

# Methoden zur Analyse der stationären Versorgung bei zeitkritischen Akuterkrankungen

---

Version 3



*Martin Scheuringer*

Hauptverband der Österreichischen Sozialversicherungsträger

Evidenzbasierte Wirtschaftliche Gesundheitsversorgung

Gesundheitsökonomie

# 1 ABSTRACT

Die Arbeiten zur stationären Versorgungsforschung stellen anhand von administrativen Daten die Versorgungssituation des stationären Sektors dar. Das Augenmerk liegt auf der Erreichbarkeit von Spitälern (Distanzen) und auf der regionalen Variabilität von Behandlungen. Es werden alle Aufenthalte, bestimmte geplante Eingriffe und im Gegensatz dazu zeitkritisch zu versorgende Akut-Erkrankungen (Schlaganfall, Herzinfarkt) analysiert.

Dabei ist eine differenzierte Betrachtung der Distanzen für die Qualität der Ergebnisse wesentlich. Insbesondere bei Akuterkrankungen, die einer unmittelbaren Versorgung bedürfen, ist das Funktionieren der Rettungs- und Behandlungskette oftmals entscheidend für den Behandlungserfolg. Die Behandlung erfolgt dabei im Idealfall in Krankenanstalten, die über eine geeignete Behandlungseinheit (z.B. Stroke-Unit, Herzkatheter) verfügen. Da jedoch im Notfall - zum Beispiel bei unklarer Symptomatik – auch Krankenanstalten angefahren/aufgesucht werden, die nicht der richtigen Versorgungsstruktur entsprechen, müssen PatientInnen mitunter in andere Krankenanstalten transferiert werden.

In den Analysen wird mit anonymisierten Daten gearbeitet. Das Fehlen eines eindeutigen Personenbezugs verhindert die Beurteilung von Behandlungsketten und führt zu Unschärfen in den Auswertungen (da die durchschnittlichen Distanzen und die regionale Variabilität verzerrt werden).

Eine Möglichkeit, die Datengrundlage zu verbessern besteht darin, die kodierten Transfers<sup>1</sup> zwischen Krankenanstalten bei der Analyse auszuschließen. Die Kodierqualität der Entlassungs- und Aufnahmekennzeichen in den verwendeten Daten (MBDS-Daten<sup>2</sup> des BMG) ist jedoch mangelhaft, weshalb es oftmals nicht möglich ist, die erstaufnehmende Krankenanstalt mit Sicherheit zu bestimmen.

Um die Validität der Ergebnisse der stationären Versorgungsforschung in Anbetracht der Einschränkungen der Datenqualität besser beurteilen zu können, werden verschiedene Modelle gerechnet. Dabei wird die Datengrundlage für die Berechnung variiert – es werden verschiedene Modellannahmen im Hinblick auf das Entlassungs- bzw. Aufnahmekennzeichens „Transfer“ getestet (welche Aufenthalte sollen in die Analyse eingeschlossen werden, welche nicht?).

Dieser Bericht dokumentiert das Bestimmen und Testen von 3 unterschiedlichen Modellen sowie die Wahl einer Vorgangsweise für die Berechnungen im Bericht „Stationäre Erstversorgung von PatientInnen mit Myokardinfarkt bzw. Schlaganfall“.

In einem zweiten Teil sind die in den fachlichen Berichten verwendeten statistischen Verfahren dargestellt.

Des Weiteren werden die SQL und R Codes dokumentiert.

---

<sup>1</sup> Im Prinzip besteht eine ähnliche Problematik bei Wiederaufnahmen. Diese wird in den vorliegenden Analysen jedoch vernachlässigt.

<sup>2</sup> Als **Minimum Basic Data Set (MBDS)** werden jene Daten bezeichnet, die von Krankenanstalten im Zuge der LKF-Abrechnung an das Bundesministerium für Gesundheit (BMG) in anonymisierter Form geliefert werden.

## 2 AUFTRAG

### *Anfrage*

---

Anfrage von: Mag.<sup>a</sup> Nina Pfeffer

Anfrage am: 30. Jänner 2012

---

### *Ersteller*

---

Bericht erstellt von: Mag. Martin Scheuringer

Bericht erstellt im: Juli 2012

---

### *Qualitätssicherung*

---

Qualitätssicherung durch: Mag.<sup>a</sup> Nina Pfeffer, Dr. Ingrid Wilbacher,

Mag. Andreas Golz, Dr. Timo Fischer

Qualitätssicherung im: Juli 2012

---

### *Kontakt für Rückfragen*

Martin Scheuringer, [martin.scheuringer@hvb.sozvers.at](mailto:martin.scheuringer@hvb.sozvers.at), 01/71132-3624

### 3 VERZEICHNIS

1	Abstract.....	2
2	Auftrag .....	3
3	Verzeichnis .....	4
3.1	Bilder .....	4
3.2	Tabellen .....	4
4	Zweck dieses Dokuments .....	5
5	Daten.....	5
5.1	Verwendete Daten .....	5
5.2	Datenherkunft/Glossar.....	6
6	Methoden.....	6
6.1	Bestimmung der burden of disease .....	6
6.2	Statistische Verfahren.....	12
7	Anhang.....	14
7.1	deskriptive Analysen .....	14
7.2	GIS-Karte je ICD .....	18
7.3	Todeswahrscheinlichkeit.....	22
7.4	Sensitivitätsanalyse .....	24

#### 3.1 Bilder

Abbildung 1: Unterschiede der Quoten (Subtraktion) zwischen Modell 1 und 2.....	10
Abbildung 2: Differenz der Quoten Modell 1-2 .....	11

#### 3.2 Tabellen

Tabelle 1: Sensitivität Median – Distanzen .....	9
--	---

## 4 ZWECK DIESES DOKUMENTS

1. Dieses Dokument beschreibt die im Rahmen der Analysen zur Stationären Versorgungsforschung verwendeten Daten hinsichtlich ihrer Herkunft und Beschaffenheit.
2. Es sind Beschreibungen der Methoden für stationäre Versorgungsforschung mit Schwerpunkt Distanzanalysen ausformuliert.
3. Die Vorgangsweise wird anhand von Beispielen erläutert. Diese beziehen sich auf Erkrankungen, bei denen davon ausgegangen werden kann, dass eine zeitnahe Versorgung durch eine geeignete Versorgungsstruktur die negativen Folgen der Erkrankung für die/den PatientIn minimiert.

## 5 DATEN

### 5.1 Verwendete Daten

Für dieses Projekt wurde eine relationale Datenbank errichtet, die mit den MBDS-Einzeldatensätzen bestückt wurde, mit regionalen Gliederungsinformationen, sowie mit Straßenentfernungsdistanzen zwischen jeder PLZ und jedem Krankenhaus – diese wurden für dieses Projekt extra ausgemessen.<sup>3</sup> Mit Hilfe dieser Datenbank kann jedem stationären Aufenthalt eine Straßen-Fahrdistanz<sup>4</sup> zu seinem Wohnort zugeordnet werden. Diese Distanzen können nach unterschiedlichen Fragestellungen gruppiert werden und so Aufschluss über die Versorgungslandschaft Österreichs geben.

Für die Gruppierung und Sortierung der Daten waren zahlreiche SQL-Abfragen und für die rechnerische Auswertung waren Programme in der statistischen Programmiersprache R notwendig. Die GIS-Karten wurden in R gezeichnet.

Ausländische Gastaufenthalte (Personen ohne Wohnsitz in Österreich) wurden aus der Analyse ausgeschlossen.

Das entscheidende Kriterium für die Zuordnung einer Leistung/Diagnose zu einer Region ist je nach Betrachtungsweise der Wohnsitz der/s PatientIn, der bzw. die den stationären Aufenthalt hatte (Quellpopulation) oder die Region der behandelnden Krankenanstalt (Zielpopulation). Für die durchzuführenden Distanzanalysen wird die Quellpopulation betrachtet, Aufenthalte werden also nach dem Wohnsitzprinzip den Regionen zugeordnet.

---

<sup>3</sup> Jeder Region und jeder Krankenanstalt wurden aus der Open-Street-Map Koordinaten zugewiesen. Zur Bestimmung der Distanz wurde die Weg- oder Straßen-Entfernung verwendet. Dazu wurde Routing-Software eingesetzt.

<sup>4</sup> Bei großen Distanzen oder dichtem Verkehr wird auch die Flugrettung eingesetzt. Hier verzerren die Straßendistanzen. Welcher Aufenthalt mit einem Hubschrauber und welcher mit einem Rettungswagen eintrifft, ist in den vorliegenden Daten zu Spitalsaufenthalten nicht dokumentiert.

