyse) auf Grundlage der Geocodierung der Patientenadressen, untersucht. Des Weiteren wurden die Ergebnisse des Fragebogens und der Serologie deskriptiv-statistisch sowie analytisch-epidemiologisch ausgewertet. Aus dem Bereich der GIS-gestützten räumlichen Analyse wurden vornehmlich Verfahren der Point Pattern Analysis (PPA-Methoden) angewendet, und zwar im einzelnen

1. zur Bestimmung der räumlichen Punkteverteilung;
2. zur Bestimmung der räumlichen Punktdichte;
3. zur Identifikation räumlicher Cluster, d.h. signifikant höherer Konzentrationen von Patienten in Teilbereichen des Untersuchungsgebietes

Insgesamt wurden 51 IgM-positive Patienten und 777 IgG-positive Patienten ermittelt, wovon 13 sowohl IgM- als auch IgG-positiv waren. Bei einem Patienten wurde bei Aufnahme innerhalb des Untersuchungszeitraums Legionella-Antigen im Urin nachgewiesen.

Von den untersuchten Umweltfaktoren war nur der Besuch von Schwimmbädern und anderen Sportstätten signifikant. Andere als typisch berichtete Risikofaktoren der allgemeinen Umwelt (Luftbefeuchter, Rückkühlwerke) konnten hingegen nicht bestätigt werden. Es muss allerdings davon ausgegangen werden, dass viele Patienten mit der richtigen Beantwortung dieser Fragen überfordert waren.

Die flächenhafte räumliche Analyse ergab erhöhte Prävalenzen serologisch positiver Patienten für einzelne Postleitzahlregionen in ländlichen Regionen des Landes Brandenburg sowie

in den Teilen der Stadt Frankfurt (Oder) sowie in Eisenhüttenstadt. Problematisch ist hierbei die Überbetonung von Gebieten mit einer geringen absoluten Patientenzahl.

Bei der räumlichen Analyse der Punktdaten ergab die *Risk-Adjusted Nearest Neighbour Hierarchical Cluster Analysis* die am besten geeigneten Ergebnisse. Dieses Verfahren identifiziert Cluster serologisch positiver Patienten unter Berücksichtigung der absoluten Patientenzahl, d.h. die Patientendichte wird bei der Ermittlung von Clustern integriert.

Für Mindestclustergrößen von 10, 15 und 20 IgG-positiven Patienten wurden 18, acht bzw. zwei Cluster identifiziert. Die adressen-scharfe Geocodierung der Patienten ermöglichte es, Straßenabschnitte mit einer Häufung IgG-positiver Patienten zu identifizieren.

Die Ursachen für die gefundenen Cluster wurden nicht näher untersucht. Möglicherweise spielen Faktoren der technisch konditionierten Umwelt eine Rolle. Für eine Bestätigung dieser Hypothese bedarf es jedoch gründlicher hygienischer Ortsinspektionen und ggf. hygienisch-mikrobiologischer Untersuchungen sowie weitergehender geostatistischer Analysen unter Einbeziehung der gewonnenen umwelthygienischen Detailinformationen.

Die identifizierten Cluster bilden kein akutes, sondern ein kumulatives oder länger zurückliegendes Geschehen ab. Ein derartiger Ausbruch könnte sich in der Infektionsstatistik des Klinikums in der Form einer Häufung wohnbenachbarter Patienten mit akuter Legionellose mitgeteilt haben. Dies ließe sich auch retrospektiv durch die Zusammenstellung der Adressen von Legionellose-Patienten im Klinikum überprüfen.

