



**Entwicklung und Anwendung von Modellen zur
Berechnung von Schwellenwerten bei
Mindestmengen für die Koronarchirurgie**

- Vorbericht -

[Auftrag B05/01b]

Version 1.0

Stand: 24.04.2006

Thema: Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für die Koronarchirurgie

Auftraggeber: Gemeinsamer Bundesausschuss
Datum des Auftrags: 22.12.2004, konkretisiert am 26.08.2005
Interne Auftragsnummer: B05/01b

Kontakt:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Dillenburger Str. 27

51105 Köln

Internet: www.iqwig.de

Tel: 0221/35685-0

Fax: 0221/35685-1

Email: B05-01@iqwig.de

Bemerkungen:

Im folgenden Text wurde bei der Angabe von Personenbezeichnungen jeweils die männliche Form angewandt. Dies erfolgte ausschließlich zur Verbesserung der Lesbarkeit.

Bei dem vorliegenden Vorbericht handelt es sich um eine **vorläufige Darstellung** der Ergebnisse zu Mindestmengen bei der Koronarchirurgie. Zu diesem Bericht können Stellungnahmen abgegeben werden, die gegebenenfalls zu einer Ergänzung und/oder Überarbeitung des Berichts führen können.

Inhaltsverzeichnis

	Seite
Inhaltsverzeichnis	2
Tabellenverzeichnis	3
Abbildungsverzeichnis.....	4
Abkürzungsverzeichnis	5
1. Ziel der Untersuchung.....	6
2. Hintergrund.....	6
2.1 Auftrag	6
2.2 Mindestmengen für die Koronarchirurgie	7
3. Projektablauf einschließlich Konkretisierung	8
4. Methoden	9
4.1 Datenquellen	9
4.2 Vorgaben der Fachexperten	9
4.3 Auswertungen	10
4.3.1 Datenbeschreibung.....	10
4.3.2 Modellbildung.....	12
4.3.3 Berechnung von Schwellenwerten.....	14
4.3.4 Multiples Testen und Signifikanzniveau	14
4.3.5 Software	14
4.4 Ergänzungen im Vergleich zum Berichtsplan	14
5. Ergebnisse	15
5.1 Deskriptive Statistik.....	15
5.2 Ergebnisse für den Indikator "Letalität"	19
5.2.1 Zusammenhang zwischen Fallzahl und Risiko (2004).....	19
5.2.2 Zusammenhang zwischen Fallzahl und Risiko (2003).....	24
5.2.3 Schlussfolgerung.....	25
5.3 Ergebnisse für den Indikator "Infektion"	26
5.3.1 Zusammenhang zwischen Fallzahl und Risiko (2004).....	26
5.3.2 Zusammenhang zwischen Fallzahl und Risiko (2003).....	30
5.3.3 Schlussfolgerung.....	31
5.4 Ergebnisse für den Indikator "Brustwandarterie"	32
5.4.1 Zusammenhang zwischen Fallzahl und Risiko (2004).....	32
5.4.2 Zusammenhang zwischen Fallzahl und Risiko (2003).....	36
5.4.3 Schlussfolgerung.....	37
5.5 Ergebnisse für den Indikator "Anastomosen"	38
5.5.1 Zusammenhang zwischen Fallzahl und Risiko (2004).....	38
5.5.2 Zusammenhang zwischen Fallzahl und Risiko (2003).....	41
5.5.3 Schlussfolgerung.....	42
6. Diskussion	43
7. Zusammenfassung	44
8. Literatur	46
Anhang: Weiterer Zeitplan.....	48

Tabellenverzeichnis

Tabelle 1	Deskriptive Übersicht
Tabelle 2	Rangkorrelationen aller Variablen
Tabelle 3	Rohe Risikorate getrennt nach Notfall ja/nein (in %)
Tabelle 4	Einfaches logistisches Regressionsmodell für "Letalität" (Modell I, 2004)
Tabelle 5	Multipl. logistisches Regressionsmodell für "Letalität" (Modell II, 2004)
Tabelle 6	Multipl. logistisches Regressionsmodell für "Letalität" mit FPs (Modell III, 2004)
Tabelle 7	Ergebnisse des GEE-Ansatzes für "Letalität" (Modell IV, 2004)
Tabelle 8	Multipl. logistisches Regressionsmodell für "Letalität" mit FPs (Modell III, 2003)
Tabelle 9	Einfaches logistisches Regressionsmodell für "Infektion" (Modell I, 2004)
Tabelle 10	Multipl. logistisches Regressionsmodell für "Infektion" (Modell II, 2004)
Tabelle 11	Multipl. logistisches Regressionsmodell für "Infektion" mit FPs (Modell III, 2004)
Tabelle 12	Ergebnisse des GEE-Ansatzes für "Infektion" (Modell IV, 2004)
Tabelle 13	Multipl. logistisches Regressionsmodell für "Infektion" mit FPs (Modell III, 2003)
Tabelle 14	Logistisches Regressionsmodell für "Brustwandarterie" mit FP für Fallzahl (Modell I, 2004)
Tabelle 15	Multipl. logistisches Regressionsmodell für "Brustwandarterie" mit FPs für Fallzahl und Kovariablen (Modell II, 2004)
Tabelle 16	Ergebnisse des GEE-Ansatzes für "Brustwandarterie" (Modell III, 2004)
Tabelle 17	Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-III für "Brustwandarterie" (2004)
Tabelle 18	Ergebnisse des GEE-Ansatzes für "Brustwandarterie" (Modell III, 2003)
Tabelle 19	Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-III für "Brustwandarterie" (2003)
Tabelle 20	Logistisches Regressionsmodell für "Anastomosen" mit FP für Fallzahl (Modell I, 2004)
Tabelle 21	Multipl. logistisches Regressionsmodell für "Anastomosen" mit FPs für Fallzahl und Kovariablen (Modell II, 2004)
Tabelle 22	Ergebnisse des GEE-Ansatzes für "Anastomosen" (Modell III, 2004)
Tabelle 23	Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-III für "Anastomosen" (2004)
Tabelle 24	Ergebnisse des GEE-Ansatzes für "Anastomosen" (Modell III, 2003)
Tabelle 25	Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-III für "Anastomosen" (2003)

Abbildungsverzeichnis

Abbildung 1	Verteilung der Fallzahl pro Krankenhaus (2003 oben, 2004 unten)
Abbildung 2	Risikoplot für "Letalität" vs. Fallzahl (2004)
Abbildung 3	Rohe Risikokurve mit 95%-Konfidenzband für "Letalität" aus Modell I (2004)
Abbildung 4	Adjustierte Risikokurve mit 95%-Konfidenzband für "Letalität" aus Modell III (2004) (Kovariablen auf Mittelwert gesetzt)
Abbildung 5	Risikoplot für "Letalität" vs. Fallzahl (2003)
Abbildung 6	Risikoplot für "Infektion" vs. Fallzahl (2004)
Abbildung 7	Rohe Risikokurve mit 95%-Konfidenzband für "Infektion" aus Modell I (2004)
Abbildung 8	Adjustierte Risikokurve mit 95%-Konfidenzband für "Infektion" aus Modell III (2004) (Kovariablen auf Mittelwert gesetzt)
Abbildung 9	Risikoplot für "Infektion" vs. Fallzahl (2003)
Abbildung 10	Risikoplot für "Brustwandarterie" vs. Fallzahl (2004)
Abbildung 11	Risikoplot für "Brustwandarterie" vs. Fallzahl (2003)
Abbildung 12	Risikoplot für "Anastomosen" vs. Fallzahl (2004)
Abbildung 13	Risikoplot für "Anastomosen" vs. Fallzahl (2003)

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AG	Arbeitsgruppe
BMI	Body Mass Index
BQS	BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH
CDC	Centers for Disease Control and Prevention
FG	Freiheitsgrad
FP	Fractional Polynomials
G-BA	Gemeinsamer Bundesausschuss
GEE	Generalized Estimating Equations
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KCH	Koronarchirurgie
KI	Konfidenzintervall
Knie-TEP	Knie-Totalendoprothese
NNIS	National Nosocomial Infections Surveillance System
OP	Operation
OR	Odds Ratio
PIN	Population Impact Number

1. Ziel der Untersuchung

Ziel des Berichts ist es zu untersuchen, ob anhand deutscher Daten Schwellenwerte oder Bezugswerte (Benchmarks) bezüglich Mindestmengen (minimale Anzahl von Leistungen pro Jahr, die notwendig ist, um diese Leistung erbringen zu dürfen) für den Bereich "Koronarchirurgie" (operative Versorgung von verengten Herzkranzgefäßen) ableitbar sind. Wenn ja, werden für diesen Zweck adäquate Regressionsmodelle entwickelt und diese zur Berechnung von Schwellenwerten bzw. Benchmarks angewendet. Außerdem wird untersucht und diskutiert, ob sich anhand deutscher Daten ein statistischer Zusammenhang zwischen der Behandlungsmenge und Indikatoren der Prozessqualität ergibt.

2. Hintergrund

2.1 Auftrag

Der Gemeinsame Bundesausschuss (G-BA) hat mit Schreiben vom 22.12.2004 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Entwicklung eines Rechenmodells zur Festlegung von Schwellenwerten für Mindestmengen beauftragt. Dieses Rechenmodell sollte unter Berücksichtigung deutscher Daten für die Leistungsbereiche Knie-Totalendoprothese (Knie-TEP) und Koronarchirurgie und darüber hinaus "im folgenden auch generell" anwendbar sein.

Im Rahmen der Auftragskonkretisierung wurde der Auftrag in Bezug auf die beiden Leistungsbereiche geteilt und am 26.08.2005 für die Koronarchirurgie neu formuliert. Der Auftragsgegenstand lautet nun wie folgt: Das IQWiG soll adäquate Rechenmodelle für den Leistungsbereich Koronarchirurgie entwickeln und diese zur Ermittlung von Schwellenwerten anwenden. Dabei ist eine enge Zusammenarbeit mit der Bundesgeschäftsstelle für Qualitätssicherung (BQS) zu gewährleisten, da diese die Datengrundlage zur Bearbeitung des Auftrags zur Verfügung stellt.

2.2 Mindestmengen für die Koronarchirurgie

In den letzten drei Jahrzehnten wurden für verschiedene Leistungen der Gesundheitsversorgung eine Vielzahl von Studien zu einer möglichen Assoziation zwischen Behandlungsmenge und Ergebnisqualität publiziert [1]. Vor diesem Hintergrund wurden im Rahmen des Fallpauschalengesetzes vom April 2002 die Partner der Selbstverwaltung aufgefordert, einen Katalog planbarer Leistungen aufzustellen, bei denen die Ergebnisqualität in besonderem Maße von der Behandlungsmenge abhängt und für diese Leistungen Mindestmengen je Arzt und Krankenhaus festzulegen [2]. Ein solcher Katalog wurde im Jahr 2003 für verschiedene operative Eingriffe vorgelegt. In der Sitzung des G-BA am 21.09.2004 wurde gemäß § 91 Abs. 7 beschlossen, diesen Mindestmengenkatalog um die Leistungen Knie-TEP und koronarchirurgische Eingriffe zu erweitern. Hierfür wurden jedoch noch keine konkreten Mindestmengen festgelegt. Dies sollte erst dann erfolgen, wenn die zur Beschlussfassung erforderlichen wissenschaftlichen Daten vorliegen. Im Dezember 2004 wurde das IQWiG damit beauftragt, auf der Basis deutscher Daten für die Indikationen Knie-TEP und Koronarchirurgie Modelle für die Berechnung von Schwellenwerten zu entwickeln und anzuwenden.

Dieser Auftrag wurde neu formuliert und am 09.06.2005 für die Knie-TEP (Auftrag B05/01a) und am 26.08.2005 für die Koronarchirurgie konkretisiert (Auftrag B05/01b). Der Abschlussbericht zur Knie-TEP wurde am 05.12.2005 dem G-BA vorgelegt und am 06.02.2006 im Internet publiziert. In dem nun vorliegenden Bericht zum Auftrag B05/01b werden die Ergebnisse zur Koronarchirurgie dargestellt.

3. Projektablauf einschließlich Konkretisierung

Die Bearbeitung des Auftrags erfolgte nach dem im Berichtsplan dargestellten Schema, das bereits während der Konkretisierung zum Teil durchlaufen war. In Klammern stehen die jeweils verantwortlichen Institutionen. Die Festlegung der Indikatoren der Ergebnis- und der Prozessqualität, die Einflussfaktoren sowie die Festlegung der klinisch relevanten Unterschiede wurden von den vom Unterausschuss "*Sonstige stationäre Qualitätssicherung*" benannten Fachexperten für die Koronarchirurgie vorgenommen. Bei diesen Entscheidungen wurde berücksichtigt, welche Daten von der BQS zur Verfügung gestellt werden können.

- 1) Fachexperten benennen
(Deutsche Krankenhausgesellschaft und Gesetzliche Krankenversicherung)
- 2) Qualitätsindikatoren und Einflussfaktoren festlegen
(Fachexperten mit BQS)
- 3) Datenbasis festlegen (BQS-Daten)
(Unterausschuss "*Sonstige stationäre Qualitätssicherung*")
- 4) Mittleres Risiko der Patienten und Quartile aus 2003er Daten berechnen
(IQWiG)
- 5) Klinisch relevante Unterschiede sowie Wichtigkeit der Indikatoren
(ggf. Rangfolge) festlegen und dokumentieren
(Fachexperten, Konsentierung in AG Mindestmengen)
- 6) Zusammenhang darstellen mit 2003er Daten (Risikoplot, Regression)
(IQWiG)
- 7) Modellwahl
(IQWiG)
- 8) Ggf. Vorgaben zur Schwellenwertbestimmung festlegen (z.B. absolute bzw. relative Risikoreduktion bzw. Gradient)
(Festlegung erfolgt verblindet bezüglich der Fallzahl von Fachexperten, Konsentierung in AG Mindestmengen)
- 9) Ggf. Schwellenwertberechnung für 2004er Daten
(IQWiG)

Der Berichtsplan zu diesem Auftrag wurde am 05.09.2005 fertig gestellt und am 14.09.2005 veröffentlicht. Die erforderlichen Daten für die Bearbeitung des Auftrags wurden von der BQS am 10.01.2006 geliefert.

4. Methoden

Die Methoden zur Erstellung des Berichts wurden im Berichtsplan vom 05.09.2005 vorab festgelegt. Sofern sich im Verlauf diesbezüglich Änderungen ergeben haben, sind diese im Abschnitt 4.4 dargestellt.