## 5.2 Datenherkunft/Glossar

Die **MBDS-Daten** sind anonymisierte Daten zu stationären Aufenthalten: Es werden Aufenthalte dokumentiert, aber nicht die Personen, die in Krankenanstalten behandelt werden. Die Daten hält das BMG. In diesen Daten sind Diagnosen und Leistungen aller stationären Aufenthalte in Landesgesundheitsfonds finanzierten Krankenanstalten (LGF) und PRIKRAF-Krankenanstalten dokumentiert. Für die gegenständlichen Auswertungen wurden nur die LGF-Krankenanstalten betrachtet.

**Regionale Informationen** zu Distanzen zwischen Postleitzahlen und deren exakte geographische Lokalisierung durch Bestimmung eines Zentrums wurden durch einen externen Auftragnehmer<sup>5</sup> ermittelt und dem HVB zur Verfügung gestellt.

Die **Arbeitslosenzahlen** je PLZ wurden vom AMS zur Verfügung gestellt.

Die **GAP-DRG-Datenbank** ist eine Datenbank, in der unterschiedliche Datenquellen für die Jahre 2006 und 2007 zusammengeführt wurden: MBDS-Daten, Medikamentendaten und Abrechnungsdaten aus dem niedergelassenen Bereich. Sie enthält pseudonymisierte Daten und ermöglicht daher die Abbildung von Behandlungsketten.

## 6 METHODEN

### 6.1 Bestimmung der burden of disease

Die nach Diagnosen aufgeschlüsselte Krankheitslast im stationären Bereich ist aus den MBDS-Daten leicht zu erkennen – es muss dabei jedoch berücksichtigt werden, dass mit anonymisierten Datensätzen gearbeitet wird und daher keine Aussagen darüber erfolgen können, wie viele Menschen tatsächlich erkrankt sind und Spitäler aufsuchen (d.h. das Verhältnis Aufenthalte zu PatientInnen ist nicht bekannt).

Am Beispiel Herzinfarkt bzw. Schlaganfall bedeutet dies: Wir können sagen wie viele Aufenthalte durch Herzinfarkte/Schlaganfälle verursacht werden, aber wir wissen nicht wie viele PatientInnen mit Herzinfarkt/Schlaganfall hinter diesen Aufenthalten stehen. Außerdem kann eine Diagnose als Haupt- oder als Nebendiagnose kodiert werden. Da Nebendiagnosen im LKF-Modell aber nicht abrechnungsrelevant sind, wird die Qualität der Kodierung in der Regel von Experten als limitiert eingeschätzt.

- Es ist nicht sichergestellt, dass es sich bei der kodierten Diagnose um das primäre Ereignis handelt (es könnte sich auch um eine Wiederaufnahme nach Entlassung handeln – diese sind jedoch schlecht kodiert)

---

<sup>5</sup> Fa. Drahtwarenhandlung Simulation Services GmbH. Dieses Projekt wurde im Jahr 2010 in Kooperation mit der Abteilung 33-EWG durchgeführt.

- Auf Grund der Anonymisierung der Daten können Behandlungsketten nicht verfolgt werden: Viele Aufenthalte enden bzw. beginnen mit Transfers von/nach anderen Krankenanstalten, diese werden jedoch nicht immer kodiert. Bei Verlegungen zwischen Krankenanstalten (z.B. in die richtige Versorgungsstufe) werden daher mehrere Aufenthalte pro Person erzeugt.
- Ein ähnliches Phänomen wie bei den Transfers besteht bei nicht kodierten Wiederaufnahmen. Auch hier erfolgt nach Expertenmeinungen sowohl die Kodierung bei der Aufnahme als auch bei der Entlassung unzuverlässig.<sup>6</sup>

### 6.1.1 Haupt- und Zusatzdiagnosen

Herzinfarkt wird vermutlich dann als Zusatzdiagnose verwendet, wenn

- es ein status post Herzinfarkt ist,
- eine andere Diagnose akut aufgetreten ist (während oder vor dem Herzinfarkt) und mehr LKF-Punkte bringt (vgl. Baumsystematik im LKF-Modell).

So etwas ist möglich, wenn

- ein/e PatientIn für eine OP aufgenommen wird und den Infarkt im Spital erleidet oder bereits irgendwann früher einmal einen Herzinfarkt erlitten hatte,
- während des Aufenthaltes, der durch den Infarkt ausgelöst wurde eine Komplikation mit „mehr Punkten“ = höherer Priorität auftritt.<sup>7</sup>

Im Jahr 2009 wurden 6455 Herzinfarkte und 6096 Schlaganfälle als Zusatzdiagnose kodiert. Diese werden aus der Analyse exkludiert, weil es darum geht, nur akute Herz- bzw. Hirninfarkte zu beobachten.

---

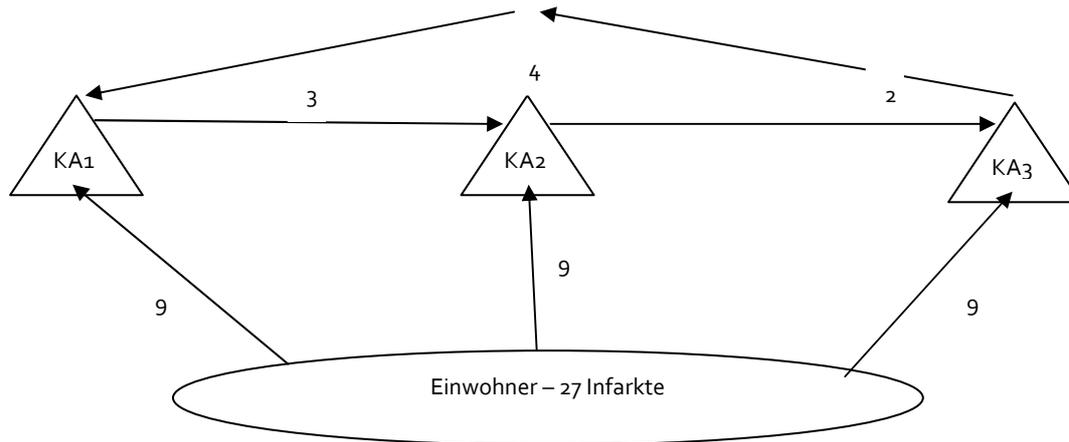
<sup>6</sup> Dabei handelt es sich um eine im LKF-Arbeitskreis und im LKF-Projektmanagement oftmals geäußerte Einschätzung der LKF-Experten des BMG.

<sup>7</sup> Auskunft Dr. Gottfried Endel (Hauptverband), per Mail am 16.5.2012

## 6.1.2 Kodierproblematik der Transfers – methodische Überlegungen

Die Entlassungen und Aufnahmen von Transfers im MBDS-Datensatz sollten dokumentiert sein. Wie bereits weiter oben ausgeführt ist leider auf Grund von Ungereimtheiten **nicht** davon auszugehen, dass dies in erforderlicher Qualität der Fall ist. Ein Myokardinfarkt kann zu mehr Aufenthalten führen, und wird daher potentiell mehrfach als Erkrankung gezählt, wenn die Transfers und die Wiederaufnahmen nicht ordentlich kodiert werden. **Dies verzerrt regionale Häufigkeiten von Krankheiten und durchschnittliche Distanzmaße.**

Zur Verdeutlichung eine schematische Darstellung:



Erläuterung der Darstellung:

In einer Region erleiden 27 Menschen einen Myokardinfarkt. Je 9 werden in KA1, KA2 und KA3 behandelt. Jedoch werden 3 nach einer Behandlung noch in KA2 verlegt, und 2 werden von KA2 nach KA3 verlegt, sowie 4 von KA3 nach KA1. So erzeugen 27 Myokardinfarkte  $27+9=36$  Aufenthalte.

Dividiert man durch die Einwohner scheint hier eine regionale Häufung von Myokardinfarkten statt zu finden, die aber durch die vielen Transfers hervorgerufen wurde. Da die Distanz pro Aufenthalt berechnet wird erhält man zusätzlich mehr Distanzen zur Bildung des Medians, wodurch eine Verzerrung nach oben bei der Distanz-Analyse der Erstversorgung entsteht.

Aus den MBDS-Daten ist aber nur die Zahl 36 Aufenthalte zu erkennen sowie jener Teil der Transfers, der auch explizit kodiert wurde. Das führt zu den Ungereimtheiten:

- Variante 1: Man nimmt alle PatientInnen weg, die als Transfer aufgenommen werden: Leider weiß ich nicht ob der Myokardinfarkt Grund des Transfers war.
- Variante 2: Man nimmt alle PatientInnen weg, die als Transfer entlassen werden; auch hier kann wegen der Schwere des Myokardinfarkts oder wegen Nachbehandlung transferiert werden.