9. Literatur

- BBR (Bundesamt für Bauwesen und Raumordnung) (1999): Die Gebietsreform der neuen Länder. Folgen für die laufende Raumbewertung des BBR. Arbeitspapiere Heft 5/1999. Bonn
- Barret, F.A. (2000): Finke's 1792 map of human diseases: the first world disease map? In: *Social Science & Medicine* 50: S. 915-921
- Bailey, T. C. (1998): A review of statistical spatial analysis in GIS. In S. Fortheringham (Hrsg.): *Spatial analysis and GIS*. London: 13-44
- Bailey, T. C. & A. C. Gatrell (1995): *Interactive spatial data analysis*. Burnt Mill, Essex, England
- Becker, N., Wahrendorf, J. (1998): *Krebsatlas der Bundesrepublik Deutschland – 1981-1990*. Heidelberg, 3. Auflage
- Bhopal, R.S. (1993): Geographical variation of Legionnaires' disease: a critique guide to future research. *Int J Epidemiol* 22: 1127-1136
- Blaschke, Th. (1997): *Landschaftsanalyse und –bewertung im GIS*. Forschungen zur deutschen Landeskunde. Band 43. Trier
- Boer, J.W. den, et al. (2002): A large outbreak of Legionnaires' disease at a flower show, the Netherlands, 1999. *Emerging Infectious Diseases* 8: 37-43
- Bolous, M.N.K. (2004): Towards evidence-based, GIS-driven spatial health information infrastructure and surveillance services in the United Kingdom. *International Journal of Health Geographics* 3 (1): 1-50
- Bhopal, R. S., Diggle, P. & B. Rowlingson (1992): Pinpointing clusters of apparently sporadic cases of Legionnaires' disease. *British Medical Journal* 304. 1022-1027
- Brody, H., Rip, M.R., Vinten-Johansen, P., Paneth, N. & S. Rachman (2000): Map-making and myth-making in Broad Street: the London cholera epidemic, 1854. *The Lancet* 356: S. 64-68
- Bühl, A., Zöfel, P. (2004): *SPSS 12. Einführung in die moderne Datenanalyse unter Windows*. München
- Burrough, P. (1986): *Principles of Geographic Information Systems for land resources assessment*. Oxford
- Burrough, P. (1998): *Principles of Geographic Information Systems*. New York
- Burt, J. E. & Barber, G. M. (1996): *Elementary statistics for Geographers*. New York, 2. ed.
- Campese, C., Decludt, B. (2002): Notified cases of legionnaires' disease in France in 2001. *Euro Surveill* 7: 121-128

- Cockings, S., Dunn, C., Bhopal, R.S. & R. D. Walker (2004): Users' perspectives on epidemiological, GIS and point pattern approaches to analysing environment and health data. *Health & Place* 10: 169-182
- Cromely, E. & S. McLafferty (2002): *GIS and Public Health*. New York
- Cressie, N. A. C. (1993): *Statistics for spatial data*. New York
- Curtis, A., Leitner, M., Hanlon, C. (2002): Using hierarchical nearest neighbour analysis and animation to investigate the spatial and temporal patterns of raccoon rabies in West Virginia. In Khan, O. A. & R. Skinner (Hrsg.): *Geographic Information Systems and Health Applications*. London: 155-171
- Dreesmann, J. (2004): Statistik für raumbezogene Daten. In Schweikart, J. & Th. Kistemann (Hrsg.): *GIS im Gesundheitswesen*. Heidelberg: 71-90
- Diggle, P. J. (1985): *Statistical analysis of spatial point patterns*. Paris
- Diggle, P. J. (1993): Point process modelling in environmental epidemiology. In Barnett V. & K. F. Turkman (Hrsg.): *Statistics for the environment*. Chichester
- Diggle, P. J. & B. Rowlingson (1994): A conditional approach to point process modelling of elevated risk. *Journal of the Royal Statistical Society Series A* 157: 433-440
- Elliot, P. & Wartenberg, D. (2004): Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives* 112 (9): 998-1006
- European Community (1991): *Atlas of Avoidable Death*. Oxford
- Exner M., Pleischl, S., Rechenburg, A., Hornei, B., Wieland, A., Kistemann, T., Dangendorf, F. (2002): *Prevention and Control of Legionnaires' Disease in all Settings*. Draft Review for WHO HQ, Geneva (unpublished)
- Exner, M. (1990): Verhütung, Erkennung und Bekämpfung von Legionellen-Infektionen im Krankenhaus. *Forum Städte-Hygiene* 42: 178-191
- Exner, M. (1997a): Legionellose. In: F. Hofmann (Hrsg.): *Infektiologie*. 6. Ergänzungslieferung 8/93: 1-18
- Exner, M. (1997b): Legionellose. In: Beck, E.G. & T. Eikmann (Hrsg.): *Hygiene in Krankenhaus und Praxis*. Ecomed Verlagsgesellschaft, Landsberg.
- Fry, C. (1999): GIS in telecommunications. In Longley, P., Goodchild, M. Maguire, D. & D. Rhind (Hrsg.): *Geographical Information System: Management issues and applications*. Band 2. New York: 819-826
- Gans, P., Kistemann, T., Schweikart, J. (2001): Regionale Unterschiede der Lebenserwartung. In: Institut für Länderkunde, Leipzig (Hrsg.): *Nationalatlas der Bundesrepublik Deutschland – Bevölkerung*. Spektrum Akademischer Verlag, Heidelberg-Berlin: 98-99
- Gatrell, A. C. & T. C. Bailey (1996): Interactive spatial data analysis in medical geography. *Social Science and Medicine* 42(6): 843-855