4.1 Datenquellen

Es wurden die Daten der BQS aus dem Bundesdatenpool 2003 und 2004 verwendet (siehe Abschnitt 4.3.1). Diese Daten wurden für den vorliegenden Auftrag von der BQS aufbereitet und dem IQWiG am 10.01.2006 so zur Verfügung gestellt, dass auf der Beobachtungseinheit "Patient" regressionsanalytische Berechnungen möglich sind. Die Fallzahl der behandelten Patienten bezieht sich hierbei jeweils auf ein Krankenhaus. Die Daten wurden in pseudonymisierter Form unter Beachtung der amtlichen Datenschutzrichtlinien aufbereitet und statistisch ausgewertet.

4.2 Vorgaben der Fachexperten

Im Rahmen der Auftragskonkretisierung wurden von den Fachexperten für die Koronarchirurgie des G-BA die Indikatoren für die Ergebnisqualität (Letalität, tiefe sternale Infektion) bestimmt, für die aus dem Zusammenhang mit der Leistungsmenge Schwellenwerte berechnet werden sollen. Außerdem wurden für deskriptive Zwecke Indikatoren der Prozessqualität (Verwendung der inneren Brustwandarterie, Anzahl peripherer Anastomosen) festgelegt, für die lediglich untersucht wird, ob es einen Zusammenhang mit der Leistungsmenge gibt. Schwellenwerte werden für die Indikatoren der Prozessqualität nicht berechnet.

Des Weiteren wurden von den Fachexperten am 01.03.2006 folgende Festlegungen bezüglich der Reihenfolge und der erforderlichen Effektstärken der Indikatoren der Ergebnisqualität für eine Mindestmengenregelung getroffen: (Genauere Informationen zu den Kodierungen der betreffenden Variablen findet man im Abschnitt 4.3.1).

A) Rangfolge der Qualitätsindikatoren bezüglich ihrer Wichtigkeit

(gemeint ist hierbei die Wichtigkeit bezüglich einer Mindestmengenregelung)

- 1) Letalität
- 2) Infektion

B) Klinisch relevante Veränderung der Population Impact Number (PIN)

(PIN gibt die erwartete Anzahl der Patienten an, von denen einer durch die Einführung einer Mindestmenge bezüglich des jeweiligen Indikators einen Nutzen hat [3].)

- 1) Letalität: PIN=500
- 2) Infektion: PIN=500

Für die Indikatoren der Prozessqualität sind keine weiteren Vorgaben notwendig, da für diese keine Schwellenwerte berechnet werden.

4.3 Auswertungen

4.3.1 Datenbeschreibung

Die primäre Datenbasis besteht aus Daten von 54335 Patienten mit isoliert koronarchirurgischen Operationen aus 77 Krankenhäusern aus dem Bundesdatenpool 2004 der BQS (54653 Operationen an 54335 Patienten) [4]. Im Rahmen dieser Auswertungen wurden nur die Daten der Erst-Operationen verwendet, so dass der Stichprobenumfang $n=54335$ beträgt. Für Untersuchungen zur Modellstabilität wurden auch die Daten von 43297 Patienten mit isoliert koronarchirurgischen Erst-Operationen aus 77 Krankenhäusern aus dem Bundesdatenpool 2003 (Datensatzversion 6.0) verwendet [5]. Der Stichprobenumfang von 2003 ist niedriger als von 2004, weil die Daten von 12411 Patienten mit isoliert koronarchirurgischen Erst-Operation aus dem Jahr 2003 in einer anderen Datenstruktur (Datensatzversion 5.01) vorliegen und nicht ohne weiteres mit den anderen Daten zusammengeführt werden können. Die beiden gewählten Indikatoren für die Ergebnisqualität sind wie folgt definiert.

- 1) Letalität: Krankenhaussterblichkeit (ja / nein)
- 2) Infektion: postoperative Mediastinitis (ja / nein)

Die beiden gewählten Indikatoren für die Prozessqualität sind wie folgt definiert.

- 1) Brustwandarterie: keine Verwendung der inneren Brustwandarterie (ja / nein)
- 2) Anastomosen: Anzahl peripherer Anastomosen geringer als 3 (ja / nein)

Der Indikator "Brustwandarterie" wurde in der Form "keine Verwendung der inneren Brustwandarterie ja/nein" kodiert, damit für alle Qualitätsindikatoren das *Risiko für ein unerwünschtes Ereignis* in Abhängigkeit der Fallzahl untersucht wird und so die Richtung einer Qualitätsverschiebung in gleicher Weise zu interpretieren ist. Außerdem ist es aus statistischen Gründen im Rahmen der Modellbildung bei der logistischen Regression vorteilhaft, als Zielvariable das Ereignis zu wählen, welches seltener eintritt.

Der Indikator "Anastomosen" liegt in drei Ausprägungen vor: 1, 2 und ≥ 3 Anastomosen. Wegen der vergleichsweise geringen Anzahl von Fällen mit einer Anastomose wurden Patienten mit 1 und 2 Anastomosen zusammengefasst zur Kategorie "<3 Anastomosen", d.h. für die weiteren Untersuchungen wurde die Variable "Anastomosen" dichotomisiert. Aus den gleichen Gründen wie oben beschrieben wird bei den regressionsanalytischen Auswertungen das Risiko für das Auftreten von weniger als 3 peripheren Anastomosen betrachtet.

Zur Bildung adäquater Teilstichproben sowie zur Durchführung einer Risiko-adjustierung, wurden die folgenden potentiellen Risikofaktoren zur Verfügung gestellt:

- (a) Euro-Score: in Prozent
- (b) KCH-Score: in Prozent
- (c) BMI: in kg/m^2
- (d) Dringlichkeit: 1 = elektiv
2 = dringlich
3 = Notfall
4 = Notfall (Reanimation / ultima ratio)
- (e) CDC-Klasse: Werte 0, 1, 2, 3
Risikoklassen des National Nosocomial Infections Surveillance Systems (NNIS) des Centers for Disease Control and Prevention (CDC) [6]

Wegen der geringen Anzahl von Patienten mit CDC-Klasse 3 (7 Fälle) wurden diese für multifaktorielle Auswertungen mit CDC-Klasse 2 zusammengefasst, so dass die Variable CDC-Klasse in den Regressionsmodellen nur drei geordnete Ausprägungen 0, 1 und 2 hat. Diese Variable wurde als quasi-stetige Variable modelliert, in Sensitivitätsanalysen aber auch als kategorielle Variable betrachtet.

4.3.2 Modellbildung

Die Modellbildung wurde für die beiden gewählten Indikatoren der Ergebnisqualität ("Letalität" und "Infektion") sowie die beiden Indikatoren der Prozessqualität ("Brustwandarterie" und "Anastomosen"), jeweils getrennt in mehreren aufeinander aufbauenden Schritten, für die Daten des Bundesdatenpools 2004 vorgenommen. Für die primäre Datenanalyse wurden Notfälle (Dringlichkeitsklassen 3 und 4) ausgeschlossen, da Mindestmengenregelungen nicht für Notfälle gelten. Zur visuellen Darstellung des Zusammenhangs zwischen Fallzahl und Risiko wurden zunächst deskriptive Risikoplots erstellt [7]. Hierzu wird der Wertebereich der Fallzahl in 20 Gruppen mit gleicher Besetzungshäufigkeit unterteilt und dann das Risiko innerhalb der Gruppen berechnet. Die Risiken werden dann gegen den Gruppenmittelwert aufgetragen. Um einer möglichen Identifizierung einzelner Krankenhäuser vorzubeugen, wurden die drei Gruppen mit den größten Fallzahlen zu einer Gruppe zusammengefasst, so dass die Risikoplots auf 18 Gruppen basieren. Diese Risikoplots geben einen ersten Eindruck, in welcher Form die betrachtete Variable sinnvoll in einem logistischen Regressionsmodell untersucht werden kann.

Im Rahmen einfacher logistischer Regressionsmodelle [8] wurde der rohe Zusammenhang zwischen dem jeweiligen Qualitätsindikator und der Fallzahl analysiert. Bei Bedarf wurde hierbei die Methode der Fractional Polynomials (FP) eingesetzt, die im Rahmen der logistischen Regression eine flexible Modellierung auch nicht monotoner Zusammenhänge ermöglicht [9].

Anschließend wurden multiple logistische Regressionsanalysen durchgeführt. Neben der Fallzahl wurden für alle Indikatoren die Risikofaktoren (a) bis (e) untersucht. Da der Euro-Score und der KCH-Score inhaltlich identisch zu interpretieren sind und eine sehr hohe Korrelation aufweisen, wurde jeweils nur einer der Scores in den multifaktoriellen Modellen betrachtet. Zur Adjustierung wurde primär der Euro-Score verwendet, in Sensitivitätsuntersuchungen aber durch den KCH-Score ersetzt. Die Variable "Dringlichkeit" wurde primär zum Ausschluss von Notfällen verwendet. In Sensitivitätsuntersuchungen wurde die nach Ausschluss von Notfällen verbliebene binäre Größe "Dringlichkeit (ja/nein)"

aber auch als Kovariable eingesetzt. Für die stetigen Kovariablen der multifaktoriellen Modelle wurde ebenfalls die FP-Methode [9] zur flexiblen Modellierung eingesetzt. Dabei wurde gemäß der Selektionsprozedur von Royston und Sauerbrei [10] vorgegangen. Von allen Kovariablen wurden jeweils alle Wechselwirkungen 1. Ordnung im Rahmen eines Backward-Eliminationsverfahrens mit Einschlussgrenze $p=0.01$ untersucht. Da vorab keine Hypothesen für Wechselwirkungen vorhanden waren, wurde diese niedrige Signifikanzgrenze gewählt, um ggf. nur sehr deutliche Wechselwirkungseffekte zu berücksichtigen.

Danach wurde mit Hilfe der Methode der Generalized Estimating Equations (GEE) [11] untersucht, ob die Zugehörigkeit der Patienten zum gleichen Krankenhaus zu einem Cluster-Effekt führt. Die mögliche Bedeutung eines Cluster-Effekts in Untersuchungen zu Volume-Outcome-Beziehungen wurde in mehreren methodischen Arbeiten unterstrichen [12,13]. Nur wenn der GEE-Ansatz deutliche Hinweise auf einen Cluster-Effekt zeigte und zu relevanten Änderungen der Parameterschätzer führte, wurde das GEE-Modell zur endgültigen Modellierung verwendet. Ansonsten wurde das vorherige multifaktorielle Modell ohne Cluster-Effekt als endgültiges Modell betrachtet.

Zur Darstellung der Effektstärke wurden Odds Ratios (ORs) aus der logistischen Regression ermittelt, und zwar jeweils für eine Änderung der Fallzahl um 50 Fälle. Zur Beschreibung des Erklärungswertes der betrachteten Risikofaktoren wurden drei Maße für erklärte Variabilität berechnet: Die c-Statistik, die der Fläche unter der ROC-Kurve entspricht [7], das adjustierte Likelihood-Ratio R^2 nach Nagelkerke [14] sowie die von Mittelböck & Schemper [15] empfohlene Squared-Pearson-Correlation r^2 .

Die Robustheit der Modelle wurde mit Hilfe regressionsdiagnostischer Verfahren überprüft [16]. Sensitivitätsuntersuchungen wurden bezüglich fehlender Werte, extremer Werte und unterschiedlicher Modellierung der Kovariablen vorgenommen. Zur Untersuchung der Ergebnisstabilität wurden die gleichen Modelle auch auf die Daten des Bundesdatenpools 2003 angewendet. Die Ergebnisse für 2003 haben jedoch aufgrund der vielen fehlenden Werte beim Euro-Score (siehe Tab. 1) nur geringe Aussagekraft.

4.3.3 Berechnung von Schwellenwerten

Da für keinen der beiden betrachteten Indikatoren der Ergebnisqualität ein statistisch signifikanter Zusammenhang mit der Leistungsmenge gefunden wurde, wurde gemäß Berichtsplan keines der möglichen Verfahren zur Berechnung von Schwellenwerten angewendet [17].

4.3.4 Multiples Testen und Signifikanzniveau

Für alle Signifikanztests wurde ein Signifikanzniveau von $\alpha=0.05$ (zweiseitig) verwendet. Eine Adjustierung für multiples Testen wurde nicht vorgenommen, da es sich hier nicht um eine confirmatorische Prüfung handelt. Alle hier durchgeführten Signifikanztests haben nur hypothesengenerierenden Charakter.

4.3.5 Software

Für die Datenanalysen wurde SAS 9.13 verwendet, insbesondere die Prozeduren LOGISTIC und GENMOD. Für spezielle Berechnungen wurde die Matrixsprache SAS/IML verwendet.

4.4 Ergänzungen im Vergleich zum Berichtsplan

Im Laufe der Bearbeitung haben sich im Vergleich zu der im Berichtsplan vorab dargestellten Methodik geringfügige Änderungen bzw. Ergänzungen ergeben. Im Berichtsplan wurden bezüglich der Risikoadjustierung explizit nur der Euro-Score, der KCH-Score sowie der BMI genannt. Es wurden bei der Modellierung jedoch alle zur Verfügung gestellten potentiellen Risikofaktoren betrachtet (Euro-Score, KCH-Score, BMI, Dringlichkeit, CDC-Klasse).

5. Ergebnisse

5.1 Deskriptive Statistik

Die Daten wurden zunächst auf Plausibilität untersucht. Dabei zeigten sich bei einer Reihe von Patienten unplausible bzw. unmögliche Angaben zum BMI-Wert. Bei 195 Patienten mit BMI-Werten größer 60 wurden diese für die Folgeauswertungen auf fehlenden Wert gesetzt. Tabelle 1 gibt eine deskriptive Übersicht aller vorhandenen Daten. Tabelle 2 zeigt die Korrelationen (Spearman Rangkorrelation) der betrachteten Kovariablen.