Es ist wichtig, bei der Auswahl der betrachteten Population Einschränkungen zu treffen, die eine möglichst realitätsnahe Abbildung der medizinischen Versorgung darstellt. Aufgrund der Kodierproblematik gibt es keine perfekte Lösung – jede der o.a. Varianten bringt Vor- und Nachteile.

Um eine möglichst objektive Entscheidungsgrundlage für die Auswahl einer Variante zu erhalten, werden Sensitivitätsanalysen durchgeführt. Das bedeutet, dass untersucht wird, wie sich die regionalen Quoten (Aufenthalte pro PatientInnen) und Distanzen (Wohnsitz – Krankenanstalt) je nach Modellvariante ändern. Haben die unterschiedlichen Populationen einen Einfluss auf das gesuchte Ergebnis?

- Modellannahme 1: Alle Aufenthalte mit der Diagnose Myokardinfarkt werden untersucht.
- Modellannahme 2: Aus der ersten Annahme werden alle Aufenthalte ausgeschlossen, die als Transfer aufgenommen wurden. (=Variante 1)
- Modellannahme 3: Es werden Aufenthalte ausgeschlossen, die als Transfer gekommen sind und die als Transfer entlassen wurden. (=Variante 2)

### 6.1.2.1 Vergleich der Distanzen

Die 3 Modellannahmen führen zu folgenden Ergebnissen in der Distanzanalyse:

Modell 1	
distanz_median	15.9 km
distanz_median_rural	22.1 km
distanz_median_urban	6.5 km
Modell 2	
distanz_median	13.9 km
distanz_median_rural	20.5 km
distanz_median_urban	6.0 km
Modell 3	
distanz_median	13.9 km
distanz_median_rural	20.5 km
distanz_median_urban	5.6 km

Tabelle 1: Sensitivität Median – Distanzen

Zwischen dem Modell 2, das nur *Transferaufnahmen* und Modell 3, das sowohl *Transferaufnahmen* und *Transferentlassungen* ausschließt, besteht kaum ein Unterschied. Die durchschnittliche Distanz verkürzt sich bei beiden Modellen jedoch im Vergleich zu Modell 1, bei dem keine Einschränkung der Aufenthalte erfolgte (im ländlichen Bereich um ca. 1,5 km bzw. im städtischen Raum um ca. 0,5 km Unterschied).

**Da es um zeitkritische Notfall-Versorgung geht legt dieser eher geringe Unterschied von 1,5 km bzw. 0,5 km nahe, Modell 2 zu nehmen. Bei diesem sind die durchschnittlichen Distanzen kürzer. Dies erscheint unter der Prämisse einer wohnortnahen Erst- und einer von dieser weiter entfernten Weiterbehandlung (z.B. in höherwertigen Versorgungsstufen) als plausibel. Modell 3 wird ausgeschlossen, da hier kaum Unterschiede in den Distanzen zu Modell 2 erkennbar sind und weil hier die Gefahr besteht, viele relevante Erstaufnahmen aus der Population auszuschließen, da Aufenthalte, die als Transfer entlassen werden ausgeschlossen, werden.**

### 6.1.2.2 Vergleich der Quoten je PLZ

Im Folgenden wird zwischen Modell 1 und 2 die regionale Variabilität verglichen: Lässt man alle Aufenthalte, die als Transfers aufgenommen wurden weg, so ergeben sich in einigen Regionen durchaus stärkere Unterschiede. Anders formuliert: Durch das Weglassen der Transferaufnahmen

verändern sich die Verteilungen der Quoten je PLZ: Quoten über 5 Aufenthalten je 1000 EW pro PLZ werden deutlich geringer. Das heißt, einige PLZ mit überdurchschnittlich hohen Quoten fallen nun nicht mehr so sehr ins Auge.

Das bedeutet: in bestimmten Bezirken wird vermutlich zuerst das nahe Spital angefahren und dann wird ins weiter entfernte transferiert.

### 6.1.2.3 Wie groß ist der absolute Unterschied zwischen den beiden Modellen auf Postleitzahlenebene?

In den folgenden Darstellungen wird der Unterschied der Quoten (Aufenthalte je 1000 EW) zwischen Modell 1 und 2 dargestellt: In den meisten PLZ passiert keine Veränderung der Quote, wenn man die beiden Quoten voneinander subtrahiert; das betrifft ca. 1000 PLZ.

In den restlichen (dargestellt in Abbildung 1: Unterschiede der Quoten (Subtraktion) zwischen Modell 1 und 2) kommt es zu Verschiebungen in der Quote von bis zu + 0,008, meist aber weniger als + 0,002 Aufenthalte pro Einwohner. Das bedeutet am Beispiel Herzinfarkt: in ca. 25 PLZ geschehen in etwa 2 Aufenthalte pro 1000 Einwohner weniger, wenn die Transferaufnahmen ausgeblendet werden.

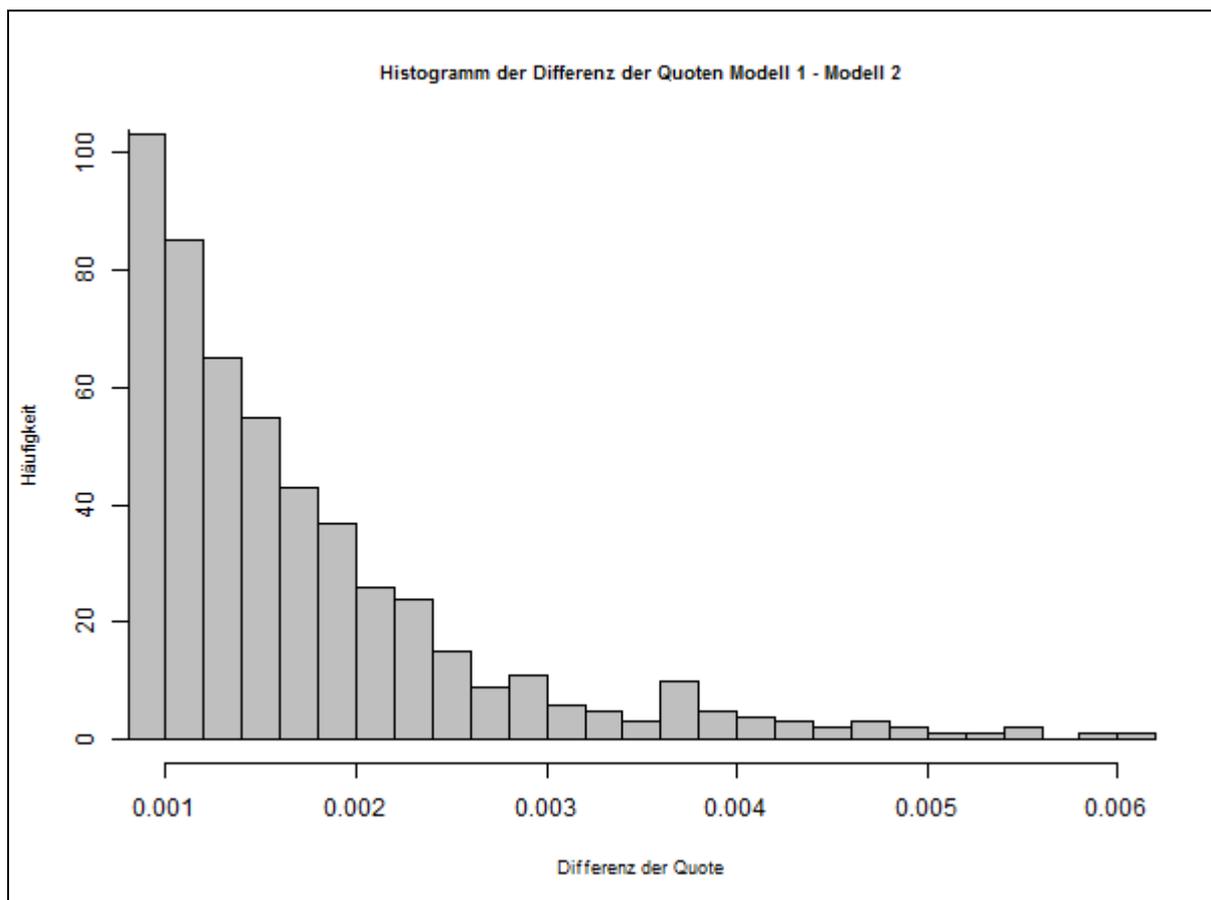


Abbildung 1: Unterschiede der Quoten (Subtraktion) zwischen Modell 1 und 2

In der GIS-Darstellung (Abbildung 2: Differenz der Quoten Modell 1-2) ist die regionale Verteilung dieser Differenzen der Quoten zu erkennen: In Bezirken mit großen Kreisen sinkt die Quote der Herzinfarkte am stärksten, wenn man alle Transferaufnahmen ausschließt.