- Gatrell, A. C., Bailey, T. C., Diggle, P. J. & B. S. Rowlingson (1996): Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers* 21: 256-274
- Gatrell, A.C. & M.L. Senior (1999): Health and healthcare applications. In: Longley, P.A., Maguire, D.J., Goodchild, M.F. & D.W. Rhind (Hrsg.): *Geographical Information Systems*. Chichester. S. 925-938
- Gatrell, A.C. & M. Loytonen (1998): *GIS and Health*. London
- Gesler, W. (1986): The uses of spatial analysis in medical geography: a review. In *Social Science and Medicine* 23/ 10: 963-973
- Gould, M. I. (1992): The use of GIS and CAC by health authorities: results from a postal questionnaire. *Area* 24: 391-401
- Goodchild, M. F. (1993): Ten years ahead: Dobson's automated Geography in 1993. *Professional Geographer* 45: 444-445
- Groothuis, D.G., Veenendaal, H.R. & H.L. Dijkstra (1985): Influence of temperature on the number of *Legionella pneumophila* in hot water systems. *Journal of Applied Bacteriology* 59 (6): 529-536
- Haraldsson, A., Rechnitzer, C., Friis-Moller, A., Briem, H. (1990): Prevalence of IgM antibodies to nine *Legionella* species in Icelandic children. *Scand J Infect Dis* 22: 445-449.
- Hartigan, J. A. (1975): *Clustering Algorithms*. New York
- Higgs, G. & Gould, M. (2001): Is there a role for GIS in the 'new NHS'? *Health and Place* 7: 249-259
- Hodder, I. & Orton, C. (1976): *Spatial analysis in archaeology*. Cambridge
- Howe, G.M. (1963): *National atlas of disease mortality in the United Kingdom*. London
- Joseph, C.A. (2004): Legionnaires' disease in Europe. *Epidemiology and Infection* 132: 417-424
- Kelsall, J. E. & Diggle, P. J. (1995): Kernel estimation of relative risk. *Bernoulli* 1: 3-16
- King, B. F. (1967): Step wise clustering procedures. *Journal of the American Statistical Association* 62: 86-101
- Kistemann, Th., Dangendorf, F., Schweikart, J. (2002): New perspectives on the use of Geographical Information Systems (GIS) in environmental health sciences. Vol. 205: 161-181
- Kistemann, Th., Stalleicken, I., Hornei, B., Fischnaller, E., Exner, M. (2004): Studie zur Legionellenproblematik am Klinikum Frankfurt (Oder) GmbH. Gutachten (unveröffentlicht). Bonn
- Kool, J. L., Bergmire-Sweat, D., Butler, J. C., Brown, E. W., Peabody, D. J., Massi, D. S., Carpenter, J. C., Pruckler, J. M., Benson, R. F., Fields, B. S. (1999): Hospital characteris-