Tab. 1: Deskriptive Übersicht

	Jahr 2003		Jahr 2004	
	fehlende Werte	Maßzahl	fehlende Werte	Maßzahl
Anzahl Patienten	-	43297	-	54335
Anzahl Krankenhäuser	-	77	-	77
Letalität (ja / nein)	-	3.2 / 96.8	-	3.1 / 96.9
Infektion (ja / nein)	-	0.8 / 99.2	-	0.8 / 99.2
keine Brustwandarterie (ja / nein)	2	14.5 / 85.5	-	13.2 / 86.8
Anzahl Anastomosen (1 / 2 / ≥3)	8	4.4 / 20.8 / 74.8	147	4.5 / 20.9 / 74.6
Präoperative Mediastinitis (ja / nein)	-	0.1 / 99.9	-	0.1 / 99.9
Vorhergegangene Koronarchirurgie (ja / nein)	644	4.0 / 96.0	68	3.7 / 96.3
Dringlichkeit (elektiv / dringlich / Notfall / Notfall: Reanimation, ultima ratio)	-	63.9 / 27.9 / 7.8 / 0.4	-	61.9 / 28.7 / 8.9 / 0.5
CDC-Klasse (0 / 1 / 2 / 3) [6]	261	34.3 / 54.0 / 11.7 / 0.03	413	26.9 / 58.1 / 15.0 / 0.01
Fallzahl	-	961.1 (512.0)	-	926.0 (517.0)
Anzahl Grafts	-	2.8 (0.9)	-	2.9 (0.9)
BMI	861	27.7 (4.0)	1226	27.8 (4.1)
Euro-Score	21415	5.5 (7.9)	6180	6.1 (8.3)
KCH-Score	21462	2.8 (4.5)	6308	3.0 (4.5)
Mittelwert (Standardabweichung) bzw. Häufigkeiten (in Prozent)				

Tab. 2: Rangkorrelationen aller Variablen

Jahr 2003							
	Fallzahl	Dringlichkeit ^a	CDC-Klasse	Anzahl Grafts	BMI	Euro-Score	KCH-Score
Fallzahl	1	-0.02	-0.10	-0.06	0.01	0.02	0.03
Dringlichkeit ^a		1	0.07	0.01	-0.03	0.28	0.28
CDC-Klasse			1	0.13	0.04	0.11	0.12
Anzahl Grafts				1	0.00	-0.02	-0.02
BMI					1	-0.10	-0.07
Euro-Score						1	0.92
Jahr 2004							
Fallzahl	1	-0.01	-0.13	-0.05	0.02	0.01	0.01
Dringlichkeit ^a		1	0.06	0.01	-0.03	0.26	0.27
CDC-Klasse			1	0.09	0.03	0.12	0.12
Anzahl Grafts				1	-0.01	-0.03	-0.04
BMI					1	-0.09	-0.06
Euro-Score						1	0.92
^a Dringlichkeit: 1=elektiv, 2=dringlich, 3=Notfall, 4=Notfall (Reanimation, ultima ratio)							

Es zeigen sich keine gravierenden Unterschiede in der Verteilung der Variablen in den Jahren 2003 und 2004. Die meisten Variablen, insbesondere auch Fallzahl und die übrigen Kovariablen, sind in beiden Jahren nur schwach miteinander korreliert. Eine Ausnahme bilden Euro-Score und KCH-Score. Bei deren Berechnung gehen im Wesentlichen identische Komponenten ein, was sich in einer sehr hohen Korrelation zwischen den Score-Werten niederschlägt ($r = 0.92$ in beiden Jahren, siehe Tabelle 1). Eine positive Korrelation zeigt sich zudem zwischen der Dringlichkeit und dem Euro-Score sowie dem KCH-Score.

Abbildung 1 zeigt die patientenbezogene Verteilung der Fallzahl pro Krankenhaus in den Jahren 2003 und 2004. Tabelle 3 zeigt das Risiko für die betrachteten Zielvariablen getrennt nach elektiv/dringlichen und Notfall-OPs. Für die Berechnungen wurden beim Indikator "Infektion" Patienten mit präoperativer Mediastinitis ausgeschlossen, beim Indikator "Brustwandarterie" wurden nur Patienten ohne vorherigen koronarchirurgischen Eingriff berücksichtigt. Das Risiko ist bei den Notfall-OPs durchweg - teils deutlich - höher.

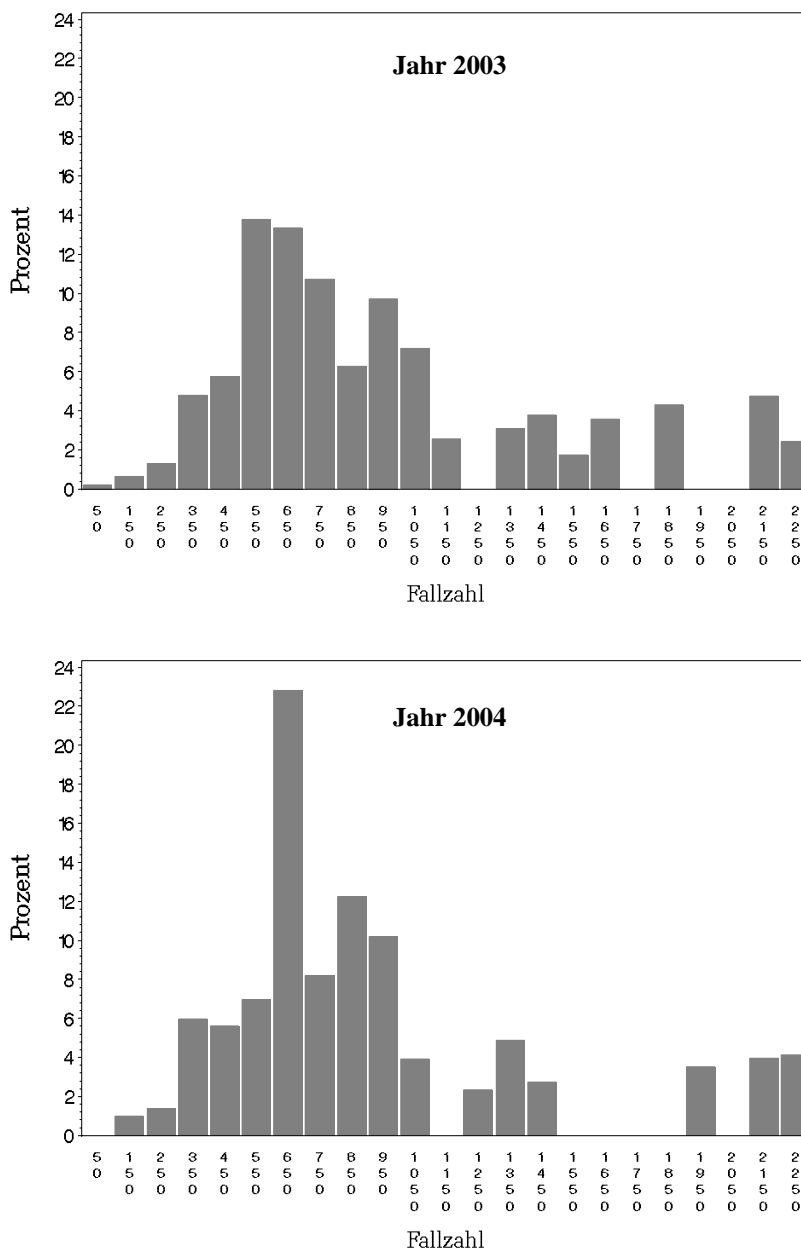


Abb. 1: Verteilung der Fallzahl pro Krankenhaus (2003 oben, 2004 unten)
 (Die Zahlen an der horizontalen Achse geben die Klassenmittelwerte an)

Tab. 3: Rohe Risikorate getrennt nach Notfall ja/nein (in %)

Jahr 2003			
Variable	gesamt	Notfall	kein Notfall
Letalität	3.2	10.3	2.5
Infektion	0.8	1.1	0.8
keine Brustwandarterie	13.1	24.9	12.0
Anastomosen<=1	4.4	5.6	4.3
Anastomosen<=2	25.2	26.7	25.1
Jahr 2004			
Variable	gesamt	Notfall	kein Notfall
Letalität	3.1	9.8	2.4
Infektion	0.8	1.2	0.8
keine Brustwandarterie	11.7	23.0	10.5
Anastomosen<=1	4.5	5.4	4.4
Anastomosen<=2	25.4	27.0	25.2

5.2 Ergebnisse für den Indikator "Letalität"

5.2.1 Zusammenhang zwischen Fallzahl und Risiko (2004)

Zur Untersuchung des Zusammenhangs zwischen Fallzahl und "Letalität" wurden nur elektive und dringliche Eingriffe herangezogen, d.h. Notfall-Operationen wurden ausgeschlossen. Es resultierten n=49227 Datensätze. Das mittlere Risiko beträgt 2.4%. Abbildung 2 zeigt den Risikoplot für den Indikator "Letalität".

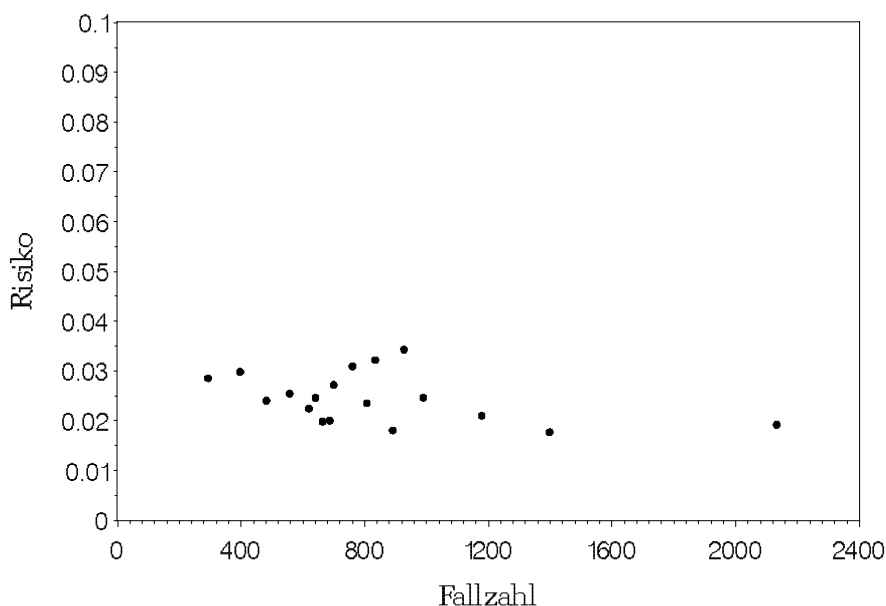


Abb. 2: Risikoplot für "Letalität" vs. Fallzahl (2004)

Es wurde zunächst ein einfaches logistisches Regressionsmodell mit der Fallzahl als Einflussfaktor berechnet. Die Ergebnisse sind in Tabelle 4 dargestellt (Modell I). Zwar ist die Assoziation zwischen Fallzahl und Letalitätsrisiko deutlich statistisch signifikant ($p < 0.0001$), aber die Fallzahl hat nur einen sehr niedrigen Erklärungswert.

Tab. 4: Einfaches logistisches Regressionsmodell für "Letalität" (Modell I, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR ^a	95%-KI
Intercept	1	-3.4765	0.0610	< 0.0001	-	-
Fallzahl	1	-0.00025	0.00006	< 0.0001	0.9876	0.9817-0.9935
Hosmer-Lemeshow-Anpassungstest: 17.1 ($p = 0.03$)						
Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.0003$; Adjusted-Likelihood-Ratio $R^2 = 0.002$, $c = 0.53$						
^a bezogen auf 50 Fälle						

Abbildung 3 zeigt den aus dem Modell geschätzten Risikoverlauf mit 95%-Konfidenzband. Der Regressionskoeffizient der Fallzahl ist nahe bei Null mit negativem Vorzeichen, d.h. das Risiko fällt monoton mit steigender Fallzahl leicht ab. Dies deckt sich mit den Ergebnissen eines zwar statistisch signifikanten Zusammenhangs zwischen Fallzahl und Risiko, aber einem sehr geringen Erklärungswert der Fallzahl.

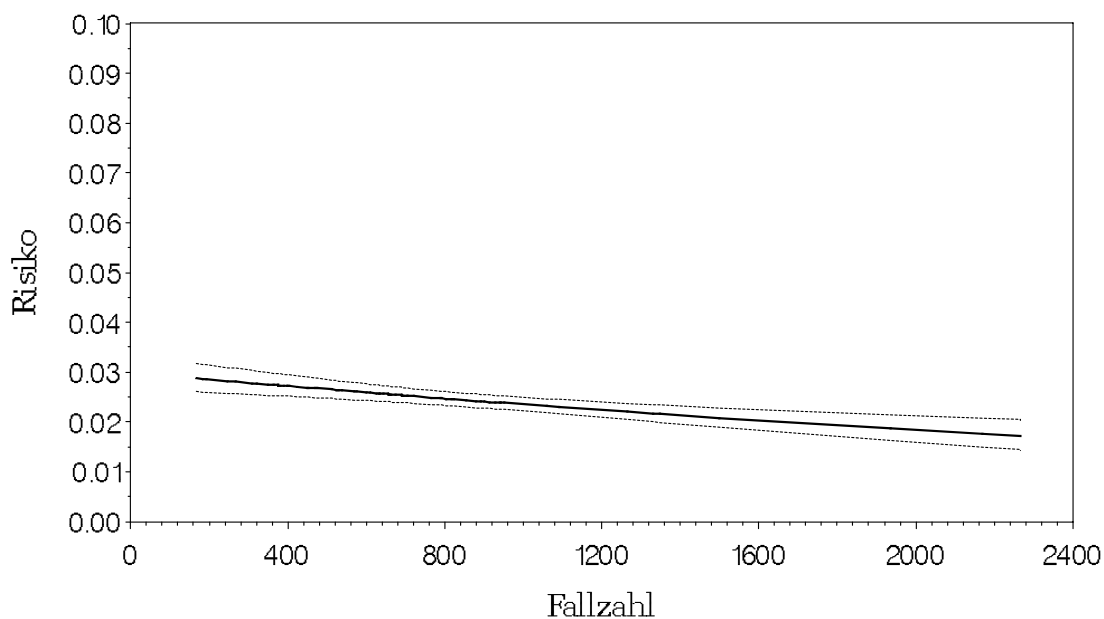


Abb. 3: Rohe Risikokurve mit 95%-Konfidenzband für "Letalität" aus Modell I (2004)

Zur Untersuchung eines möglichen Einflusses potentieller Risikofaktoren wurde in einem nächsten Schritt eine entsprechende Adjustierung im Rahmen eines multiplen logistischen Regressionsmodells vorgenommen. Von den 49227 elektiven und dringlichen Eingriffen enthielten 42674 Datensätze (86.7%) vollständige Angaben zu den relevanten Kovariablen. Hiermit wurde eine Complete-Case-Analyse durchgeführt. Zur Adjustierung wurde zunächst der Euro-Score verwendet (siehe Abschnitt 4.3.2). Die Ergebnisse des multifaktoriellen Modells sind in Tabelle 5 dargestellt (Modell II).