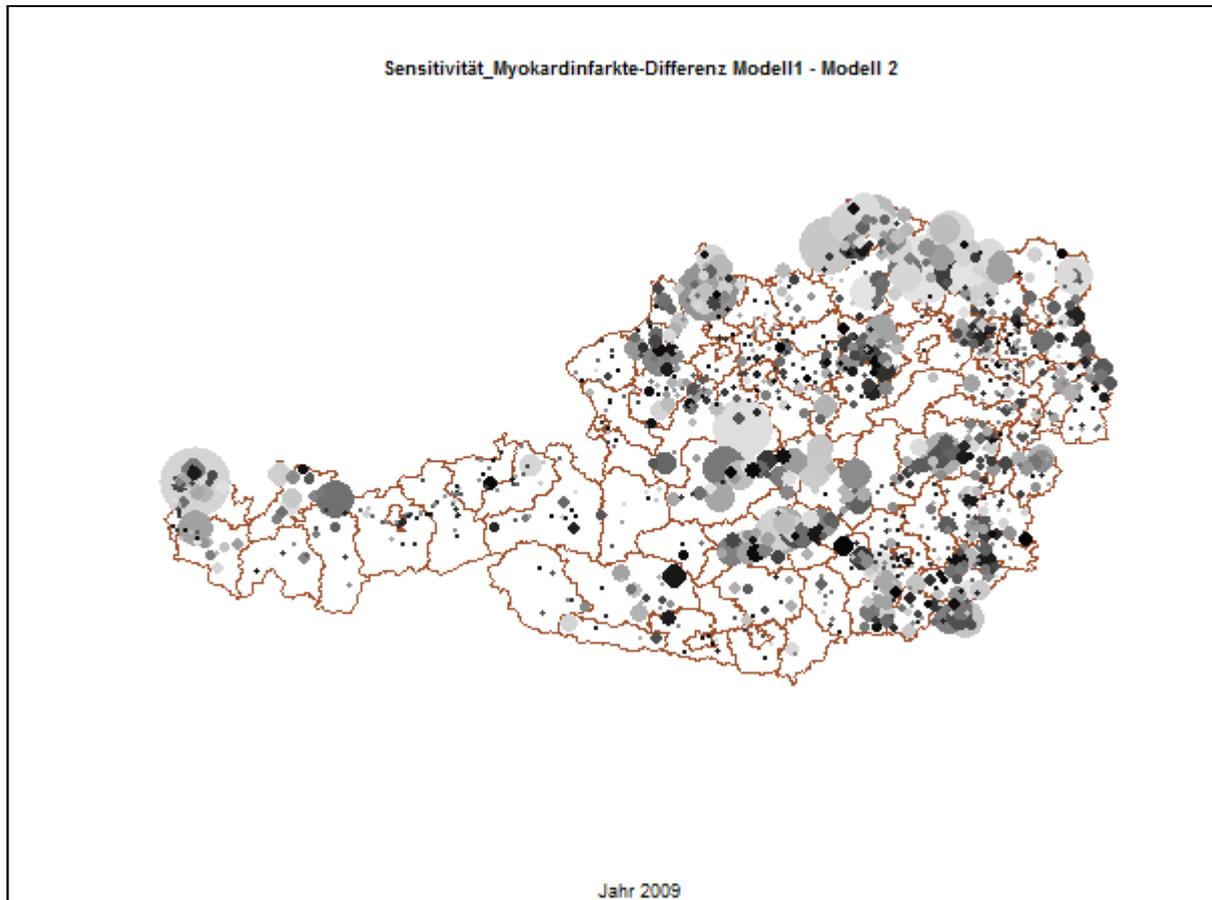


Abbildung 2: Differenz der Quoten Modell 1-2

### 6.1.3 Kontrolle von Modell 2 mit personenbezogenen Daten aus GAP-DRG

Nachdem die Sensitivitätsanalyse zur Auswahl von Modell 2 geführt hat, kann aus der GAP-DRG-Datenbank anhand pseudonymisierter Daten gezeigt werden, ob mehrere Aufenthalte aus den MBDS-Daten einer Person zuzuordnen sind.<sup>8</sup> Es kann so ein Faktor angegeben werden, der beschreibt wie viele Aufenthalte durch eine/n Myokardinfarkt-PatientIn im Schnitt verursacht werden. Die Zahlen beziehen sich auf 2007.

In der GAP-DRG finden sich 17.511 Aufenthalte von 16.101 Personen – inklusive Ausländer. Das ergibt einen Faktor von 1,095. Das bedeutet, dass eine Person im Jahr 2007 im Schnitt **1,095** Aufenthalte mit Myokardinfarkt als Hauptdiagnose hatte. Im Gegensatz dazu führt die Modellwahl zu folgendem Faktor: 14.744 Aufenthalte (Modell 1) und 11.942 Aufenthalte ohne Transferaufnahmen (Modell 2) – ohne Aufenthalte von ausländischen GastPatientInnen – ergibt einen Faktor von **1,23**.

<sup>8</sup> Die MBDS-Daten wurden im Rahmen eines Forschungsprojekts durch ein Record Linkage Verfahren in den Datenbestand der GAP-DRG integriert und den pseudonymisierten Personen zugeordnet.

Da auf Grund von Datenbankunterschieden von zwei unterschiedlichen Populationen in zwei unterschiedlichen Jahren auszugehen ist, hat der Unterschied in den Faktoren keine allzu große Bedeutung. Ein Vergleich zwischen den pseudonymisierten GAP-DRG-Daten und den anonymisierten MBDS-Daten kann unter diesen Umständen nicht zur Validierung der Modellwahl beitragen.

#### 6.1.4 Modellauswahl

Für die Durchführung der Analysen wird Modell 2 empfohlen, das heißt, dass alle *Transferaufnahmen* aus der Untersuchungspopulation ausgeschlossen werden, um die Verfälschungen zu minimieren und die Erstversorgung präzise darzustellen.

## 6.2 Statistische Verfahren

Es kommen verschiedene statistische Verfahren zur Anwendung; diese werden in den folgenden Abschnitten kurz erläutert.

### 6.2.1 Maße der Zentralen Tendenz

Der *Median* teilt eine Reihe von Daten in zwei gleich große Teile. Die Hälfte der Beobachtungen ist über, die andere unter dem Wert. Wenn man Daten in gleich große Teile teilt, so erhält man Quantile. Zb. 4 Teile: 0 - 25%, 25% - 50%, 50% - 75%, 75% - 100% der Fälle.

Der *arithmetische Mittelwert* wird nach einer Formel berechnet:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

### 6.2.2 Verfahren für Zusammenhänge von Daten

#### 6.2.2.1 Regression

In einer linearen Regressionsgleichung wird untersucht wie stark der Einfluss bestimmter Variablen auf eine Variable ist.

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_nx_{in} + e_i$$

#### Beispiel:

Hängt die Menge der Herzinfarkte/ Schlaganfälle einer Postleitzahl von bestimmten Gegebenheiten dieser Postleitzahl ab? Ich kann vermuten, dass das durchschnittliche Alter, der Anteil der Männer und die Arbeitslosigkeit einen Einfluss auf die Menge der Herzinfarkte/ Schlaganfälle ausüben. Mit Hilfe des Regressionsverfahrens kann ich nun prüfen ob dieser Einfluss auch wirklich besteht und wie stark er ist.

In einer binären LOGIT-Regressionsfunktion werden die Einflussfaktoren für das Fallen eines Individuums in eine bestimmte Gruppe geprüft.

$$p(y = 1) = \frac{1}{1 + e^z} \text{ mit } z = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_nx_{in}$$

**Beispiel:**

Die Wahrscheinlichkeit an einem Myokardinfarkt zu sterben hängt von verschiedenen Variablen ab: Ihr empirischer Einfluss wird gemessen, wobei der Einfluss anderer die Todeswahrscheinlichkeit bestimmender Faktoren kontrolliert werden kann.

## 7 ANHANG

Dieser Anhang stellt die in R programmierten Abfragen dar, mit denen die Berechnungen durchgeführt wurden und die Karten gezeichnet wurden. Dabei wurde immer auf eine eigens dafür errichtete Datenbank, die in die Datenbank der EWG eingegliedert ist, zugegriffen.