- tics associated with the colonization of water systems by Legionella and risk of nosocomial Legionnaires' disease: a cohort study of 15 hospitals. *Infection Control and Hospital Epidemiology* 20/12: 798-805
- Kreienbock, Schach, S. (1995): *Epidemiologische Methoden*. Spektrum Verlag, Heidelberg
- Krafft, Th., König, A. & Kortevoß, A. (2004): Optimierung der Versorgung chronisch Kranker – GIS-basierte Analyse eines Disease Management-Programms (DMP) für Diabetes. In: Schweikart, J. und Th. Kistemann (Hrsg.): *GIS im Gesundheitswesen*. Heidelberg: 211-224
- Lawson, A.B. (2001): *Statistical methods in spatial epidemiology*. New York
- Levine, N. (2002): *Crimestat II. A spatial statistics program für the analysis of crime incident locations*. The National Institute of Justice. Washington
- Longley, P. A. & G. Clarke (1995): *GIS for Business and service planning* GeoInformation International. Cambridge
- Longley, P. A., Goodchild, M. F., Maguire, M. J. & D. W. Rhind (Hrsg.) (2001): *Geographical Information System and Science*. New York
- Lorup, E. (2004): *Point Pattern Analysis – ein Überblick*. Unveröffentl. Manuskript zum Hochschullehrgang der Universität Salzburg. Salzburg
- De Lepper, M. J. C. , Scholten, H. J. und R. M. Stern (1994): The added value of Geographical Information Systems in Public Health and Environmental Health. *The GeoJournal Library* 24. Dordrecht/ Boston/ London
- Maheswaran, R. & M. Craglia (2004): *GIS in Public Health Practice*. London
- Marston, B.J., Lipman, H.B. & Breiman, R.F. (1994) Surveillance for Legionnaires' disease. Risk factors for morbidity and mortality, *Arch Intern Med*, 154(21): S. 2417-2422
- Mott, K. E., Nutall, I., Desjeux, P. & P. Cattand (1995): New geographical approaches to control of some zoonoses. *Bulletin of the World Health Organization* 73 (2): 247-257
- Mitchell, A. (1999): *GIS Analysis*. Redlands, California
- McQuitty, L. L. (1960): Hierarchical syndrome analysis. *Educational and Psychological Measurement* 20: 293-304
- McCoy, J. & K. Johnston (2001): *Using ArcGIS spatial analyst*. Handbuch zur ArcGIS-Extension spatial analyst
- O'Sullivan, D. (2003): *Geographic Information Analysis*. New Jersey
- Pagano, E. & Raeber, A. (1994): *Legionellen und Legionellose*. Bern. <http://www.bag.admin.ch/infekt/pub/wissenschaft/d/legiod1.pdf> (11.12.2004)
- Parzen, E. (1962): On the estimation of a probability density and mode. *Annals of Mathematical Statistics* 33: 1065-1076

- Pleischl, S., Engelhart, S., Bailly, E., Exner, M. (2001) Klinische und rechtliche Aspekte der Prävention und Kontrolle von Legionella-Infektionen. *Umweltmed Forsch Prax* 6: 193-201
- Pleischl, S., Engelhardt, S. & M. Exner (o.A.): Arbeitspapier Legionellose: Medizinische und umwelthygienische Gesichtspunkte. <http://www.meb.uni-bonn.de/hygiene/forschung/legionellose.pdf> (08.01.2005)
- Reader, S. (1995): The present state of GIS and future trends. In de Savigny, D. & Wijayarathne, P. (Hsrg.): *GIS for health and the environment (Proceedings of an international workshop held in Colombo, Sri Lanka, 5 -10 September 1994)*. Ottawa: 1-12
- Ripley, B. D. (1976): The second-order analysis of stationary point processes. *Journal of Applied Probability* 13: 255-266
- Ripley, B. D. (1977): Modelling spatial patterns. *Journal of the Royal Statistical Society, Series B* 39: 172-192
- Ripley, B. D. (1981): *Spatial statistics*. Chichester
- Rissland, J., van Treek, U., Traeger, D. & Baumeister, H. G. (2003): Infektionssurveillance in Nordrhein- Westfalen. *Standardberichte, Barometer und Frühwarnsystem Gesundheitswesen* 65: 719-723
- RKI (1999): Ratgeber Infektionskrankheiten. 10. Folge: Legionellose. *Epidemiologisches Bulletin* 49: S. 369-372
- RKI (2003): Legionellose im Jahr 2002. *Epidemiologisches Bulletin* Nr. 45: 361-365
- RKI (2004): Zu zwei nosokomialen Legionellose-Ausbrüchen in einem Klinikum im Land Brandenburg. *Epidemiologisches Bulletin* 11: S. 89-91
- Rodenwaldt, E. & Jusatz, H. (1952-1961) : *World Atlas of Epidemic Diseases*. Vol. 1-3. Hamburg
- Rosenblatt, M. (1956): Remarks on some non-parametric estimates of a density function. *Annals of Mathematical Statistics* 27: 832-837
- Schweikart, J. (2004): GIS: ein Modell der Welt mit Raumbezug – Grundlagen der Geoinformationssysteme. In Schweikart, J. & Kistemann, Th. (Hrsg.): *Geoinformationssysteme im Gesundheitswesen*. Heidelberg: S. 17-36
- Schweikart, J. & Kistemann, Th. (2004): Geoinformation in der Gesundheit – Entwicklung und Potenziale. In: Schweikart, J. & Th. Kistemann (Hsrg.): *Geoinformationssysteme im Gesundheitswesen*. Heidelberg: 3-14
- Smallman-Raynor, M.R., Cliff, A.D. & P. Hagett (1992): *London international atlas of AIDS*. Oxford/ Cambridge/ Massachusetts
- Sneath, P. H. A. (1957): The application of computers to taxonomy. *Journal of General Microbiology* 17: 201-226
- Snow, J. (1855): *On the mode of communication of cholera*. London, 2. Auflage