Tab. 5: Multiples logistisches Regressionsmodell für "Letalität" (Modell II, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-4.2496	0.2493	< 0.0001	-	-
Fallzahl	1	0.00002	0.00007	0.7680	1.001 ^a	0.995-1.007
Euro-Score	1	0.0659	0.00271	< 0.0001	1.068	1.062-1.074
BMI	1	-0.0281	0.00841	0.0008	0.972	0.956-0.988
CDC-Klasse	1	0.7532	0.0542	< 0.0001	2.124	1.910-2.362
Hosmer-Lemeshow-Anpassungstest: 67.7 ($p < 0.0001$)						
Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.025$; Adjusted-Likelihood-Ratio $R^2 = 0.092$, $c = 0.74$						
^a bezogen auf 50 Fälle						

Durch Berücksichtigung der Risikofaktoren verbessert sich die erklärte Variabilität deutlich, ist aber immer noch gering. Der Hosmer-Lemeshow-Test ($p < 0.0001$) weist jedoch immer noch auf Abweichungen des Modells von den Daten hin. Die betrachteten Kovariablen weisen alle einen signifikanten Einfluss ($p < 0.0001$) auf. Bei Berücksichtigung der relevanten Kovariablen zeigt die Fallzahl keinen statistisch signifikanten Einfluss mehr ($p = 0.77$).

Die Risikoplots für die Kovariablen BMI und Euro-Score weisen aber auch auf einen nicht-linearen Zusammenhang mit der Fallzahl hin (Abbildungen nicht gezeigt). Um zu überprüfen, ob eine geeignete Transformation dieser Kovariablen zu einer Verbesserung der Modellanpassung führt, wurde die FP-Methode angewendet. Für die Kovariablen BMI und Euro-Score wurden die jeweils besten FPs im Rahmen eines logistischen Regressionsmodells unter Berücksichtigung der Fallzahl ermittelt. Für den BMI ergab sich ein FP 2. Ordnung mit $FP1(\text{BMI}) = \log(\text{BMI})$ und $FP2(\text{BMI}) = \sqrt{\text{BMI}}$. Für den Euro-Score ergab sich ein FP 1. Ordnung mit $FP1(\text{Euro-Score}) = \log(\text{Euro-Score})$. Für die Fallzahl selbst zeigte sich der lineare Term als am besten geeignet. Die Ergebnisse sind in Tabelle 6 dargestellt. Die Modellanpassung bzgl. der erklärten Variabilität und dem Hosmer-Lemeshow-Test ist gegenüber Modell II deutlich verbessert. Der nicht-signifikante Einfluss der Fallzahl auf das Letalitätsrisiko ($p = 0.72$) wird durch dieses Modell bestätigt.

Tab. 6: Multiples logistisches Regressionsmodell für "Letalität" mit FPs
(Modell III, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	15.7832	3.9468	< 0.0001	-	-
Fallzahl	1	-0.00002	0.00007	0.7245	0.999 ^a	0.993-1.005
log(Euro-Score) [FP1]	1	0.9531	0.0356	< 0.0001	2.594	2.419-2.781
log(BMI) [FP1]	1	-15.6107	2.9657	< 0.0001	-	-
$\sqrt{\text{BMI}}$ [FP2]	1	5.6987	1.1256	< 0.0001	-	-
CDC-Klasse	1	0.7015	0.0550	< 0.0001	2.017	1.811-2.246
Hosmer-Lemeshow-Anpassungstest: 10.9 ($p=0.20$)						
Erklärte Variabilität: Squared-Pearson-Correlation $r^2=0.037$; Adjusted-Likelihood-Ratio $R^2=0.126$, $c=0.78$						
^a bezogen auf 50 Fälle						

In einem nächsten Schritt wurde das Modell mit FPs (Modell III) um Wechselwirkungen 1. Ordnung erweitert und dann ein Backward-Eliminationsverfahren mit Einschlussgrenze $p=0.01$ angewendet. Keine der betrachteten Wechselwirkungen erwies sich als signifikant.

Zur Überprüfung von Clustereffekten wurde schließlich ein GEE-Modell basierend auf Modell III angepasst. Tabelle 7 zeigt die Ergebnisse des GEE-Ansatzes (Modell IV). Die Korrelation innerhalb der Krankenhäuser ist sehr gering ($r=0.003$), die Parameterschätzer verändern sich kaum im Vergleich zu Modell III, der p -Wert für die Fallzahl beträgt $p=0.97$.

Tab. 7: Ergebnisse des GEE-Ansatzes für "Letalität" (Modell IV, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	15.7193	3.9016	< 0.0001	-	-
Fallzahl	1	-0.000004	0.0001	0.9734	0.9998 ^a	0.9901-1.0097
log(Euro-Score) [FP1]	1	0.9414	0.0469	< 0.0001	2.5636	2.3384-2.8104
log(BMI) [FP1]	1	-15.5653	2.8651	< 0.0001	-	-
$\sqrt{\text{BMI}}$ [FP2]	1	5.6716	1.0685	< 0.0001	-	-
CDC-Klasse	1	0.7739	0.0691	< 0.0001	2.1682	1.8936-2.4827
Exchangeable Kovarianzstruktur: $r=0.003$						
Erklärte Variabilität: Squared-Pearson-Correlation $r^2=0.038$						
^a bezogen auf 50 Fälle						

Zur Beschreibung des Zusammenhangs zwischen Fallzahl und dem Indikator "Letalität" wählen wir somit Modell III. Die auf Modell III basierende adjustierte Risikofunktion ist in Abbildung 4 dargestellt.

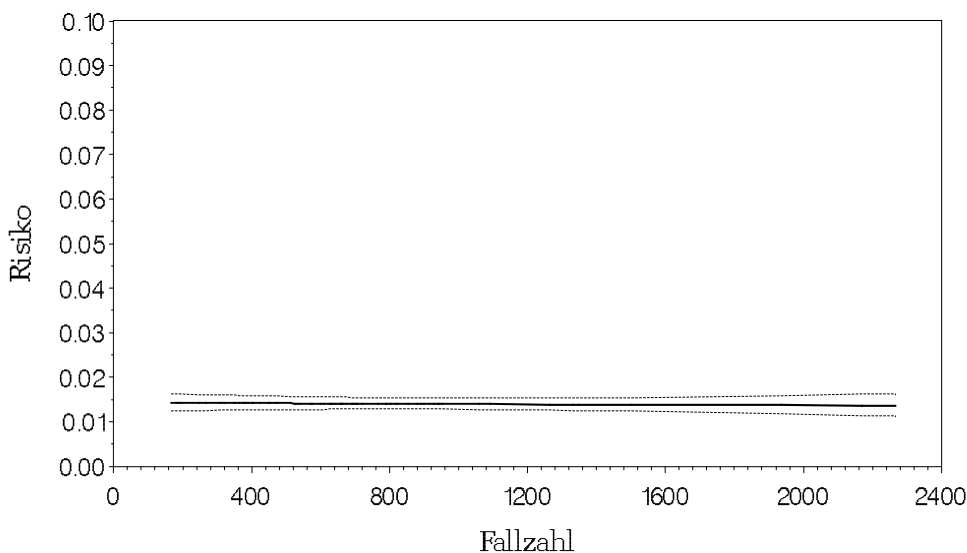


Abb. 4: Adjustierte Risikokurve mit 95%-Konfidenzband für "Letalität" aus Modell III (2004) (Kovariablen auf Mittelwert gesetzt)

Abschließend wurde die Robustheit der Ergebnisse mit Hilfe von Sensitivitätsanalysen überprüft. Hierzu wurden zunächst in Modell III zum einen der Euro-Score durch den KCH-Score ersetzt und zum anderen eine Modellierung der CDC-Klassen als kategorielle Variable vorgenommen. Die resultierenden Ergebnisse entsprachen den Resultaten von Modell III. Es gab keine relevante Veränderung des Regressionskoeffizienten der Fallzahl sowie des zugehörigen p -Werts ($p=0.73$) im Vergleich zu Modell III. Auch die Berücksichtigung der Dringlichkeit als zusätzliche Kovariable ergab keine relevante Veränderung der Ergebnisse.

Von den 49227 Datensätzen wiesen 13.3% der Daten fehlende Werte bei den Kovariablen auf. Um eine mögliche Verzerrung der Ergebnisse der Complete-Case-Analyse abschätzen zu können, wurden das Letalitätsrisiko und die Verteilung der Fallzahl getrennt für die Patienten mit und ohne fehlende Werte bei den Kovariablen berechnet. Für Patienten ohne fehlende Werte betrug das Letalitätsrisiko 2,3%, die mediane Fallzahl war 803 (25%- und 75%-Quartil: 632

und 1060), für Patienten mit fehlenden Kovariablenwerten waren die entsprechenden Werte 3.2% und 765 (632 und 1090). Die Daten liefern somit keinen Hinweis auf einen gravierenden Unterschied zwischen den Patientengruppen mit und ohne fehlende Kovariablenwerte.

Schließlich wurden regressionsdiagnostische Verfahren zur Identifikation extremer Werte eingesetzt, die stark vom angepassten Modell abweichen, und eine neue Modellanpassung ohne diese Extremwerte durchgeführt. Auch hier ergaben sich keine Veränderungen der Ergebnisse.

5.2.2 Zusammenhang zwischen Fallzahl und Risiko (2003)

Das mittlere Risiko für den Indikator „Letalität“ beträgt für die $n=39729$ Datensätze aus dem Jahr 2003 2.5%. Abbildung 5 zeigt den Risikoplot in Abhängigkeit von der Fallzahl.

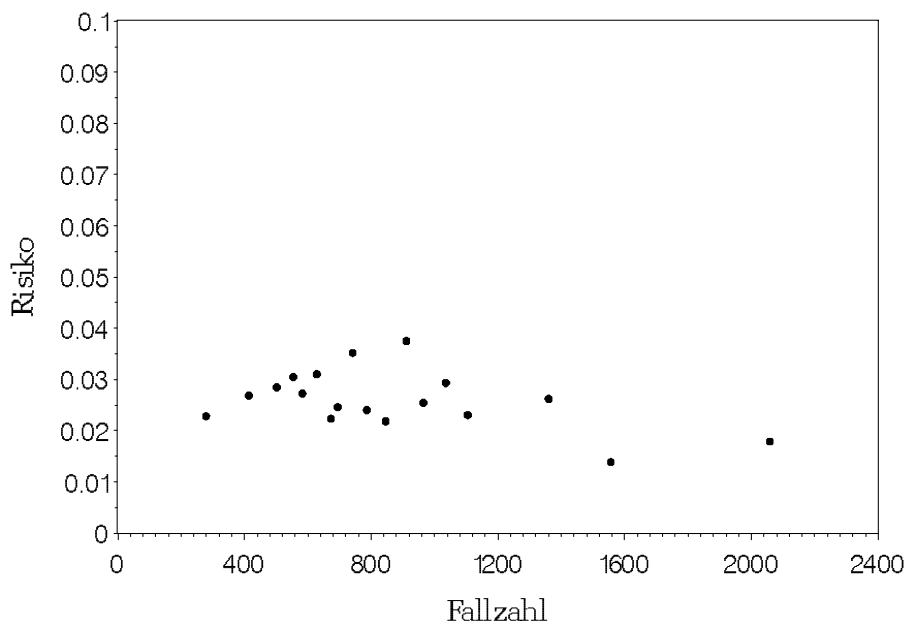


Abb. 5: Risikoplot für "Letalität" vs. Fallzahl (2003)

Zur Untersuchung der Ergebnisstabilität wurden die anhand der Daten aus dem Jahr 2004 hergeleiteten Modelle auf die 2003er Daten angewendet. Es werden hier nur die Ergebnisse des multiplen logistischen Regressionsmodells mit FPs dargestellt (Modell III, Tabelle 8).

Tab. 8: Multiples logistisches Regressionsmodell für "Letalität" mit FPs
(Modell III, 2003)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	12.9633	6.2137	0.0370	-	-
Fallzahl	1	-0.00018	0.0001	0.1146	0.991 ^a	0.981-1.001
log(Euro-Score) [FP1]	1	0.9070	0.0527	< 0.0001	2.477	2.234-2.747
log(BMI) [FP1]	1	-14.4382	4.6158	0.0018	-	-
$\sqrt{\text{BMI}}$ [FP2]	1	5.5561	1.7393	0.0014	-	-
CDC-Klasse	1	0.7610	0.0804	< 0.0001	2.140	1.829-2.506
Hosmer-Lemeshow-Anpassungstest: 7.3 ($p=0.50$)						
Erklärte Variabilität: Squared-Pearson-Correlation $r^2=0.039$; Adjusted-Likelihood-Ratio $R^2=0.124$, $c=0.77$						
^a bezogen auf 50 Fälle						

Das Modell liefert eine gute Anpassung an die Daten. Auch bei den Daten aus dem Jahr 2003 zeigt die Fallzahl nach Adjustierung keinen signifikanten Einfluss auf das Letalitätsrisiko ($p=0.11$).

5.2.3 Schlussfolgerung

Die Ergebnisse der multifaktoriellen Modellierung zeigen eindeutig, dass sich nach Adjustierung bezüglich relevanter Kovariablen und unter Berücksichtigung möglicher nicht-linearer Zusammenhänge kein signifikanter Einfluss der Fallzahl auf das Letalitätsrisiko nachweisen lässt. Somit wurde für den primären Qualitätsindikator "Letalität" keine Schwellenwertberechnung vorgenommen.

5.3 Ergebnisse für den Indikator "Infektion"

5.3.1 Zusammenhang zwischen Fallzahl und Risiko (2004)

Zur Untersuchung des Zusammenhangs zwischen Fallzahl und "Infektion" wurden Notfall-Operationen sowie Patienten mit präoperativer Mediastinitis ausgeschlossen, so dass n=49163 Datensätze in die Auswertung eingingen. Das mittlere Infektions-Risiko beträgt 0.76%. Abbildung 6 zeigt den Risikoplot für den Indikator "Infektion" in Abhängigkeit von der Fallzahl.

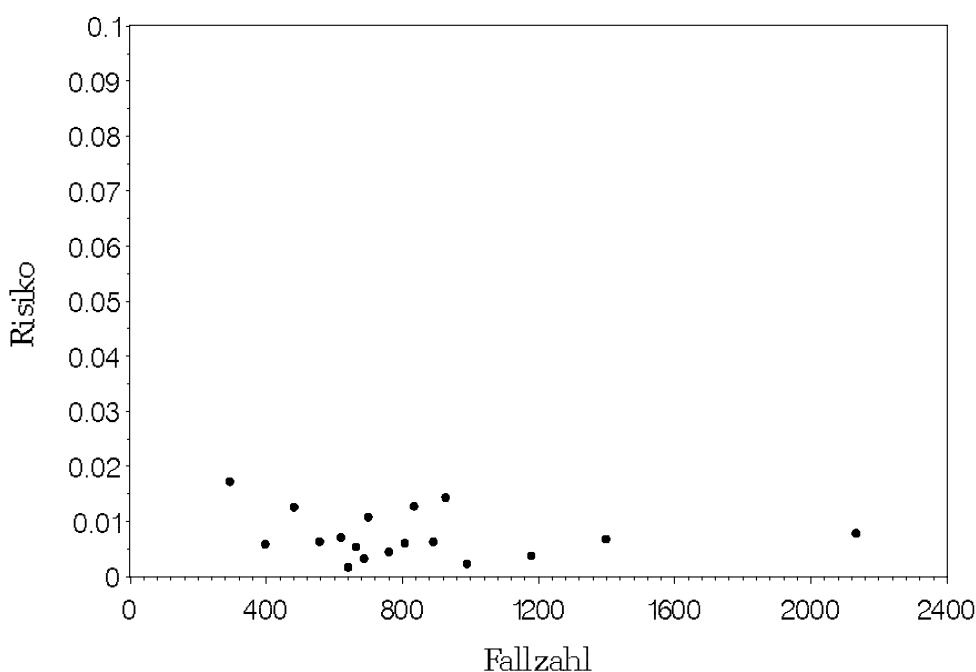


Abb. 6: Risikoplot für "Infektion" vs. Fallzahl (2004)

Zunächst wurde ein einfaches logistisches Regressionsmodell mit der Fallzahl als Einflussfaktor berechnet. Die Ergebnisse sind in Tabelle 9 dargestellt (Modell I).