### 7.1 deskriptive Analysen

```
##### stationäre Versorgungsforschung
##### Martin Scheuringer
```

```
### Deskriptive Analysen
```

```
## libraries und packages
rm(list = ls())
library(RODBC) # SQL-Datenbank-Connenction
require(stats) # für mehr Farben bei den Diagrammen
```

```
# Zugang zur Datenbank und Abfrage
conn <- odbcConnect("geo")
```

```
##### Einzugsbereich allgemein
# Daten laden
dist_ka <- sqlQuery(conn, "SELECT t1.id, t1.ka, t1.plz, substr(t1.icd_sub,1,3), t1.road_distance FROM
full_2 t1
      INNER JOIN diag_sa1_2009 t2
      USING(id)
      WHERE t2.auf_art <> 'T'")
```

```
dist_ka_rural <- sqlQuery(conn, "SELECT t1.id, t1.ka, t1.plz, substr(t1.icd_sub,1,3), t1.road_distance
FROM full_2 t1
      INNER JOIN diag_sa1_2009 t2
      USING(id)
      WHERE t2.auf_art <> 'T'
      AND t1.plz NOT IN (SELECT unnest(postcode) FROM villages_all
      WHERE province_name IN ('Wien')
      OR district_name IN ('Graz Stadt', 'Linz Stadt', 'Salzburg Stadt',
      'Innsbruck Stadt', 'Klagenfurt Stadt', 'Villach Stadt', 'St. Pölten', 'Dornbirn',
      'Wiener Neustadt', 'Feldkriech', 'Wels Stadt', 'Steyr Stadt',
      'Bregenz')
      OR municipal_name IN
('Klosterneuburg', 'Leonding', 'Baden', 'Wolfsberg', 'Leoben', 'Krems', 'Traun', 'Amstetten', 'Kapfenberg',
Lustenau', 'Mödling')
      )")
```

```
dist_ka_urban <- sqlQuery(conn, "SELECT t1.id, t1.ka, t1.plz, substr(t1.icd_sub,1,3), t1.road_distance
FROM full_2 t1
  INNER JOIN diag_sa1_2009 t2
  USING(id)
  WHERE t2.auf_art <> 'T'
  AND t1.plz IN (SELECT unnest(postcode) FROM villages_all
  WHERE province_name IN ('Wien')
  OR district_name IN ('Graz Stadt','Linz Stadt','Salzburg Stadt',
  'Innsbruck Stadt','Klagenfurt Stadt','Villach Stadt','St.Pölten','Dornbirn',
  'Wiener Neustadt','Feldkriech','Wels Stadt','Steyr Stadt',
  'Bregenz')
  OR municipal_name IN
('Klosterneuburg','Leonding','Baden','Wolfsberg','Leoben','Krems','Traun','Amstetten','Kapfenberg','
Lustenau','Mödling')
)")
```

```
# Arbeitsverzeichnis für Diagramm
```

```
setwd("J://Ref36/Referat/36.2 - Krankenanstalten/2 (9.21) - WEISSES BUCH/2 (9.21.11) - Weisses
Buch 2011/Projekte/stationäre Versorgungsforschung/6_Diagramme")
```

```
#####
```

```
# Maße berechnen
```

```
# generell
```

```
#####
```

```
# Durchschnittliche Distanzen ausmessen
```

```
distanz_mean <- mean(dist_ka$road_distance)
distanz_median <- median(dist_ka$road_distance)
distanz_sd <- sd(dist_ka$road_distance)
```

```
# rural
```

```
distanz_mean_rural <- mean(dist_ka_rural$road_distance)
distanz_median_rural <- median(dist_ka_rural$road_distance)
distanz_sd_rural <- sd(dist_ka_rural$road_distance)
```

```
#urban
```

```
distanz_mean_urban <- mean(dist_ka_urban$road_distance)
distanz_median_urban <- median(dist_ka_urban$road_distance)
distanz_sd_urban <- sd(dist_ka_urban$road_distance)
```

```
#####
```

```

#Quantile
qu <- quantile(dist_ka$road_distance, probs = seq(0, 1, 0.05))
qu_r <- quantile(dist_ka_rural$road_distance, probs = seq(0, 1, 0.05))
qu_u <- quantile(dist_ka_urban$road_distance, probs = seq(0, 1, 0.05))

#####
#Diagramme

# 150 km
win.metafile(file = "Distanzen_generell_2009_150km_t.wmf", width = 7, height = 4)
op <- par(bg = "white", fg = "black")
hist(dist_ka$road_distance,
      main = "Verteilung der Auto-Distanzen in Österreich 2009",
      cex.main = .75,
      axes = FALSE,
      col = "gray75",
      xlab = "Entfernung",
      cex.lab = .75,
      ylab = "Häufigkeit",
      xlim = c(0,140000),
      breaks = 500)
axis(2, at = seq(0,300000,50000), tcl = 0.1, labels = c("0","50k","100k","150k","200k","250k","300k"),
     cex.axis = .66, las = 1, lwd = 1)
axis(1, at = seq(0,140000,10000), tcl = 0.1, labels = c("0 km","10 km","20 km","30 km","40 km","50
km","60 km","70 km","80 km","90 km","100 km","110 km","120 km","130 km","140 km"), cex.axis =
.66, las = 2, lwd = 1)
lines(x = c(distanz_median,distanz_median), y = c(0,290000), lty = 2)
text(x = distanz_median, y = 310000, "Median", cex = .5)
rect(xleft = qu[1],ybottom = 0, xright = qu[length(qu)-1], ytop = 50000)
text(x = qu[length(qu)-1], y = 51000, paste("95% aller Aufenthalte bei ", round((qu[length(qu)-
1])/1000, digits = 1), sep = "")), cex = .5, pos = 4)
rect(xleft = qu[1],ybottom = 0, xright = qu[length(qu)-5], ytop = 100000)
text(x = qu[length(qu)-5], y = 110000, paste("75% aller Aufenthalte bei ", round((qu[length(qu)-
5])/1000, digits = 1), sep = "")), cex = .5, pos = 4)
par(op)
dev.off()

length(dist_ka$road_distance[dist_ka$road_distance > 45000])

# 150 km rural
win.metafile(file = "Distanzen_generell_2009_150km_rural_urban_t.wmf", width = 7, height = 8)
#png(file="Distanzen_generell_2009_150km_rural_urban.png", height=600, width=600)
op <- par(bg="white", fg="black", mfrow = c(2,1))
hist(dist_ka_rural$road_distance,
      main = "Vergleich der Distanzverteilungen 2009",
      cex.main = .75,

```

```

axes = FALSE,
sub = "ohne Städte mit mehr als 20.000 Einwohnern",
font.sub = 3,
col = "gray75",
xlab = "Entfernung",
ylab = "Häufigkeit",
cex.lab = .75,
xlim = c(0,140000),
breaks = 500)
axis(2, at = seq(0,100000,50000), tcl = 0.1, labels = c("0", "50k", "100k"), cex.axis = .66, las = 1, lwd = 1)
axis(1, at = seq(0,140000,10000), tcl = 0.1, labels = c("0 km", "10 km", "20 km", "30 km", "40 km", "50
km", "60 km", "70 km", "80 km", "90 km", "100 km", "110 km", "120 km", "130 km", "140 km"), cex.axis =
.66, las = 2, lwd = 1)
lines(x = c(distanz_median_rural, distanz_median_rural), y = c(0,100000), lty = 2)
text(x = distanz_median_rural, y = 105000, "Median", cex = .5)
rect(xleft = qu_r[1], ybottom = 0, xright = qu_r[length(qu_r)-1], ytop = 25000)
text(x = qu_r[length(qu_r)-1], y = 26000, paste("95% aller Aufenthalte bei ",
round((qu_r[length(qu_r)-1])/1000, digits = 1), sep = ""), cex = .5, pos = 4)
rect(xleft = qu_r[1], ybottom = 0, xright = qu_r[length(qu_r)-5], ytop = 50000)
text(x = qu_r[length(qu_r)-5], y = 51000, paste("75% aller Aufenthalte bei ",
round((qu_r[length(qu_r)-5])/1000, digits = 1), sep = ""), cex = .5, pos = 4)

# 150 km urban
hist(dist_ka_urban$road_distance,
main = NULL,
axes = FALSE,
sub = "nur Städte mit mehr als 20.000 Einwohnern",
font.sub = 3,
col = "gray75",
xlab = "Entfernung",
ylab = "Häufigkeit",
xlim = c(0,110000),
cex.lab = .75,
breaks = 500)
axis(2, at = seq(0,200000,50000), tcl = 0.1, labels = c("0", "50k", "100k", "150k", "200k"), cex.axis = .66,
las = 1, lwd = 1)
axis(1, at = seq(0,140000,10000), tcl = 0.1, labels = c("0 km", "10 km", "20 km", "30 km", "40 km", "50
km", "60 km", "70 km", "80 km", "90 km", "100 km", "110 km", "120 km", "130 km", "140 km"), cex.axis =
.66, las = 2, lwd = 1)
lines(x = c(distanz_median_urban, distanz_median_urban), y = c(0,200000), lty = 2)
text(x = distanz_median_urban, y = 205000, "Median", cex = .5)
rect(xleft = qu_u[1], ybottom = 0, xright = qu_u[length(qu_u)-1], ytop = 25000)
text(x = qu_u[length(qu_u)-1], y = 26000, paste("95% aller Aufenthalte bei ",
round((qu_u[length(qu_u)-1])/1000, digits = 1), sep = ""), cex = .5, pos = 4)
rect(xleft = qu_u[1], ybottom = 0, xright = qu_u[length(qu_u)-5], ytop = 50000)
text(x = qu_u[length(qu_u)-5], y = 51000, paste("75% aller Aufenthalte bei ",
round((qu_u[length(qu_u)-5])/1000, digits = 1), sep = ""), cex = .5, pos = 4)

par(op)
dev.off()