- Sokal, R. R. & P. H. A. Sneath (1963): Principles of Numerical Taxonomy. San Francisco
- Sokal, R. R. & C.D. Michener (1958): A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 38: 1409-1438
- Stahl, R. (1997): Wie alles begann: Ein kurzer historischer Überblick. GIS-Tutorial. <http://www.gis-tutor.de> (02.01.2005)
- Stalleicken, I. (2004): Überwachung der Influenza. In: Schweikart, J. und Kistemann Th. (Hrsg): Geoinformationssysteme im Gesundheitswesen. Grundlagen und Anwendungen. Heidelberg: S. 173-188
- Stevenson, L. (1965): Putting disease on the map: the early use of spot maps in the study of yellow fever. In: Journal of the History of Medicine and Allied Sciences 20: S. 226-261
- Strobl, J. (1988): Digitale Forstkarte und Forsteinrichtung. Salzburger Geographische Hefte 12. Salzburg
- Strobl, J. (1992): Datenmanipulation und Datenanalyse. In A. Kilchenmann (Hrsg.): Technologie Geographischer Informationssysteme. Berlin: 47-56
- Strobl, J. (1994-2003): Räumliche Interpolationsmethoden. Unveröffentl. Manuskript zum Hochschullehrgang der Universität Salzburg. Salzburg
- Tiefenbrunner, F. (2002): Legionellen im Krankenhaus: Vorkommen, Risikofaktoren und Hygienemaßnahmen. Krankenhaushygiene und Infektionsverhütung 24/5: S 173-183
- Traub, K.-P. (2004): Die Untersuchung der räumlichen Verbreitung von Gesundheitseinrichtungen mit GIS am Beispiel Hamburg. In: Schweikart, J. und Kistemann Th. (Hrsg): Geoinformationssysteme im Gesundheitswesen. Grundlagen und Anwendungen. Heidelberg: 241-255
- Uphoff, H., Stalleicken, I., Bartelds, A., Phiesel, B., Kistemann, Th. (2004): Are influenza data useful for mapping presentations? Virus Research 103: 35-46
- Upton, G. J. & B. Fingleton (1985): Spatial statistics by example. Vol.1: Point pattern and quantitative data. New York
- Van Steenberg, J.E., Slijkerman, F.A. & P. Speelman (1999): The first 48 hours of investigation and intervention of an outbreak of legionellosis in the Netherlands. Euro Surveillance 4 (11): 111-115
- Walter, S. D. & S. E. Birnie (1991): Mapping mortality and morbidity patterns: an international comparison. International Journal of Epidemiology 20: 678-689
- Ward, J. H. (1963): Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58: 236-244
- Waters, N. (1999): Transportation GIS: GIS-T. In Longley, P., Goodchild, M. Maguire, D. & D. Rhind (Hrsg.): Geographical Information Systems: Management issues applications. Band 2. New York: 827-844

- Whittle, P. (1958): On the smoothing of probability density functions. *Journal of the Royal Statistical Society, Series B* 55: 549-557
- WHO (2004): GIS and public health mapping. <http://www.who.int/csr/mapping/gisandphm/en/print.html> (26.08.2004)
- Wright, D. J., Goodchild, M. F., Proctor, J. D. (1997): Demystifying the persistent ambiguity of GIS as "tool" versus "science". *The annals of the Association of American Geographers* 87: 346-362
- Yeh, A.-G.-O. (1999): Urban planning and GIS. In Longley, P., Goodchild, M. Maguire, D. & D. Rhind (Hrsg.): *Geographical Information Systems: Management issues applications*. Band 2. New York: 877-888
- Yu, V.L. (1995): *Legionella pneumophila (Legionnaires' Disease)*. In: Mandell, G.L., Bennett, J.E., Dolin, R. (Hrsg.): *Principles and Practice of Infectious Diseases*. New York
- Zöllner, I. (2004): Sachdatenerhebung und methodische Probleme bei kleinräumigen epidemiologischen Studien. In Schweikart, J. & Th. Kistemann (Hrsg.): *GIS im Gesundheitswesen*. Heidelberg: 55-70