Tab. 9: Einfaches logistisches Regressionsmodell für "Infektion" (Modell I, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-4.7076	0.1069	< 0.0001	-	-
Fallzahl	1	-0.00018	0.0001	0.0937	0.9910 ^a	0.9814-1.0008
Hosmer-Lemeshow-Anpassungstest: 48.0 ($p < 0.0001$) Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.0001$; Adjusted-Likelihood-Ratio $R^2 = 0.0007$, $c = 0.53$						
^a bezogen auf 50 Fälle						

Abbildung 7 zeigt den aus dem Modell geschätzten Risikoverlauf mit 95%-Konfidenzband. Die Risikokurve zeigt einen mit steigender Fallzahl minimal fallenden Verlauf. Ein signifikanter Einfluss der Fallzahl auf das Infektions-Risiko ist jedoch nicht nachzuweisen ($p=0.09$).

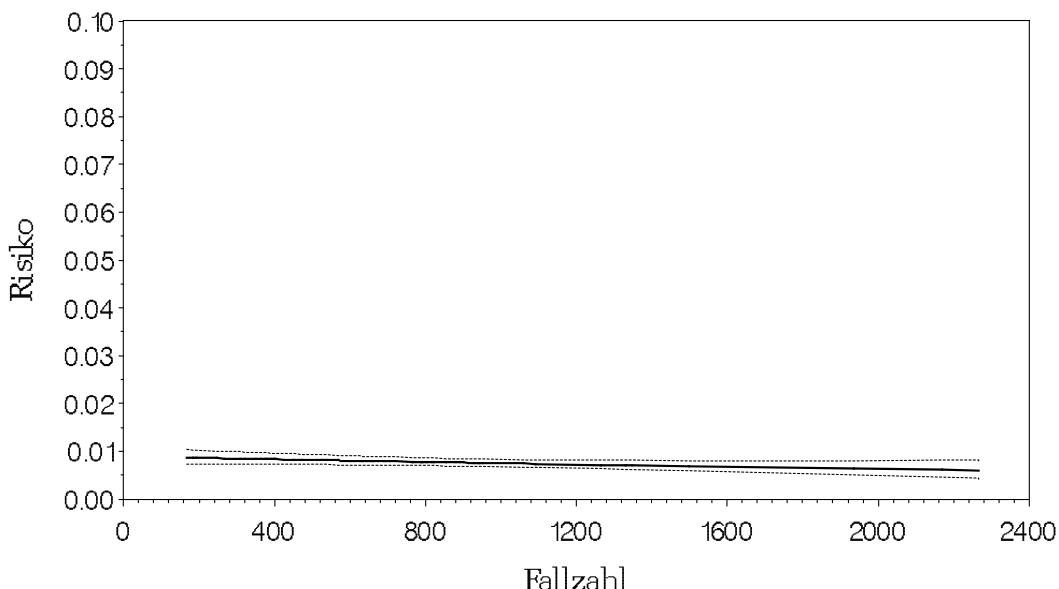


Abb. 7: Rohe Risikokurve mit 95%-Konfidenzband für "Infektion" aus Modell I (2004)

In nächsten Schritt wurde zur Adjustierung bzgl. potentieller Risikofaktoren ein multiples logistisches Regressionsmodell angepasst. Wie bei der Analyse der Letalität wurde eine Complete-Case-Analyse basierend auf 42616 Datensätzen (86.7 %) mit vollständigen Angaben zu den relevanten Kovariablen durchgeführt. Die Ergebnisse des multiplen Modells sind in Tabelle 10 dargestellt (Modell II).

Tab. 10: Multiples logistisches Regressionsmodell für "Infektion" (Modell II, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-7.8710	0.3559	< 0.0001	-	-
Fallzahl	1	-0.00012	0.0001	0.2743	0.994 ^a	0.984-1.004
Euro-Score	1	0.0398	0.0051	< 0.0001	1.041	1.030-1.051
BMI	1	0.0935	0.0110	< 0.0001	1.098	1.075-1.122
CDC-Klasse	1	0.2504	0.0867	0.0039	1.285	1.084-1.522
Hosmer-Lemeshow-Anpassungstest: 3.3 ($p=0.92$)						
Erklärte Variabilität: Squared-Pearson-Correlation $r^2=0.003$; Adjusted-Likelihood-Ratio $R^2=0.03$, $c=0.66$						
^a bezogen auf 50 Fälle						

Durch Berücksichtigung der Risikofaktoren erhöht sich die erklärte Variabilität, die allerdings immer noch gering ist. Der Hosmer-Lemeshow-Test weist auf eine sehr gute Modellanpassung hin ($p=0.92$). Alle betrachteten Kovariablen zeigen mit Ausnahme der Fallzahl ($p=0.27$) einen signifikanten Einfluss auf das Infektionsrisiko.

Die Modellanpassung des Modells mit Haupteffekten (Modell II) ist sehr gut. Dementsprechend führte die Modellierung mit Hilfe von FPs zu keinen großen Veränderungen. Lediglich für den Euro-Score ergab sich ein FP 1. Ordnung mit $FP1(\text{Euro-Score}) = (\text{Euro-Score})^{-0.5}$. Die Ergebnisse sind in Tabelle 11 dargestellt. Der nicht-signifikante Einfluss der Fallzahl auf das Letalitätsrisiko wird auch durch dieses Modell bestätigt ($p=0.18$).

Tab. 11: Multiples logistisches Regressionsmodell für "Infektion" mit FPs (Modell III, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-6.4843	0.3693	< 0.0001	-	-
Fallzahl	1	-0.00014	0.0001	0.1835	0.993 ^a	0.983-1.003
(Euro-Score) ^{-0.5} [FP1]	1	-2.2092	0.2686	< 0.0001	0.110	0.065-0.186
BMI	1	0.0971	0.0110	< 0.0001	1.102	1.078-1.126
CDC-Klasse	1	0.2270	0.0873	0.0093	1.255	1.057-1.489
Hosmer-Lemeshow-Anpassungstest: 3.6 ($p=0.89$)						
Erklärte Variabilität: Squared-Pearson-Correlation $r^2=0.004$; Adjusted-Likelihood-Ratio $R^2=0.039$, $c=0.67$						
^a bezogen auf 50 Fälle						

In nächsten Schritt wurde das Modell III um Wechselwirkungen 1. Ordnung erweitert und dann ein Backward-Eliminationsverfahren mit Einschlussgrenze $p=0.01$ angewendet. Keine Wechselwirkung erwies sich als signifikant.

Schließlich wurde zur Überprüfung von Clustereffekten ein GEE-Modell basierend auf Modell III angepasst. Tabelle 12 zeigt die Ergebnisse des GEE-Ansatzes (Modell IV). Die Korrelation innerhalb der Krankenhäuser ist sehr gering ($r=0.003$). Die Fallzahl ist auch hier die einzige nicht-signifikante Kovariable ($p=0.13$). Es zeigen sich keine relevanten Veränderungen in den Parameterschätzern im Vergleich zu Modell III, weshalb wir Modell III als endgültiges Modell wählen.

Tab. 12: Ergebnisse des GEE-Ansatzes für "Infektion" (Modell IV, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-6.2366	0.3607	< 0.0001	-	-
Fallzahl	1	-0.0004	0.0003	0.1278	0.9802 ^a	0.9518-1.0094
(Euro-Score) ^{-0.5} [FP1]	1	-2.1538	0.2716	< 0.0001	0.1160	0.0681-0.1976
BMI	1	0.0943	0.0096	< 0.0001	1.0989	1.0784-1.1198
CDC-Klasse	1	0.2292	0.0822	0.0053	1.2576	1.0705-1.4774
Exchangeable Kovarianzstruktur: r=0.003 Erklärte Variabilität: Squared-Pearson-Correlation r ² =0.004						
^a bezogen auf 50 Fälle						

Die auf Modell III basierende adjustierte Risikofunktion für den Indikator "Infektion" ist in Abbildung 8 dargestellt.

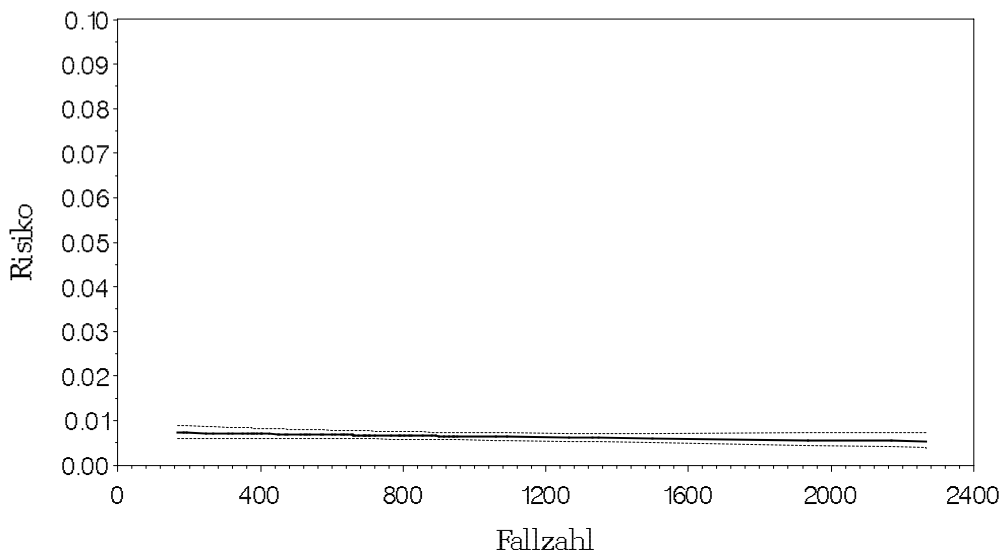


Abb. 8: Adjustierte Risikokurve mit 95%-Konfidenzband für "Infektion" aus Modell III (2004) (Kovariablen auf Mittelwert gesetzt)

Die Robustheit der Ergebnisse wurde mit Hilfe von Sensitivitätsanalysen überprüft. Zunächst wurden in Modell III zum einen der Euro-Score durch den KCH-Score ersetzt und zum anderen eine Modellierung der CDC-Klassen als kategorielle Variable vorgenommen. Die resultierenden Ergebnisse entsprachen den Resultaten von Modell III. Es gab keine relevanten Veränderungen des Regressionskoeffizienten der Fallzahl sowie des zugehörigen p-Werts ($p=0.17$). Auch die Berücksichtigung der Dringlichkeit als zusätzliche Kovariable ergab keine relevante Veränderung der Ergebnisse.

Von den 49163 Datensätzen wiesen 13.3% der Daten fehlende Werte bei den Kovariablen auf. Ein Vergleich des Infektionsrisikos und der Verteilung der Fallzahl bei Patienten mit und ohne fehlende Werte bei den Kovariablen ergab ein Infektionsrisiko von 0.80% und eine mediane Fallzahl von 803 (25%- und 75%-Quartil: 632 und 1060) bei Patienten ohne fehlende Werte und ein Risiko von 0.53% und eine mediane Fallzahl von 765 (25%- und 75%-Quartil: 632 und 1090) für Patienten mit fehlenden Kovariablenwerten. Somit ergibt sich kein Anhaltspunkt für einen relevanten Unterschied zwischen den Patientengruppen mit und ohne fehlende Kovariablenwerte. Eine Ausreißeranalyse mit Hilfe regressionsdiagnostischer Verfahren führte zu keiner Veränderung der Resultate.

5.3.2 Zusammenhang zwischen Fallzahl und Risiko (2003)

Zur Validierung der Ergebnisse wurden die Daten aus dem Jahr 2003 ausgewertet. Das mittlere Risiko beträgt 0.79%. Der Risikoplot für den Indikator "Infektion" in Abhängigkeit von der Fallzahl ist in Abbildung 9 dargestellt.

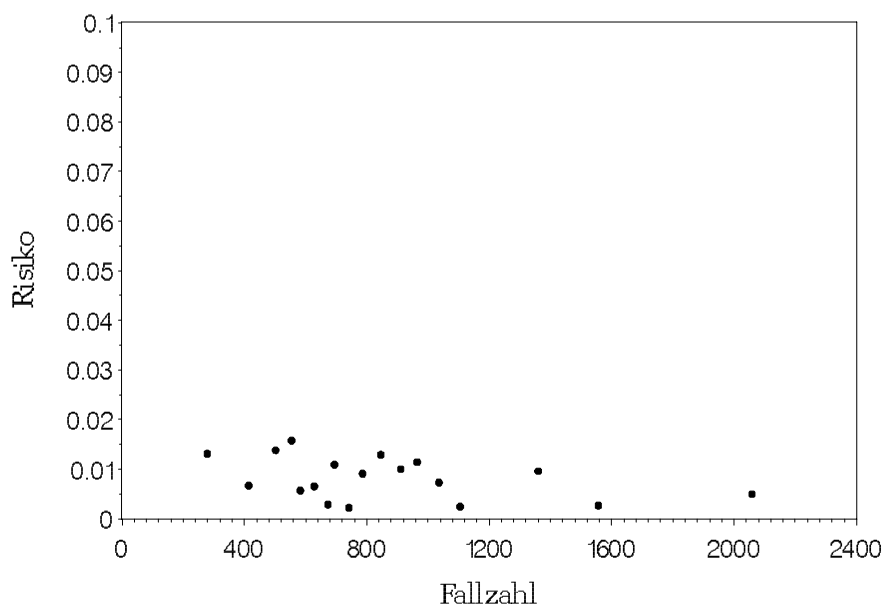


Abb. 9: Risikoplot für "Infektion" vs. Fallzahl (2003)

Die Ergebnisse des multiplen Modells mit FPs (Modell III) sind in Tabelle 13 zusammengefasst.