```

## 7.2 GIS-Karte je ICD

```
#####  
# Einzugsbereiche bei Operationen  
# Martin Scheuringer  
  
#####  
#####  
#1. Vorbereitungen  
  
# Hauptspeicher leeren  
rm(list = ls())  
  
# Packages laden  
library(raster) # GIS  
library(sp) # GIS  
library(lattice) # Graphik  
library(RODBC) # Datenbankzugang  
library(classInt) # Benötigt für die Gruppierung der Werte  
require(graphics)  
  
# Daten fñg Österreicher Grenzen laden - 2 bedeutet mit Bezirksgrenzen --> siehe  
http://www.gadm.org/country  
# Dann laden  
  
setwd("C:/Users/Scheuringer/GIS")  
load(file = "AUT_adm2.RData")  
  
# Zugang zur Datenbank und Abfrage  
conn <- odbcConnect("geo")  
  
# Daten für Punkte in GIS-Karte laden  
plz <- sqlQuery(conn,"SELECT plz, ew, longitude, latitude FROM ew_2009_plz") # Für PLZ-  
Koordinaten  
ka <- sqlQuery(conn,"SELECT hospid, beds, longitude, latitude FROM hospitals") #  
Krankenanstandenstandorte  
  
# Leistungen definieren  
sql_leist <- "('I63','I64')"  
  
# Speicherpfade definieren - müssen angelegt sein!  
speicherpfad_grafik <- "J://Ref36/Referat/36.2 - Krankenanstanden/2 (9.21) - WEISSES BUCH/2  
(9.21.11) - Weisses Buch 2011/Projekte/stationäre  
Versorgungsforschung/6_Diagramme/GIS_Karten"  
speicherpfad_tabs <- "J://Ref36/Referat/36.2 - Krankenanstanden/2 (9.21) - WEISSES BUCH/2  
(9.21.11) - Weisses Buch 2011/Projekte/stationäre Versorgungsforschung/4_R/tabs_GIS_Produktion"  
# Name der Grafiken definieren "Leistungsname"  
name <- "Hirninfirmkte - ICD I63,I64"
```

```
# Strukturen aus DIAG holen
#! Achtung ohne Österreichsummen laden!
#struktur <- "vzae_myo.csv"
#diag_pfad_vzae <- "J://Ref36/Referat/36.2 - Krankenanstalten/2 (9.21) - WEISSES BUCH/2 (9.21.11)
- Weisses Buch 2011/Projekte/stationäre Versorgungsforschung/1_Daten/Struktur_Diag"
# In Zeile 315 die Intervalle für die Brüche definieren!
```

```
#####
#####
```

```
# 2. Berechnung der Quote je plz
```

```
# absolute Häufigkeit der Eingriffe je plz
```

```
sql_statment_plz_a <- paste("
SELECT t2.plz, COUNT(substring(t1.icd_sub from 1 for 3)) AS anz
FROM diag_sa2_2009 t1
INNER JOIN diag_sa1_2009 t2 ON t1.id = t2.id
WHERE t2.auf_art <> 'T'
AND t1.diagn_typ = 'H'
AND substring(t1.icd_sub from 1 for 3) IN", sql_leist, sep = "")
```

```
sql_statment_plz <- paste(sql_statment_plz_a,"AND t2.plz IN (SELECT plz FROM ew_2009_plz)
GROUP BY t2.plz", sep = "")
```

```
abs_eingr <- sqlQuery(conn, sql_statment_plz)
```

```
# merge: zur Anzahl der Eingriffe die Geo-Daten geben
eingr_pro_ew <- merge(plz,abs_eingr, by = 'plz', all.x = TRUE)
quote <- eingr_pro_ew$anz/eingr_pro_ew$ew
ew <- eingr_pro_ew$ew
```

```
# Festlegung der Grenzen der Brüche
brueche <- classIntervals(ew, n=90, style="quantile") #90 Intervalle für 90 Graustufen
```

```
# Hinzufügen der Größenkategorie zur plz: Nimm Quote und füge Kategorie der Quote hinzu
ew <- cbind(ew, "kategorie" = cut(ew, breaks = c(0, brueche$brks[2:length(brueche$brks)]), labels =
FALSE))
```

```
# in einen dataframe speichern
eg_plz <- data.frame(eingr_pro_ew,ew,quote)
```

```
#####
#####
```

```
#2. Eingriffe in Krankenanstalten
```

```
sql_statment_ka_a <- paste("SELECT t2.ka, COUNT(substring(icd_sub from 1 for 3))
AS anz FROM diag_sa2_2009 t1 INNER JOIN diag_sa1_2009 t2 ON t1.id = t2.id
WHERE substring(icd_sub from 1 for 3) IN", sql_leist, sep = "")
```

```
sql_statment_ka <- paste(sql_statment_ka_a,"AND t2.plz in (SELECT plz FROM ew_2009_plz)
```

```

AND t2.auf_art <> 'T' GROUP BY t2.ka", sep = "")

abs_eingr_ka <- sqlQuery(conn,sql_statement_ka)

#merge: zur Anzahl die Geo-Daten
eingr_pro_ka <- merge(ka,abs_eingr_ka, by.x = 'hospid', by.y = 'ka', all.x = TRUE)
quote_ka <- eingr_pro_ka$anz/eingr_pro_ka$beds
beds <- ka$beds

# Festlegung der Grenzen der Brüche
brueche_ka <- classIntervals(beds, n=4, style="quantile")

# Hinzufügen der Größenkategorie zur plz

beds <- cbind(beds,"kategorie" = cut(beds,breaks = c(0,
brueche_ka$brks[2:length(brueche_ka$brks)]),labels = FALSE))

#Dataframe speichern
eg_ka <- data.frame(eingr_pro_ka, quote_ka, beds)

#####
#####
# Daten sortieren und Farben definieren

ord1 <- order(eg_plz$quote, decreasing = TRUE)
eg_plz_plot <- eg_plz[ord1, ]

#ord2 <-order(eg_ka$quote_ka, decreasing = TRUE)
#eg_ka_plot <- eg_ka[ord2, ]

# Grauwerte definieren

grauwert <- eg_plz_plot$kategorie*(-1)+90
graustufe <- paste("grey",grauwert,sep="")

#####
# regionale Variabilität Output

# Grafikdatei
setwd(speicherpfad_grafik)
png(filename = paste(name,"quoten_je_plz_2009.png", sep =""), width = 1200, height = 900)
#pdf(file = paste(name,"quoten_je_plz_2009.pdf", sep =""), width = 11.96, height = 8.27)
plot(gadm, bg = "white", border = "sienna")
title(main = paste(name,"-Aufenthalte je PLZ",sep=""), sub = "Jahr 2009", cex.main = 1.6, cex.sub =
1.4)
points(x = eg_plz_plot$longitude, y = eg_plz_plot$latitude, pch = 16, col = graustufe, bg = FALSE,
cex = eg_plz_plot$quote*600, type = "p")