10. Anhang

1. a) Ergebnis der Berechnung der K-Funktion für IgG positive Patienten

BIN	DISTANCE	L(t)	L(CSR)	L(t) MIN	L(t) MAX	L(t) 0,5	L(t) 2,5	L(t) 97,5	L(t) 99,5
1	0,99	22,94	0,00	2,66	2,77	2,66	2,67	2,76	2,77
2	1,97	35,91	0,00	2,01	2,22	2,01	2,02	2,19	2,22
3	2,96	43,83	0,00	1,50	1,82	1,50	1,55	1,78	1,82
4	3,95	45,80	0,00	1,19	1,55	1,19	1,20	1,48	1,55
5	4,94	46,43	0,00	0,92	1,34	0,92	0,96	1,26	1,34
6	5,92	46,66	0,00	0,74	1,09	0,74	0,76	1,09	1,09
7	6,91	46,53	0,00	0,62	0,99	0,62	0,63	0,91	0,99
8	7,90	46,29	0,00	0,46	0,87	0,46	0,48	0,85	0,87
9	8,88	46,27	0,00	0,35	0,78	0,35	0,38	0,72	0,78
10	9,87	46,56	0,00	0,22	0,73	0,22	0,25	0,59	0,73
11	10,86	47,13	0,00	0,12	0,69	0,12	0,12	0,48	0,69
12	11,85	47,81	0,00	0,00	0,61	0,00	0,01	0,43	0,61
13	12,83	48,37	0,00	-0,15	0,47	-0,15	-0,12	0,38	0,47
14	13,82	48,63	0,00	-0,28	0,36	-0,28	-0,23	0,27	0,36
15	14,81	48,51	0,00	-0,41	0,27	-0,41	-0,34	0,22	0,27
16	15,79	48,15	0,00	-0,48	0,15	-0,48	-0,45	0,14	0,15
17	16,78	47,78	0,00	-0,58	0,07	-0,58	-0,57	0,04	0,07
18	17,77	47,61	0,00	-0,72	0,01	-0,72	-0,71	-0,09	0,01
19	18,76	47,56	0,00	-0,89	-0,12	-0,89	-0,81	-0,16	-0,12
20	19,74	47,55	0,00	-1,04	-0,17	-1,04	-0,91	-0,24	-0,17
21	20,73	47,65	0,00	-1,16	-0,25	-1,16	-1,06	-0,37	-0,25
22	21,72	48,37	0,00	-1,33	-0,38	-1,33	-1,19	-0,50	-0,38
23	22,70	49,68	0,00	-1,47	-0,50	-1,47	-1,31	-0,62	-0,50
24	23,69	51,36	0,00	-1,61	-0,62	-1,61	-1,44	-0,74	-0,62
25	24,68	52,42	0,00	-1,77	-0,72	-1,77	-1,60	-0,85	-0,72
26	25,66	52,85	0,00	-1,91	-0,85	-1,91	-1,73	-0,96	-0,85
27	26,65	52,85	0,00	-2,07	-0,91	-2,07	-1,90	-1,08	-0,91
28	27,64	52,85	0,00	-2,21	-1,02	-2,21	-2,06	-1,19	-1,02
29	28,63	52,68	0,00	-2,36	-1,14	-2,36	-2,21	-1,33	-1,14
30	29,61	52,33	0,00	-2,52	-1,30	-2,52	-2,37	-1,47	-1,30
31	30,60	51,86	0,00	-2,69	-1,45	-2,69	-2,52	-1,62	-1,45
32	31,59	51,51	0,00	-2,86	-1,57	-2,86	-2,67	-1,75	-1,57
33	32,57	51,38	0,00	-3,03	-1,74	-3,03	-2,84	-1,88	-1,74
34	33,56	51,18	0,00	-3,20	-1,88	-3,20	-3,05	-1,99	-1,88
35	34,55	50,69	0,00	-3,36	-2,01	-3,36	-3,23	-2,16	-2,01
36	35,54	50,08	0,00	-3,51	-2,18	-3,51	-3,45	-2,33	-2,18
37	36,52	49,42	0,00	-3,70	-2,36	-3,70	-3,63	-2,50	-2,36
38	37,51	48,74	0,00	-3,94	-2,56	-3,94	-3,85	-2,65	-2,56
39	38,50	48,02	0,00	-4,15	-2,72	-4,15	-4,02	-2,81	-2,72
40	39,48	47,30	0,00	-4,35	-2,91	-4,35	-4,25	-2,94	-2,91
41	40,47	46,58	0,00	-4,56	-3,09	-4,56	-4,48	-3,10	-3,09
42	41,46	45,83	0,00	-4,78	-3,24	-4,78	-4,70	-3,29	-3,24
43	42,45	45,07	0,00	-5,03	-3,41	-5,03	-4,92	-3,46	-3,41
44	43,43	44,32	0,00	-5,27	-3,61	-5,27	-5,15	-3,64	-3,61
45	44,42	43,64	0,00	-5,50	-3,80	-5,50	-5,38	-3,84	-3,80
46	45,41	43,02	0,00	-5,74	-3,98	-5,74	-5,61	-4,06	-3,98
47	46,39	42,56	0,00	-5,99	-4,18	-5,99	-5,86	-4,28	-4,18
48	47,38	42,05	0,00	-6,23	-4,35	-6,23	-6,13	-4,48	-4,35
49	48,37	41,43	0,00	-6,48	-4,56	-6,48	-6,37	-4,69	-4,56
50	49,36	40,73	0,00	-6,75	-4,76	-6,75	-6,63	-4,89	-4,76