Tab. 13: Multiples logistisches Regressionsmodell für "Infektion" mit FPs
(Modell III, 2003)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-6.6216	0.5788	< 0.0001	-	-
Fallzahl	1	-0.00038	0.0002	0.0463	0.981 ^a	0.962-1.001
(Euro-Score) ^{-0.5} [FP1]	1	-0.7046	0.3560	0.0478	0.494	0.246-0.993
BMI	1	0.0758	0.0178	< 0.0001	1.079	1.042-1.117
CDC-Klasse	1	0.4921	0.1253	< 0.0001	1.636	1.280-2.091
Hosmer-Lemeshow-Anpassungstest: 5.9 ($p=0.66$)						
Erklärte Variabilität: Squared-Pearson-Correlation $r^2=0.003$; Adjusted-Likelihood-Ratio $R^2=0.026$, $c=0.65$						
^a bezogen auf 50 Fälle						

Der Regressionskoeffizient der Fallzahl ist negativ und betragsmäßig etwas größer als bei den Daten aus dem Jahr 2004 (leichter Abfall des Risikos mit steigender Fallzahl), aber immer noch sehr nahe bei Null. Der p -Wert ist mit $p=0.046$ knapp unterhalb des 5%-Signifikanz-Niveaus. Im entsprechenden GEE-Ansatz ist die Fallzahl jedoch nicht mehr signifikant ($p=0.11$), so dass wir auch bei den Daten von 2003 keinen eindeutigen signifikanten Effekt der Fallzahl auf das Infektionsrisiko feststellen können.

5.3.3 Schlussfolgerung

Die Ergebnisse der multifaktoriellen Modellierung zeigen eindeutig, dass sich nach Adjustierung bezüglich relevanter Kovariablen und unter Berücksichtigung möglicher nicht-linearer Zusammenhänge kein signifikanter Einfluss der Fallzahl auf das Infektionsrisiko nachweisen lässt. Somit wurde auch für den sekundären Qualitätsindikator "Infektion" keine Schwellenwertberechnung vorgenommen.

5.4 Ergebnisse für den Indikator "Brustwandarterie"

5.4.1 Zusammenhang zwischen Fallzahl und Risiko (2004)

Wie in Abschnitt 4.3.1 beschrieben, wurde der Indikator "Brustwandarterie" als "keine Verwendung der inneren Brustwandarterie ja/nein" kodiert. Notfall-OPs und Patienten mit vorausgegangener Koronarchirurgie wurden ausgeschlossen, so dass 47272 Datensätze verblieben. Das mittlere Risiko für den Indikator "Brustwandarterie" beträgt 10.5%. In Abbildung 10 ist der Risikoplot dargestellt.

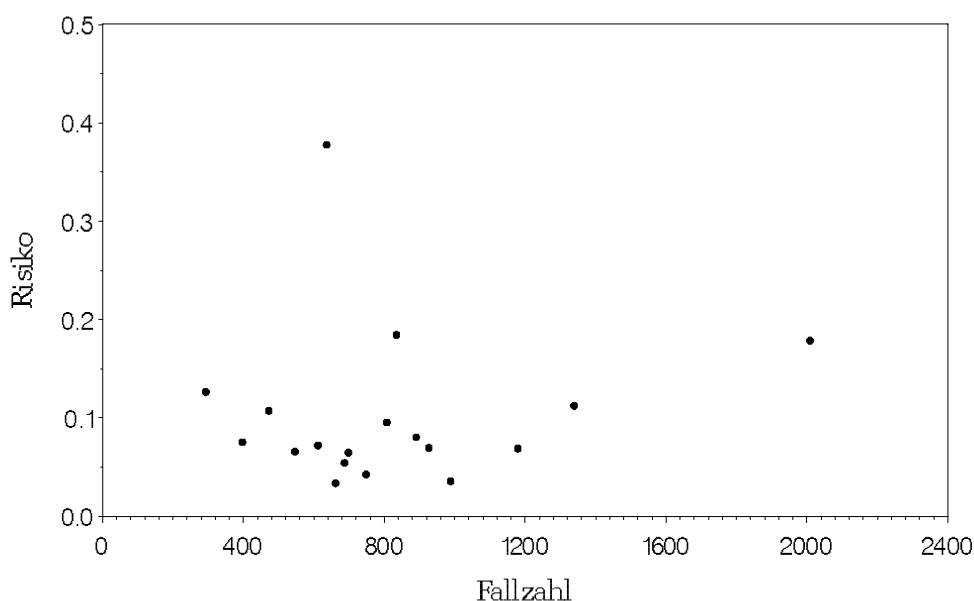


Abb. 10: Risikoplot für "Brustwandarterie" vs. Fallzahl (2004)

Der Risikoplot zeigt einen nicht-monotonen Risikoverlauf mit hoher Variabilität. Ein einfaches logistisches Regressionsmodell, das nur den linearen Term der Fallzahl berücksichtigt, liefert dementsprechend eine sehr schlechte Modellanpassung (Hosmer-Lemshow-Test: $p < 0.0001$) und kann für eine adäquate Datenbeschreibung nicht verwendet werden.

Zur Modellierung wurden daher FPs betrachtet. Die beste Anpassung lieferte das Polynom 2. Grades mit $FP1 = \text{Fallzahl}^2$ und $FP2 = \text{Fallzahl}^3$. Die Ergebnisse dieses Modells (Modell I) sind in Tabelle 14 aufgeführt. Die Modellanpassung ist besser als im einfachen logistischen Regressionsmodell nur mit linearem Haupteffekt, aber gemäß Hosmer-Lemeshow-Test immer noch deutlich signifikant abzulehnen.

Tab. 14: Logistisches Regressionsmodell für "Brustwandarterie" mit FP für Fallzahl (Modell I, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-2.0063	0.0343	< 0.0001	-	-
Fallzahl ² [FP1]	1	-9.17E-7	7.897E-8	< 0.0001	siehe Tab. 17	
Fallzahl ³ [FP2]	1	4.9E-10	3.38E-11	< 0.0001		
Hosmer-Lemeshow-Anpassungstest: 856.6 ($p < 0.0001$)						
Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.02$; Adjusted-Likelihood-Ratio $R^2 = 0.03$, $c = 0.59$						

In nächsten Schritt wurde eine Adjustierung bezüglich potentieller Risikofaktoren im Rahmen eines multiplen logistischen Regressionsmodells vorgenommen. Von den 47272 elektiven und dringlichen Eingriffen enthielten 41015 Datensätze (86.8%) vollständige Angaben zu den relevanten Kovariablen. Hiermit wurde eine Complete-Case-Analyse durchgeführt. Das Selektionsverfahren für die Anpassung von FPs führte zu folgenden FPs: $FP1(BMI) = BMI^{-0.5}$ und $FP2(BMI) = \log(BMI)$ sowie $FP1(Euro-Score) = \log(Euro-Score)$. Die Ergebnisse dieses Modellansatzes (Modell II) sind in Tabelle 15 dargestellt.

Tab. 15: Multiples logistisches Regressionsmodell für "Brustwandarterie" mit FPs für Fallzahl und Kovariablen (Modell II, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-57.1263	9.2247	< 0.0001		
Fallzahl ² [FP1]	1	-7.93E-7	8.493E-8	< 0.0001	siehe Tab. 17	
Fallzahl ³ [FP2]	1	4.3E-10	3.64E-11	< 0.0001		
$\log(Euro-Score)$ [FP1]	1	0.4346	0.0190	< 0.0001	1.544	1.488-1.603
$BMI^{-0.5}$ [FP1]	1	106.5	18.25	< 0.0001	-	-
$\log(BMI)$ [FP2]	1	10.31	1.7321	< 0.0001	-	-
CDC-Klasse	1	0.0085	0.0266	0.7495	1.009	0.957-1.063
Hosmer-Lemeshow-Anpassungstest: 10.2 ($p = 0.25$)						
Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.034$; Adjusted-Likelihood-Ratio $R^2 = 0.056$, $c = 0.64$						

Für dieses modifizierte Modell zeigt der Hosmer-Lemeshow-Test keine bedeutsame Abweichung des Modells von den Daten ($p = 0.25$). Alle betrachteten Kovariablen mit Ausnahme der CDC-Klasse sind statistisch signifikant (Tab. 15).

Die Betrachtung von Wechselwirkungen 1. Ordnung führte zu keiner weiteren Verbesserung der Modellanpassung. Zur Überprüfung von Cluster-Effekten wurde schließlich ein GEE-Modell basierend auf Modell II angepasst. Tabelle 16 zeigt die Ergebnisse des GEE-Ansatzes (Modell III). Die Korrelation innerhalb der Krankenhäuser beträgt $r=0.12$, so dass hier ein deutlicher Cluster-Effekt zu beobachten ist und wir Modell III als endgültiges Modell verwenden.

Tab. 16: Ergebnisse des GEE-Ansatzes für "Brustwandarterie" (Modell III, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-60.3754	9.4355	< 0.0001		
Fallzahl ² [FP1]	1	-3.6E-7	9.25E-7	0.6937	siehe Tab. 17	
Fallzahl ³ [FP2]	1	3.4E-10	4.28E-10	0.4312		
log(Euro-Score) [FP1]	1	0.4393	0.0522	< 0.0001	1.552	1.401-1.719
BMI ^{-0.5} [FP1]	1	113.45	18.2547	< 0.0001	-	-
log(BMI) [FP2]	1	10.7979	1.7772	< 0.0001	-	-
CDC-Klasse	1	-0.0360	0.0553	0.5149	0.965	0.866-1.075
Exchangeable Kovarianzstruktur: $r=0.12$ Erklärte Variabilität: Squared-Pearson-Correlation $r^2=0.04$						

Die Berücksichtigung des Cluster-Effekts führt dazu, dass der Einfluss der Fallzahl nicht länger statistisch signifikant ist. Dies spiegelt sich auch in den zugehörigen Odds Ratios wieder, die nach Adjustierung bzgl. des Cluster-Effekts mit einer sehr großen Unsicherheit behaftet sind (Tab. 17).

Tab. 17: Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-III für "Brustwandarterie" (2004)

Intervall	Referenzwert	OR ^a (95%-KI)		
		Modell I	Modell II	Modell III
1-632	316	1.328 (1.254-1.407)	1.273 (1.196-1.355)	1.031 (0.531-2.004)
633-1060	846	1.000	1.000	1.000
1061-2266	1663	1.078 (1.013-1.147)	1.098 (1.027-1.173)	1.822 (0.663-5.008)
^a bezogen auf Differenz der Intervallmittelpunkte				

Abschließend wurde die Robustheit der Ergebnisse mit Hilfe von Sensitivitätsanalysen überprüft. Hierzu wurden zunächst in Modell III der Euro-Score durch den KCH-Score ersetzt und die CDC-Klasse als kategorielle Variable modelliert. Die Resultate waren vergleichbar mit denen von Modell III. Auch die Berücksichtigung der Dringlichkeit als zusätzliche Kovariable ergab keine relevante Veränderung der Ergebnisse.

Der Risikoplot in Abb. 10 zeigt einen Punkt mit deutlich abweichendem Risiko bei einer Fallzahl von etwa 600. Dieser Punkt basiert auf einer Klinik mit auffallend hohem Anteil an Eingriffen ohne Verwendung der inneren Brustwandarterie. Um den Einfluss dieser extremen Daten zu untersuchen, wurden die Auswertungen erneut ohne diese Daten durchgeführt. Dieses führte zu unterschiedlichen FPs für Fallzahl und BMI. Die Modellanpassung des multiplen logistischen Regressionsmodells (Modell III) verbesserte sich ohne Berücksichtigung dieser Ausreißer (Hosmer-Lemeshow-Test: $p=0.45$), an den Ergebnissen bezüglich des Einflusses der Fallzahl änderte sich jedoch nichts. Die Anpassung eines GEE-Modells ohne Ausreißer führte im Vergleich mit Modell IV zu einer niedrigeren Korrelation ($r=0.06$) und einer Reduktion der p -Werte für die FPs der Fallzahl auf $p=0.12$ (FP1) und $p=0.02$ (FP2) sowie zu einer Verkleinerung der Konfidenzintervallbreite ohne jedoch deutliche Signifikanz zu erreichen. Ohne die Daten der Klinik mit auffallend hohem Anteil an Eingriffen ohne Verwendung der inneren Brustwandarterie scheint sich somit ein möglicher U-förmiger Zusammenhang zwischen der Fallzahl und dem Indikator "Brustwandarterie" anzudeuten. Insgesamt verbleibt der Kurvenverlauf jedoch unklar.

Von den 47272 Datensätzen wiesen 13.2% der Daten fehlende Werte bei den Kovariablen auf. Um eine mögliche Verzerrung der Ergebnisse der Complete-Case-Analyse abschätzen zu können, wurden das Risiko für den Indikator "Brustwandarterie" und die Verteilung der Fallzahl getrennt für die Patienten mit und ohne fehlende Werte bei den Kovariablen berechnet. Für Patienten ohne fehlende Werte betrug das Risiko 10.5%, die mediane Fallzahl war 792 (25%- und 75%-Quartil: 632 und 1060), für Patienten mit fehlenden Kovariablenwerten waren die entsprechenden Werte 10.4% und 765 (25%- und 75%-Quartil: 632 und 1090). Die Daten liefern somit keinen Hinweis auf einen relevanten Unterschied zwischen den Patientengruppen mit und ohne fehlende Werte.

Schließlich wurden noch regressionsdiagnostische Verfahren zur Identifikation extremer Werte eingesetzt, die stark vom angepassten Modell abweichen und eine neue Modellanpassung ohne diese Extremwerte durchgeführt. Auch hier ergaben sich keine relevanten Veränderungen der Ergebnisse.

5.4.2 Zusammenhang zwischen Fallzahl und Risiko (2003)

Das mittlere Risiko für den Indikator "Brustwandarterie" beträgt 12.0%. Der Risikoplot in Abhängigkeit von der Fallzahl ist in Abbildung 11 dargestellt. Die Struktur ist vergleichbar mit den Daten aus dem Jahr 2004.