```

```

dev.off()
#####
# Abfrage für die linien

# ich brauche von jeder Ka jede linie zur plz, wobei ich summe je plz
# das ist: anzahl der Aufnahmen je ka je plz

kanr <- levels(abs_eingr_ka$ka)
teilo <- paste("SELECT t2.plz, COUNT(substring(icd_sub from 1 for 3)) AS anz
FROM diag_sa2_2009 t1
INNER JOIN diag_sa1_2009 t2 ON t1.id = t2.id
WHERE t2.ent_art <> 'T'
AND substring(icd_sub from 1 for 3) IN", sql_leist, sep="")
teil1 <- paste(teilo, "AND t2.ka = ", sep="")
teil2 <- " GROUP BY t2.plz"

setwd(speicherpfad_tabs)

for (i in 1:length(kanr)) {          # Schleife erstellt für jede KA eine Datei mit 2 Spalten plz, anz
statement1.i <- paste(teil1,kanr[i], sep="")
statement.i <- paste(statement1.i,teil2, sep="")
ka_plz.i <- sqlQuery(conn, statement.i)
filename.i <- paste(kanr[i],"plz.csv", sep="")
write.table(ka_plz.i,filename.i, sep=";", dec=".", row.names=FALSE)
}

##### Lininen

# x-Vektor und y-Vektor der Koordinaten

#x <- jede ungerade Zahl muss die longitude der KA sein
# die geraden Zahlen sind die plz longituden
#y <- jede ungerade Zahl muss die latitude der KA sein
#png(filename = "katarakt2009_einzug.png", width = 1200, height = 900)

setwd(speicherpfad_grafik)
#pdf(file = paste(name,"_einzug_2009.pdf", sep =""), width = 11.96, height = 8.27)
png(filename = paste(name,"_einzug_2009.png", sep =""), width = 1200, height = 900)
plot(gadm, bg = "white", border = "sienna")
title(main = paste(name,"- Aufenthalte - Einzugsgebiete der Krankenanstalten"), sub = "Jahr 2009",
cex.main = 1.6, cex.sub = 1.4)

setwd(speicherpfad_tabs)
for (i in 1:length(kanr)) {
filename.i <- paste(kanr[i],"plz.csv", sep="")
ka_werte.i <- read.csv(filename.i, sep=";", dec=".",)
# alle kleiner x raus!

```

```

ka_werte.i <- ka_werte.i[ka_werte.i$anz > 1,]
ka_werte_koord.i <- merge(ka_werte.i, plz, by = "plz")
# einen x und y Vektor für die Koordianten der Linien
ka_long.i <- ka[(as.character(levels(ka$hospid))[ka$hospid]) == (kanr[i]),3]
ka_lat.i <- ka[(as.character(levels(ka$hospid))[ka$hospid]) == (kanr[i]),4]

color <- rainbow(length(kanr))
for (j in 1:length(ka_werte_koord.i[,1])) {
  x_vec.j <- c(ka_long.i,ka_werte_koord.i[,4])
  y_vec.j <- c(ka_lat.i,ka_werte_koord.i[,5])
  lines(x_vec.j,y_vec.j, lwd = 1, col = color[i])
}
}
points(x = eingr_pro_ka$longitude, y = eingr_pro_ka$latitude, pch = 16, col = 1)
#points(x = eg_plz_plots$longitude, y = eg_plz_plots$latitude, pch = 16, col = 1, bg = FALSE, cex = .01,
type = "p")
# auch wenn o schas

dev.off()

```

## 7.3 Todeswahrscheinlichkeit

```

rm(list = ls())
library(RODBC) # SQL-Datenbank-Connenction
require(stats) # für mehr Farben bei den Diagrammen
require(vcd)
require(rmeta)
require(fmsb)
#####
## Funktion McGill
#####
logistic.regression.or.ci <- function(regress.out, level=0.95)
{
#####
##
##
# This function takes the output from a glm #
# (logistic model) command in R and provides not #
# only the usual output from the summary command, but #
# adds confidence intervals for all coefficients and OR's. #
##
# This version accommodates multiple regression parameters #
##
#####
##
usual.output <- summary(regress.out)
z.quantile <- qnorm(1-(1-level)/2)
number.vars <- length(regress.out$coefficients)
OR <- exp(regress.out$coefficients[-1])
temp.store.result <- matrix(rep(NA, number.vars*2), nrow=number.vars)

```

```
for(i in 1:number.vars)
{
temp.store.result[i,] <- summary(regress.out)$coefficients[i] +
c(-1, 1) * z.quantile * summary(regress.out)$coefficients[i+number.vars]
}
intercept.ci <- temp.store.result[1,]
slopes.ci <- temp.store.result[-1,]
OR.ci <- exp(slopes.ci)
output <- list(regression.table = usual.output, intercept.ci = intercept.ci,
slopes.ci = slopes.ci, OR=OR, OR.ci = OR.ci)
return(output)
}
```

```
setwd("J:/Ref36/Referat/36.2 - Krankenanstalten/2 (9.21) - WEISSES BUCH/2 (9.21.11) - Weisses
Buch 2011/Projekte/stationäre Versorgungsforschung/1_Daten")
# Daten Laden
speicherpfad_grafik <- "J:/Ref36/Referat/36.2 - Krankenanstalten/2 (9.21) - WEISSES BUCH/2
(9.21.11) - Weisses Buch 2011/Projekte/stationäre
Versorgungsforschung/6_Diagramme/GIS_Karten"
conn <- odbcConnect("geo")
#read.table(file="myo_tod.csv", header=TRUE, sep=";")
```

```
todwsk <- sqlQuery(conn, "
SELECT t1.id, t2.road_distance, t1.ent_art, t1.alter_e, t1.sex
FROM diag_sa1_2009 t1
INNER JOIN distances t2
ON t1.ka = t2.hospid
AND t1.plz = t2.postcode
INNER JOIN diag_sa2_2009 t3
USING(id)
WHERE t1.plz IS NOT NULL
AND t1.auf_art <> 'T'
AND t3.diagn_typ = 'H'
AND substring(t3.icd_sub from 1 for 3) IN ('I63','I64')
")
```

```
# Daten umcodieren
```

```
for (i in 1:length(todwsk$ent_art)) {
if (is.na(todwsk[i,3])) {todwsk[i,3] <- "E"}
}
```

```
tot <- rep(1,length(todwsk$ent_art))
for (i in 1:length(todwsk$ent_art)) {
if (as.character(todwsk[i,3]) == 'S') {tot[i] <- 1}
else {tot[i] <- 0}
}
```

```
todwsk_d <- data.frame(id = todwsk$id, road_distance = todwsk$road_distance, tod = tod, alter = todwsk$alter_e, sex = todwsk$sex)
```

```
#Regression:
```

```
output <- glm(tod ~ road_distance + alter + as.factor(sex), family=binomial, data=todwsk_d)
```

```
oroutput <- logistic.regression.or.ci(output) # Funktion, damit du Konfidenzintervalle für die Odds-Ratios hast
```

```
orundci <- cbind(oroutput$OR, oroutput$OR.ci[,1], oroutput$OR.ci[,2]) # Odds Ratio und oberes und unteres Konfidenzintervall)
```

```
rownames(orundci)[1] <- "Distanz (pro km)"
```

```
rownames(orundci)[2] <- "Alter (pro Jahr)"
```

```
rownames(orundci)[3] <- "Geschlecht W"
```

```
tabletext <- cbind(c(rownames(orundci)), c(rep("",3)), c(rep("",3)), c(format(orundci[,1], digits=4)))
```

```
setwd(speicherpfad_grafik)
```

```
png(filename = "HERZ_todwsk_2009.png", width = 300*2, height = 450)
```

```
forestplot(tabletext[1:3,], orundci[1:3,1], orundci[1:3,2], orundci[1:3,3],
```

```
title(main="Odds Ratios für Tod nach Myokardinfarkt",
```

```
sub = "binäre logistische Regression, unabhängige Variablen: Distanz, Alter, Geschlecht"),
```

```
zero = 1,
```

```
xlab="Odds Ratio",
```

```
xlog = FALSE,
```

```
summary = TRUE,
```

```
boxsize = .75,
```

```
col = meta.colors(box="royalblue", line="darkblue", zero="red"))
```

```
dev.off()
```

## 7.4 Sensitivitätsanalyse

```
#####
```

```
# Einzugsbereiche bei Operationen
```

```
# Martin Scheuringer
```

```
# Modell mit Residuen nicht mit Quoten
```

```
#####  
#####
```