BIN	DISTANCE	L(t)	L(CSR)	L(t) MIN	L(t) MAX	L(t) 0,5	L(t) 2,5	L(t) 97,5	L(t) 99,5
51	50,34	40,06	0,00	-7,05	-4,94	-7,05	-6,92	-5,12	-4,94
52	51,33	39,47	0,00	-7,33	-5,16	-7,33	-7,20	-5,35	-5,16
53	52,32	38,89	0,00	-7,60	-5,42	-7,60	-7,49	-5,57	-5,42
54	53,30	38,20	0,00	-7,90	-5,63	-7,90	-7,77	-5,81	-5,63
55	54,29	37,40	0,00	-8,17	-5,86	-8,17	-8,05	-6,05	-5,86
56	55,28	36,62	0,00	-8,45	-6,08	-8,45	-8,35	-6,30	-6,08
57	56,27	35,84	0,00	-8,76	-6,29	-8,76	-8,61	-6,57	-6,29
58	57,25	35,09	0,00	-9,05	-6,52	-9,05	-8,91	-6,81	-6,52
59	58,24	34,30	0,00	-9,34	-6,78	-9,34	-9,23	-7,08	-6,78
60	59,23	33,56	0,00	-9,64	-7,04	-9,64	-9,52	-7,36	-7,04
61	60,21	32,82	0,00	-9,96	-7,29	-9,96	-9,82	-7,61	-7,29
62	61,20	32,09	0,00	-10,26	-7,55	-10,26	-10,15	-7,90	-7,55
63	62,19	31,39	0,00	-10,61	-7,83	-10,61	-10,46	-8,20	-7,83
64	63,17	30,74	0,00	-10,96	-8,10	-10,96	-10,76	-8,46	-8,10
65	64,16	30,08	0,00	-11,28	-8,40	-11,28	-11,09	-8,76	-8,40
66	65,15	29,36	0,00	-11,65	-8,67	-11,65	-11,41	-9,06	-8,67
67	66,14	28,53	0,00	-12,00	-8,95	-12,00	-11,77	-9,38	-8,95
68	67,12	27,67	0,00	-12,34	-9,24	-12,34	-12,09	-9,69	-9,24
69	68,11	26,80	0,00	-12,67	-9,55	-12,67	-12,45	-10,02	-9,55
70	69,10	25,92	0,00	-13,04	-9,84	-13,04	-12,78	-10,33	-9,84
71	70,08	25,06	0,00	-13,37	-10,14	-13,37	-13,16	-10,67	-10,14
72	71,07	24,21	0,00	-13,75	-10,46	-13,75	-13,52	-10,99	-10,46
73	72,06	23,36	0,00	-14,12	-10,79	-14,12	-13,88	-11,33	-10,79
74	73,05	22,49	0,00	-14,52	-11,12	-14,52	-14,26	-11,65	-11,12
75	74,03	21,61	0,00	-14,86	-11,46	-14,86	-14,64	-12,00	-11,46
76	75,02	20,73	0,00	-15,23	-11,84	-15,23	-15,00	-12,33	-11,84
77	76,01	19,83	0,00	-15,61	-12,21	-15,61	-15,39	-12,70	-12,21
78	76,99	18,91	0,00	-16,01	-12,58	-16,01	-15,78	-13,04	-12,58
79	77,98	18,00	0,00	-16,39	-12,94	-16,39	-16,16	-13,40	-12,94
80	78,97	17,09	0,00	-16,78	-13,29	-16,78	-16,58	-13,78	-13,29
81	79,96	16,17	0,00	-17,17	-13,65	-17,17	-16,96	-14,17	-13,65
82	80,94	15,25	0,00	-17,58	-14,04	-17,58	-17,37	-14,53	-14,04