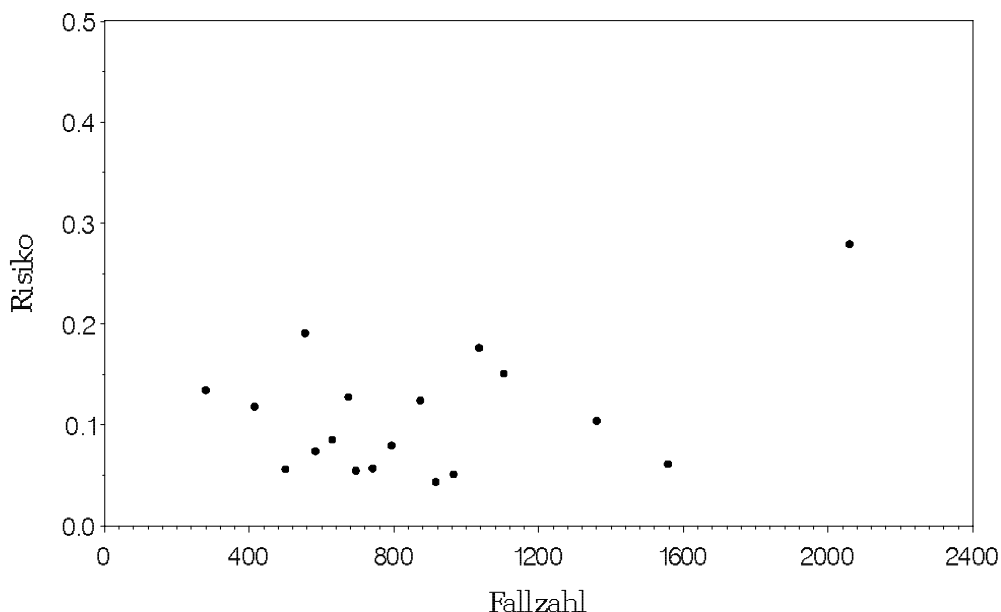


Abb. 11: Risikoplot für "Brustwandarterie" vs. Fallzahl (2003)

Zur Validierung der Ergebnisse wurden die Modelle für die Daten aus dem Jahr 2004 auf die n=37512 Datensätze aus dem Jahr 2003 angewendet. Wir zeigen hier nur die wichtigsten Ergebnisse des GEE-Ansatzes mit FPs (Modell III, Tab. 18) sowie die Odds Ratios für die Fallzahl aus den Modellen I-III (Tab. 19).

Tab. 18: Ergebnisse des GEE-Ansatzes für "Brustwandarterie" (Modell III, 2003)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-43.7337	11.3998	< 0.0001	-	-
Fallzahl ² [FP1]	1	-9.09E-7	9.85E-7	0.3562	siehe Tab. 19	
Fallzahl ³ [FP2]	1	5.21E-10	4.77E-10	0.2744		
log(Euro-Score) [FP1]	1	0.4908	0.0437	< 0.0001	1.634	1.500-1.780
BMI ^{-0.5} [FP1]	1	80.18	22.29	0.0003	-	-
log(BMI) [FP2]	1	7.8191	2.1274	0.0002	-	-
CDC-Klasse	1	-0.0405	0.0485	0.4043	0.960	0.873-1.056
Exchangeable Kovarianzstruktur: r=0.10						
Erklärte Variabilität: Squared-Pearson-Correlation r ² =0.056						

Tab. 19: Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-III für "Brustwandarterie" (2003)

Intervall	Referenzwert	OR ^a (95%-KI)		
		Modell I	Modell II	Modell III
1-598	299	1.158 (1.092-1.228)	1.266 (1.160-1.381)	1.319 (0.646-2.696)
599-1140	869	1.000	1.000	1.000
1141-2243	1692	1.480 (1.400-1.564)	1.407 (1.290-1.534)	1.305 (0.622-2.741)
^a bezogen auf Differenz der Intervallmittelpunkte				

Wie bei den Daten aus 2004 zeigt die Fallzahl bei Berücksichtigung von Kovariablen und Cluster-Effekten keinen signifikanten Einfluss auf den Indikator "Brustwandarterie". Der Verlauf der Risikokurve verbleibt damit unklar.

5.4.3 Schlussfolgerung

Die Ergebnisse der multifaktoriellen Modellierung mit Hilfe des GEE-Ansatzes zeigen eindeutig, dass bezüglich des Zusammenhangs zwischen der Fallzahl und dem Indikator "Brustwandarterie" eine sehr große Variabilität sowie ein starker Cluster-Effekt vorhanden sind. Es scheint sich am ehesten ein U-förmiger Zusammenhang anzudeuten. Bei der vorliegenden geringen Zahl an Krankenhäusern (n=77) ist jedoch der Zusammenhang zwischen der Fallzahl und dem Indikator "Brustwandarterie" von dem Cluster-Effekt überlagert, so dass die Form des Zusammenhangs nicht ermittelt werden kann und unklar bleibt.

5.5 Ergebnisse für den Indikator "Anastomosen"

5.5.1 Zusammenhang zwischen Fallzahl und Risiko (2004)

Notfall-Operationen wurden nicht berücksichtigt, so dass $n=49091$ Datensätze in die Auswertung eingingen. Das mittlere Risiko beträgt 25.2%. Abbildung 12 zeigt den Risikoplot für den Indikator "Anastomosen" in Abhängigkeit von der Fallzahl.

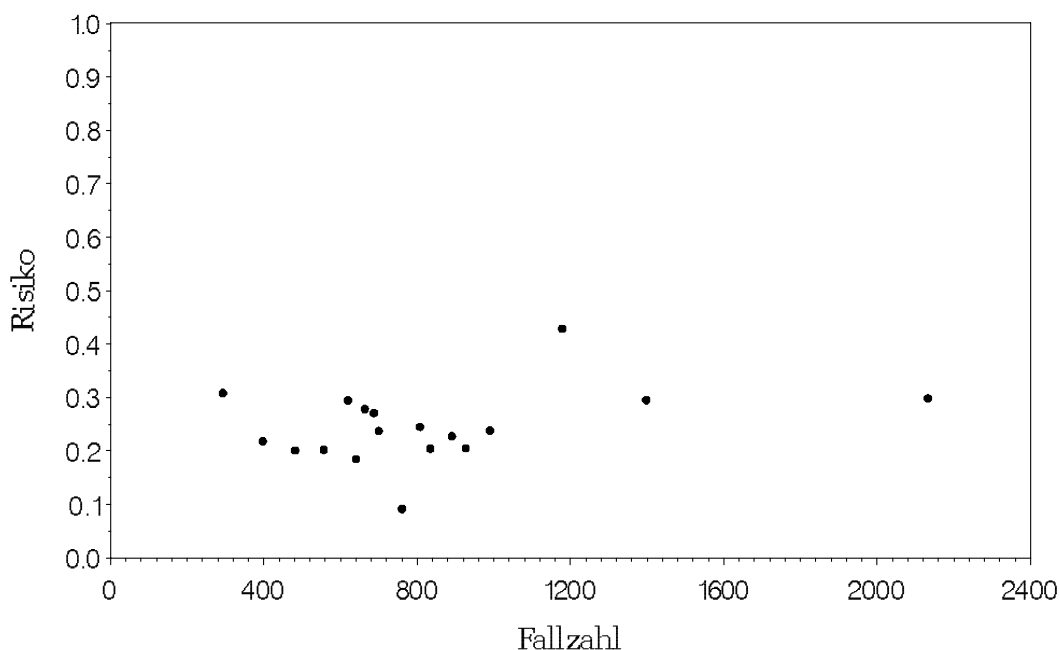


Abb. 12: Risikoplot für "Anastomosen" vs. Fallzahl (2004)

Der Risikoplot weist auf einen nicht-monotonen Risikoverlauf hin. Ein einfaches logistisches Regressionsmodell mit linearem Term zeigt einen leichten Anstieg der Risikokurve, wobei die Modellanpassung nicht adäquat ist (Hosmer-Lemshow-Test: $p < 0.0001$). Es wurde daher wieder die FP-Methode angewendet. Die beste Anpassung lieferte das Polynom 2. Grades mit $FP1 = 1/\text{Fallzahl}$ und $FP2 = (1/\text{Fallzahl}) \times \log(\text{Fallzahl})$. Die Ergebnisse dieses Modells (Modell I) zeigt Tabelle 20. Die Fallzahl ist in diesem Modell statistisch signifikant. Die Modellanpassung gemäß Hosmer-Lemshow-Test ist besser als im einfachen logistischen Regressionsmodell, aber noch immer unbefriedigend.

Tab. 20: Logistisches Regressionsmodell für "Anastomosen" mit FP für Fallzahl (Modell I, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-0.3386	0.0545	< 0.0001	-	-
1/Fallzahl [FP1]	1	2198.2	159.5	< 0.0001	siehe Tab. 23	
(1/Fallzahl)×log(Fallzahl) [FP2]	1	-424.9	30.36	< 0.0001		
Hosmer-Lemeshow-Anpassungstest: 219 ($p < 0.0001$) Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.002$; Adjusted-Likelihood-Ratio $R^2 = 0.01$, $c = 0.58$						

In nächsten Schritt wurde eine Adjustierung bezüglich potentieller Risikofaktoren im Rahmen eines multiplen logistischen Regressionsmodells vorgenommen. Es gingen 42622 elektive und dringliche Eingriffe im Rahmen einer Complete-Case-Analyse in die Auswertungen ein. Das Selektionsverfahren für FPs führte beim Euro-Score zu den folgenden FPs: $FP1(\text{Euro-Score}) = (\text{Euro-Score})^{-0.5}$ und $FP2(\text{Euro-Score}) = \log(\text{Euro-Score})$. Beim BMI lieferte der lineare Term eine adäquate Anpassung. Die Ergebnisse dieses Modellansatzes (Modell II) sind in Tabelle 21 dargestellt. Alle betrachteten Kovariablen sind signifikant.

Tab. 21: Multiples logistisches Regressionsmodell für "Anastomosen" mit FPs für Fallzahl und Kovariablen (Modell II, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-1.9834	0.2063	< 0.0001	-	-
1/Fallzahl [FP1]	1	1473.4	192.6	< 0.0001	siehe Tab. 23	
(1/Fallzahl)×log(Fallzahl) [FP2]	1	-277.2	36.13	< 0.0001		
(Euro-Score) ^{-0.5} [FP1]	1	1.3680	0.1971	< 0.0001	-	-
log(Euro-Score) [FP2]	1	0.4731	0.0532	< 0.0001	-	-
BMI	1	0.00658	0.0028	0.0179	1.007	1.001-1.012
CDC-Klasse	1	-0.3064	0.0182	< 0.0001	0.736	0.710-0.763
Hosmer-Lemeshow-Anpassungstest: 149 ($p < 0.0001$) Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.009$; Adjusted-Likelihood-Ratio $R^2 = 0.016$, $c = 0.57$						

Zur Überprüfung von Cluster-Effekten wurde ein GEE-Modell basierend auf Modell II angepasst. Tabelle 22 zeigt die Ergebnisse des GEE-Ansatzes (Modell III). Die Korrelation innerhalb der Krankenhäuser beträgt $r = 0.06$ und führt zu einer deutlichen Änderung der Ergebnisse von Modell II, so dass wir Modell III als endgültiges Modell verwenden.

Tab. 22: Ergebnisse des GEE-Ansatzes für "Anastomosen" (Modell III, 2004)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-2.6219	0.7287	0.0003	-	-
1/Fallzahl [FP1]	1	-142.03	1390.9	0.9187	siehe Tab. 23	
(1/Fallzahl)×log(Fallzahl) [FP2]	1	48.7661	285.25	0.8643		
(Euro-Score) ^{-0.5} [FP1]	1	1.3550	0.2217	< 0.0001	-	-
log(Euro-Score) [FP2]	1	0.4572	0.0610	< 0.0001	-	-
BMI	1	0.0080	0.0027	0.0031	1.008	1.003-1.013
CDC-Klasse	1	-0.3562	0.0451	< 0.0001	0.700	0.641-0.765
Exchangeable Kovarianzstruktur: r=0.07 Erklärte Variabilität: Squared-Pearson-Correlation r ² =0.009						

Nach Berücksichtigung von Cluster-Effekten ist der Einfluss der Fallzahl nicht länger statistisch signifikant. In Modell III nimmt das Risiko mit steigender Fallzahl ab, jedoch sind die Schätzungen mit extremer Unsicherheit behaftet, was sich in den sehr hohen p-Werten für die Fallzahl in diesem Modell sowie den breiten Konfidenzintervallen für die Odds Ratios (Tabelle 23) widerspiegelt.

Tab. 23: Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-III für "Anastomosen" (2004)

Intervall	Referenzwert	OR ^a (95%-KI)		
		Modell I	Modell II	Modell III
1-632	316	1.004 (0.955-1.056)	1.084 (1.019-1.153)	1.244 (0.811-1.907)
633-1060	846	1.000	1.000	1.000
1061-2266	1663	1.239 (1.200-1.278)	1.124 (1.086-1.163)	0.915 (0.621-1.349)
^a bezogen auf Differenz der Intervallmittelpunkte				

Es wurden wieder Wechselwirkungen 1. Ordnung untersucht, die jedoch auf dem 1%-Niveau nicht statistisch signifikant waren. Abschließend wurde die Robustheit der Ergebnisse mit Hilfe von Sensitivitätsanalysen überprüft. Eine Ersetzung des Euro-Scores durch den KCH-Score und eine Modellierung der CDC-Klassen als kategorielle Variable in Modell III führte zu vergleichbaren Ergebnissen. Auch die Berücksichtigung der Dringlichkeit als zusätzliche Kovariable ergab keine relevante Veränderung der Ergebnisse.

Von den 49091 Datensätzen wiesen 13.2% der Daten fehlende Werte bei den Kovariablen auf. Ein Vergleich der Risiken für Patienten ohne und mit fehlenden Werten lieferte keinen Hinweis auf einen gravierenden Unterschied zwischen den Patientengruppen. Das Risiko für den Indikator "Anastomosen" betrug 24.8% für

Patienten ohne fehlende Werte und 28.1% für Patienten mit fehlenden Werten. Die entsprechenden medianen Fallzahlen waren 803 (25%- und 75%-Quartil: 632 und 1060) und 765 (25%- und 75%-Quartil: 632 und 1090). Somit ergibt sich kein Anhaltspunkt für einen relevanten Unterschied zwischen den Patientengruppen mit und ohne fehlende Kovariablenwerte. Eine Auswertung ohne Berücksichtigung von Extremwerten führte zu keiner relevanten Veränderung der Ergebnisse. Eine alternative Dichotomisierung des Indikators "Anastomosen" mit Cut-point bei einer Anastomose (d.h. ≤ 1 vs. > 1) führte ebenso zu vergleichbaren Ergebnissen.

5.5.2 Zusammenhang zwischen Fallzahl und Risiko (2003)

Das mittlere Risiko für den Indikator "Anastomosen" beträgt 25.1%. Abbildung 13 zeigt den Risikoplot in Abhängigkeit von der Fallzahl. Die Struktur ist den Daten aus dem Jahr 2004 sehr ähnlich.