## #1. Vorbereitungen

```
# Hauptspeicher leeren  
rm(list = ls())
```

```
# Packages laden  
library(raster) # GIS  
library(sp) # GIS  
library(lattice) # Graphik  
library(RODBC) # Datenbankzugang  
library(classInt) # Benötigt für die Gruppierung der Werte  
require(graphics)
```

```
# Daten fg sterreichs Grenzen laden - 2 bedeutet mit Bezirksgrenzen --> siehe  
http://www.gadm.org/country  
# Dann laden
```

```
setwd("C:/Users/Scheuringer/GIS")  
load(file = "AUT_adm2.RData")
```

```
# Zugang zur Datenbank und Abfrage  
conn <- odbcConnect("geo")
```

```
# Daten fr Punkte in GIS-Karte laden  
plz <- sqlQuery(conn,"SELECT plz, ew, longitude, latitude FROM ew_2009_plz") # Fr PLZ-  
Koordinaten  
ka <- sqlQuery(conn,"SELECT hospid, beds, longitude, latitude FROM hospitals") #  
Krankenanstaltenstandorte
```

```
# Leistungen definieren  
sql_leist <- "('l21','l22')"
```

```
# Speicherpfade definieren - mssen angelegt sein!  
speicherpfad_grafik <- "J://Ref36/Referat/36.2 - Krankenanstalten/2 (9.21) - WEISSES BUCH/2  
(9.21.11) - Weisses Buch 2011/Projekte/stationre  
Versorgungsforschung/6_Diagramme/GIS_Karten"  
speicherpfad_tabs <- "J://Ref36/Referat/36.2 - Krankenanstalten/2 (9.21) - WEISSES BUCH/2  
(9.21.11) - Weisses Buch 2011/Projekte/stationre Versorgungsforschung/4_R/tabs_GIS_Produktion"  
# Name der Grafiken definieren "Leistungsname"  
name <- "Sensitivitt_Myokrard"
```

```
#####  
#####
```

## # 2. Berechnung der Quote je plz

```
# absolute Hufigkeit der Eingriffe je plz
```

```
sql_statment_plz_a <- paste("  
SELECT t2.plz, COUNT(substring(t1.icd_sub from 1 for 3)) AS anz
```

```

FROM diag_sa2_2009 t1
INNER JOIN diag_sa1_2009 t2 ON t1.id = t2.id
WHERE t2.auf_art <> 'T'
AND substring(t1.icd_sub from 1 for 3) IN", sql_leist, sep = "")

sql_statment_plz <- paste(sql_statment_plz_a,"AND t2.plz IN (SELECT plz FROM ew_2009_plz)
GROUP BY t2.plz", sep = "")

abs_eingr <- sqlQuery(conn, sql_statment_plz)

# merge: zur Anzahl der Eingriffe die Geo-Daten geben
eingr_pro_ew <- merge(plz,abs_eingr, by = 'plz', all.x = TRUE)
quote <- eingr_pro_ew$anz/eingr_pro_ew$ew
ew <- eingr_pro_ew$ew

# Festlegung der Grenzen der Brüche
brueche <- classIntervals(ew, n=90, style="quantile") #90 Intervalle für 90 Graustufen

# Hinzufügen der Größenkategorie zur plz: Nimm Quote und füge Kategorie der Quote hinzu
ew <- cbind(ew, "kategorie" = cut(ew, breaks = c(0, brueche$brks[2:length(brueche$brks)]), labels =
FALSE))

# in einen dataframe speichern
eg_plz <- data.frame(eingr_pro_ew,ew,quote)

#####
#####
# Daten sortieren und Farben definieren

ord1 <- order(eg_plz$quote, decreasing = TRUE)
eg_plz_plot <- eg_plz[ord1, ]

grauwert <- eg_plz_plot$kategorie*(-1)+90
graustufe <- paste("grey",grauwert,sep="")

#####
# regionale Variabilität Output

# Grafikdatei
setwd(speicherpfad_grafik)
png(filename = paste(name,"quoten_je_plz_2009_2.png", sep =""), width = 1200, height = 900)
#pdf(file = paste(name,"quoten_je_plz_2009.pdf", sep =""), width = 11.96, height = 8.27)
plot(gadm, bg = "white", border = "sienna")
title(main = paste(name,"-Aufenthalte je PLZ_2",sep=""), sub = "Jahr 2009", cex.main = 1.6, cex.sub
= 1.4)
points(x = eg_plz_plot$longitude, y = eg_plz_plot$latitude, pch = 16, col = graustufe, bg = FALSE,
cex = eg_plz_plot$quote*200, type = "p")

```

```
dev.off()
```

```
png(filename = paste(name,"hist_quoten_je_plz_2009_2.png", sep =""), width = 1200, height = 900)
hist(eg_plz_plot$quote*1000,
     main = "Histogramm der Quoten Modell 2",
     sub = name,
     cex.main = .75,
     axes = TRUE,
     col = "gray75",
     xlab = "Quote",
     cex.lab = .75,
     ylab = "Häufigkeit",
     xlim = c(0,20),
     ylim = c(0,150),
     breaks = 250
    )
dev.off()
```

```
sql_statment_plz_a <- paste("
SELECT t2.plz, COUNT(substring(t1.icd_sub from 1 for 3)) AS anz
FROM diag_sa2_2009 t1
INNER JOIN diag_sa1_2009 t2 ON t1.id = t2.id
--WHERE t2.auf_art <> 'T'
WHERE substring(t1.icd_sub from 1 for 3) IN", sql_leist, sep = "")
```

```
sql_statment_plz <- paste(sql_statment_plz_a,"AND t2.plz IN (SELECT plz FROM ew_2009_plz)
GROUP BY t2.plz", sep = "")
```

```
abs_eingr1 <- sqlQuery(conn, sql_statment_plz)
```

```
# merge: zur Anzahl der Eingriffe die Geo-Daten geben
eingr_pro_ew1 <- merge(plz,abs_eingr1, by = 'plz', all.x = TRUE)
quote1 <- eingr_pro_ew1$anz/eingr_pro_ew1$ew
ew1 <- eingr_pro_ew1$ew
```

```
# Festlegung der Grenzen der Brüche
brueche1 <- classIntervals(ew1, n=90, style="quantile") #90 Intervalle für 90 Graustufen
```

```
# Hinzufügen der Größenkategorie zur plz: Nimm Quote und füge Kategorie der Quote hinzu
ew1 <- cbind(ew1, "kategorie" = cut(ew1, breaks = c(0, brueche1$brks[2:length(brueche1$brks)]),
labels = FALSE))
```

```
# in einen dataframe speichern
eg_plz1 <- data.frame(eingr_pro_ew1,ew1,quote1)
```

```
png(filename = paste(name,"hist_quoten_je_plz_2009_1.png", sep =""), width = 1200, height = 900)
hist(eg_plz1$quote1*1000,
     main = "Histogramm der Quoten Modell 1",
     sub = name,
     cex.main = .75,
     axes = TRUE,
```

```

col = "gray75",
xlab = "Quote",
cex.lab = .75,
ylab = "Häufigkeit",
xlim = c(0,20),
ylim = c(0,150),
breaks = 250
)
dev.off()
####

#Differenz der Quoten
#length(eg_plz1$plz)
#head(eg_plz1)

# Modell 1 - Modell 2 --> In Modell 2 kommt es zu geringeren Quoten: Wer also einen hohen Wert
hat hat viele Verlegungen
diff <- eg_plz1$quote1 - eg_plz$quote

eg_plz_diff <-
data.frame(eg_plz$plz, eg_plz$ew, eg_plz$longitude, eg_plz$latitude, eg_plz$kategorie, diff)

ord1 <- order(eg_plz_diff$diff, decreasing = TRUE)
eg_plz_plot_diff <- eg_plz_diff[ord1, ]

# Grafikdatei
setwd(speicherpfad_grafik)
png(filename = paste(name, "quotendifferenz_je_plz_2009_1-2.png", sep = ""), width = 1200, height
= 900)
#pdf(file = paste(name, "quoten_je_plz_2009.pdf", sep = ""), width = 11.96, height = 8.27)
plot(gadm, bg = "white", border = "sienna")
title(main = paste(name, "-Differenz Modell1 - Modell 2", sep = ""), sub = "Jahr 2009", cex.main = 1.6,
cex.sub = 1.4)
points(x = eg_plz_plot_diff$eg_plz.longitude, y = eg_plz_plot_diff$eg_plz.latitude, pch = 16, col =
graustufe, bg = FALSE, cex = eg_plz_plot_diff$diff*1000, type = "p")

dev.off()

png(filename = paste(name, "hist_quotendifferenz_je_plz_2009_2.png", sep = ""), width = 1200,
height = 900)
hist(eg_plz_plot_diff$diff,
main = "Histogramm der Differenz der Quoten Modell 1 - 2",
sub = name,
cex.main = .75,
axes = TRUE,
col = "gray75",
xlab = "Differenz der Quote",
cex.lab = .75,
ylab = "Häufigkeit",

```

```
xlim = c(0,0.01),  
breaks = 250  
)  
dev.off()
```