83	81,93	14,37	0,00	-17,98	-14,42	-17,98	-17,80	-14,91	-14,42
84	82,92	13,48	0,00	-18,38	-14,80	-18,38	-18,21	-15,30	-14,80
85	83,90	12,56	0,00	-18,79	-15,19	-18,79	-18,62	-15,69	-15,19
86	84,89	11,62	0,00	-19,19	-15,59	-19,19	-19,05	-16,10	-15,59
87	85,88	10,68	0,00	-19,60	-15,99	-19,60	-19,47	-16,49	-15,99
88	86,87	9,77	0,00	-20,01	-16,41	-20,01	-19,90	-16,88	-16,41
89	87,85	8,85	0,00	-20,45	-16,81	-20,45	-20,34	-17,33	-16,81
90	88,84	7,94	0,00	-20,89	-17,25	-20,89	-20,80	-17,74	-17,25
91	89,83	6,99	0,00	-21,35	-17,69	-21,35	-21,25	-18,19	-17,69
92	90,81	6,02	0,00	-21,79	-18,12	-21,79	-21,70	-18,61	-18,12
93	91,80	5,05	0,00	-22,24	-18,56	-22,24	-22,17	-19,07	-18,56
94	92,79	4,08	0,00	-22,71	-19,03	-22,71	-22,64	-19,51	-19,03
95	93,78	3,11	0,00	-23,17	-19,50	-23,17	-23,12	-19,99	-19,50
96	94,76	2,15	0,00	-23,65	-19,97	-23,65	-23,60	-20,46	-19,97
97	95,75	1,19	0,00	-24,12	-20,46	-24,12	-24,06	-20,93	-20,46
98	96,74	0,22	0,00	-24,59	-20,94	-24,59	-24,52	-21,40	-20,94
99	97,72	-0,75	0,00	-25,08	-21,42	-25,08	-25,00	-21,88	-21,42
100	98,71	-1,73	0,00	-25,57	-21,90	-25,57	-25,47	-22,36	-21,90
101	99,70	-1,72	0,00	-25,07	-21,40	-25,07	-24,97	-21,89	-21,40

1. b) Ergebnis der Berechnung der K-Funktion für IgM positive Patienten

BIN	DISTANCE	L(t)	L(csr)	L(t) min	L(t) max	L(t) 0,5	L(t) 2,5	L(t) 97,5	L(t) 99,5
1	20,98	45,19	0,00	18,16	26,94	18,16	18,92	26,31	26,94
2	41,97	44,05	0,00	23,36	38,65	23,36	25,39	37,01	38,65
3	62,95	32,56	0,00	21,53	32,32	21,53	23,93	32,28	32,32
4	83,93	14,66	0,00	12,05	14,78	12,05	13,44	14,74	14,78
5	104,92	-0,04	0,00	-0,04	0,00	-0,04	0,00	0,00	0,00

Erklärung

Ich versichere, dass ich die Arbeit selbstständig verfasst habe, dass ich keine anderen Quellen und Hilfsmittel als die angegebenen benutzt und die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, in jedem Fall als Entlehnung kenntlich gemacht habe. Das gleiche gilt auch für beigegebene Zeichnungen, Kartenskizzen und Abbildungen.

Bonn, den 08.02.2005