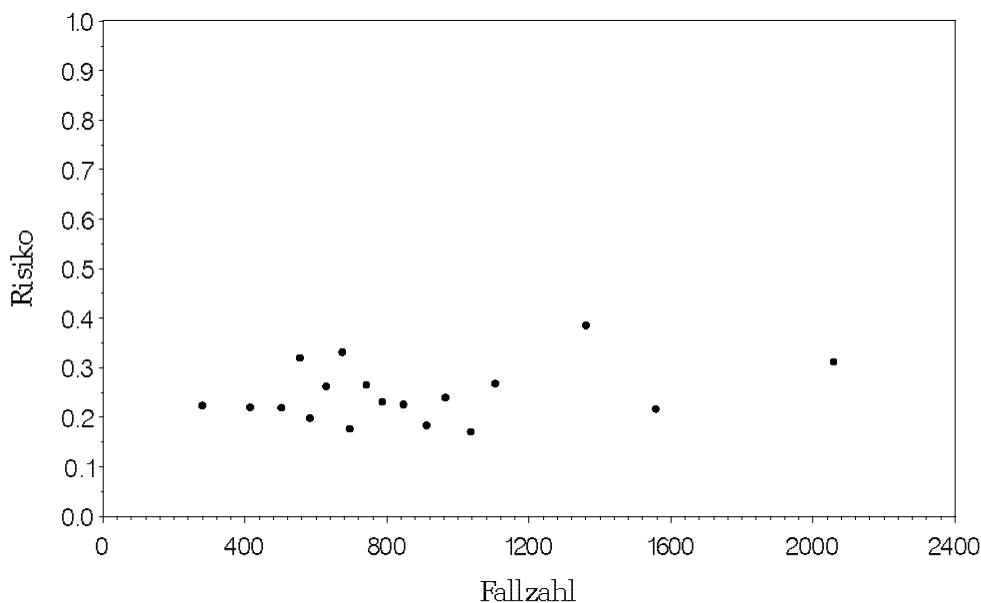


Abb. 13: Risikoplot für "Anastomosen" vs. Fallzahl (2003)

Zur Validierung der Ergebnisse wurden die Modelle für die Daten aus dem Jahr 2004 auf die $n=39729$ Datensätze aus dem Jahr 2003 angewendet. Wir zeigen hier nur die wichtigsten Ergebnisse des GEE-Ansatzes mit FPs (Modell III, Tab. 24) sowie die Odds Ratios für die Modell I-III (Tab. 25).

Tab. 24: Ergebnisse des GEE-Ansatzes für "Anastomosen" (Modell III, 2003)

Parameter	FG	Parameterschätzung	Standardfehler	p-Wert	OR	95%-KI
Intercept	1	-2.2694	0.4350	< 0.0001	-	-
1/Fallzahl [FP1]	1	-475.90	459.84	0.3007	siehe Tab. 25	
(1/Fallzahl)×log(Fallzahl) [FP2]	1	98.99	99.02	0.3175		
(Euro-Score) ^{-0.5} [FP1]	1	1.2700	0.3020	< 0.0001	-	-
log(Euro-Score) [FP2]	1	0.3770	0.0896	< 0.0001	-	-
BMI	1	-0.0006	0.0036	0.8622	0.999	0.992-1.006
CDC-Klasse	1	-0.3725	0.0431	< 0.0001	0.689	0.633-0.750
Exchangeable Kovarianzstruktur: $r=0.05$ Erklärte Variabilität: Squared-Pearson-Correlation $r^2=0.01$						

Tab. 25: Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-III für "Anastomosen" (2003)

Intervall	Referenzwert	OR ^a (95%-KI)		
		Modell I	Modell II	Modell III
1-598	299	0.819 (0.774-0.867)	0.911 (0.839-0.990)	1.075 (0.804-1.438)
599-1140	869	1.000	1.000	1.000
1141-2243	1692	1.119 (1.087-1.151)	1.033 (0.989-1.079)	0.933 (0.792-1.099)
^a bezogen auf Differenz der Intervallmittelpunkte				

Wie bei den Daten von 2004 scheint sich ein fallendes Risiko mit steigender Fallzahl anzudeuten, jedoch sind erneut die Ergebnisse nicht statistisch signifikant.

5.5.3 Schlussfolgerung

Die Ergebnisse der multifaktoriellen Modellierung mit Hilfe des GEE-Ansatzes zeigen eindeutig, dass bezüglich des Zusammenhangs zwischen der Fallzahl und dem Indikator "Anastomosen" eine hohe Variabilität sowie ein Cluster-Effekt vorhanden sind. Es scheint sich am ehesten ein mit steigender Fallzahl sinkendes Risiko anzudeuten. Bei der vorliegenden geringen Zahl an Krankenhäusern (n=77) ist jedoch der Zusammenhang zwischen der Fallzahl und dem Indikator "Anastomosen" von dem Cluster-Effekt überlagert, so dass die Form des Zusammenhangs nicht ermittelt werden kann und unklar bleibt.

6. Diskussion

Bei der Bewertung der in diesem Bericht beschriebenen Ergebnisse sind vor allem zwei wichtige Aspekte zu berücksichtigen. Die zugrunde liegenden Daten sind routinemäßig erhobene Daten der BQS Bundesauswertungen 2003 und 2004. Die Daten sind also nicht für den Zweck dieses Berichts, nämlich die Untersuchung von Volume-Outcome-Beziehungen bei der Koronarchirurgie erhoben worden, wodurch sie für die vorliegende Untersuchung zwangsläufig nur eingeschränkte Aussagekraft besitzen. Die Qualität dieser Daten ist unklar, da sie von den Krankenhäusern selbst berichtet werden und im Rahmen dieses Auftrags bezüglich ihrer Validität nicht überprüfbar waren. Eine solche Überprüfung der Krankenhausangaben wäre jedoch notwendig, um Analysen auf der Basis einer gesicherten Datenqualität durchzuführen. Einen deutlichen Hinweis auf eine fragwürdige Qualität der Berichterstattung der Krankenhäuser liefert ein von der BQS durchgeführter Datenabgleich von BQS-Daten mit den zugehörigen Krankenakten [18]. In einer Stichprobe von 685 Krankenakten zeigte sich, dass von allen behandlungsbedürftigen, chirurgischen und allgemeinen Komplikationen gemäß Krankenakte nur 65.6% (84 von 128) auch an die BQS berichtet wurden. Auf der anderen Seite fanden sich nur 63.2% (84 von 133) der an die BQS berichteten Komplikationen auch in der Krankenakte wieder [18]. Mit anderen Worten: Bei der dem Datenabgleich zugrunde liegenden Stichprobe fehlten in den BQS-Daten ca. 34% der Komplikationen aus den Krankenakten und ca. 37% der an die BQS berichteten Komplikationen waren in den Krankenakten nicht dokumentiert. Die Berichtsqualität der Krankenhäuser speziell für die Koronarchirurgie und für die in diesem Bericht ausgewerteten Daten ist unbekannt. Sollte die oben beschriebene allgemeine schlechte Berichtsqualität auch für die hier gewählten Indikatoren der Ergebnisqualität für die Koronarchirurgie zutreffen, so kann dies je nach Verteilung der fehlenden und irrtümlich berichteten Komplikationen auf den Zusammenhang zwischen Leistungsmenge und Komplikationsrisiko einen deutlichen Einfluss haben. Aus diesem Grund sind die Ergebnisse dieses Berichts entsprechend vorsichtig zu interpretieren.

Unabhängig von der fragwürdigen Datenqualität gilt Folgendes: Hätte man in der vorliegenden Auswertung in der Beziehung zwischen Ergebnisqualität und Fallzahl deutliche Schwellenwerte gefunden, so wäre dies keinesfalls ein ausreichender Beleg dafür gewesen, dass die Verwendung dieses Schwellenwerts als Mindestmenge zu einer entsprechenden Qualitätsverbesserung führt. Ein solcher Beleg kann zuverlässig nur über eine kontrollierte Interventionsstudie geführt werden und nicht über den Nachweis einer statistischen Assoziation aus Routinedaten.

7. Zusammenfassung

Mit Hilfe von Daten des Bundesdatenpools der BQS aus den Jahren 2004 und 2003 von Patienten mit isoliert koronarchirurgischen Erst-Operationen aus 77 Krankenhäusern wurde der Zusammenhang zwischen dem Risiko für "Letalität" (primärer Qualitätsindikator) bzw. dem Risiko für "Infektion" (sekundärer Qualitätsindikator) und der Fallzahl des betreffenden Krankenhauses pro Jahr mit Hilfe von logistischen Regressionsmodellen untersucht. Für keinen der beiden Indikatoren wurde nach Berücksichtigung relevanter Confounder (Euro-Score, BMI, CDC-Klasse) ein signifikanter Zusammenhang zwischen der Fallzahl und dem entsprechenden Risiko statistisch nachgewiesen. Ein deutlicher Cluster-Effekt, der die Beziehung zwischen Risiko und Fallzahl verändert, wurde bei den beiden Indikatoren der Ergebnisqualität nicht gefunden. Die vorliegende Auswertung unterstützt damit nicht die Hypothese, dass es in Deutschland bei der Koronarchirurgie einen Zusammenhang zwischen der erbrachten Leistungsmenge und der Ergebnisqualität gibt. Folglich wurden auch keine Methoden zur Berechnung von Schwellenwerten angewendet. Es ist hierbei jedoch zu beachten, dass in den zur Verfügung stehenden BQS-Daten die Leistung Koronarchirurgie nur von 77 Krankenhäusern angeboten wurde und im Jahr 2004 die minimale Leistungsmenge 166 Fälle betrug. Es ist somit möglich, dass es sehr wohl einen Zusammenhang zwischen der Leistungsmenge und der Ergebnisqualität gibt, sich dieser aber auf geringere Fallzahlen beschränkt, die es in Deutschland im Jahr 2004 gar nicht gab.

Auf der gleichen Datenbasis (Bundesdatenpools der BQS aus den Jahren 2004 und 2003) wurde mit Hilfe logistischer Regressionsmodelle auch untersucht, ob es bei der Koronarchirurgie einen Zusammenhang zwischen der Leistungsmenge und den Indikatoren der Prozessqualität "Verwendung der Brustwandarterie" sowie "Anzahl peripherer Anastomosen" gibt. Auch für die Indikatoren der Prozessqualität wurde nach Berücksichtigung relevanter Confounder sowie eines Cluster-Effekts kein statistisch signifikanter Zusammenhang mit der Leistungsmenge gefunden. Hier zeigte sich bei beiden Indikatoren ein deutlicher Cluster-Effekt. Bezüglich des Risikos der Nichtverwendung der Brustwandarterie deutete sich ein U-förmiger Zusammenhang, bezüglich des Risikos einer Anzahl peripherer Anastomosen unterhalb von drei ein mit steigender Fallzahl sinkendes Risiko an. Aufgrund der vorliegenden geringen Zahl an Krankenhäusern (n=77) und den überlagernden Cluster-Effekten konnte aber die Form des Zusammenhangs zwischen der Fallzahl und den beiden Indikatoren der Prozessqualität nicht mit ausreichender Sicherheit ermittelt werden und bleibt daher unklar.

Aus den vorliegenden Daten lassen sich keine Hypothesen ableiten, dass es in Deutschland für Patienten mit isoliert koronarchirurgischen Erst-Operationen einen Zusammenhang zwischen der erbrachten Leistungsmenge der Krankenhäuser und der Ergebnisqualität gibt. Um den wissenschaftlichen Nachweis zu führen, dass eine Mindestmengenregelung zu einer Verbesserung der Ergebnisqualität führt, ist grundsätzlich eine kontrollierte Interventionsstudie notwendig.

8. Literatur

- [1] Rathmann W, Windeler J. *Zusammenhang zwischen Behandlungsmenge und Behandlungsqualität*. Evidenzbericht, Fachbereich Evidenz-basierte Medizin des MDS, Essen, 2002.
- [2] Geraedts M. *Evidenz zur Ableitung von Mindestmengen in der Medizin*. Gutachten im Auftrag der Bundesärztekammer, 2003.
- [3] Heller RF, Dobson A. Disease impact number and population impact number: Population perspectives to measures of risk and benefit. *BMJ* 2000; 321: 950-952.
- [4] Mohr VD, Bauer J, Döbler K, Eckert O, Fischer B, Woldenga C (Hrsg.). *Qualität sichtbar machen. BQS-Qualitätsreport 2004*. Düsseldorf: BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH, 2005.
- [5] Mohr VD, Bauer J, Döbler K, Eckert O, Fischer B, Woldenga C (Hrsg.). *Qualität sichtbar machen. BQS-Qualitätsreport 2003*. Düsseldorf: BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH, 2004.
- [6] Culver DH, Horan TC, Gaynes RP, Martone WJ, Jarvis WR, Emori TG, Banerjee SN et al. Surgical wound infection rates by wound class, operative procedure, and patient risk index. National Nosocomial Infections Surveillance System. *Am J Med* 1991; 91 (Suppl 3B): 152S-157S.
- [7] Hosmer DW, Lemeshow S. *Applied Logistic Regression (2nd Ed.)*. New York: Wiley, 2000.
- [8] Bender R, Ziegler A, Lange S. Logistische Regression. *Dtsch Med Wochenschr* 2002; 127: T11-T13.
- [9] Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 1999; 28: 964-974.
- [10] Royston P, Sauerbrei W. Building multivariable regression models with continuous covariates in clinical epidemiology - with an emphasis on fractional polynomials. *Methods Inf Med* 2005; 44: 561-571.
- [11] Kleinbaum DG, Klein M. *Logistic Regression: A Self-Learning Text (2nd Ed.)*. New York: Springer, 2002.

- [12] Panageas KS, Schrag D, Riedel E, Bach PB, Begg CB. The effect of clustering of outcomes on the association of procedure volume and surgical outcomes. *Ann Intern Med* 2003; 139: 658-666.
- [13] Urbach DR, Austin PC. Conventional models overestimate the statistical significance of volume-outcome associations, compared with multilevel models. *J Clin Epidemiol* 2005; 58: 391-400.
- [14] Nagelkerke NJD. A note on the general definition of the coefficient of determination. *Biometrika* 1991; 78: 691-692.
- [15] Mittlböck M, Schemper M. Explained variation for logistic regression. *Stat Med* 1996; 15: 1987-1997.
- [16] Pregibon D. Logistic regression diagnostics. *Ann Stat* 1981; 9: 705-724.
- [17] Bender R, Grouven U. Möglichkeiten und Grenzen statistischer Regressionsmodelle zur Berechnung von Schwellenwerten für Mindestmengen. *Z ärztl Fortbild Qual Gesundheitswes* 2006; 100: 93-98.
- [18] Fischer, B. Datenvalidierung im BQS-Verfahren: Ergebnisse eines Probebetriebs. Vortrag beim Workshop "Qualität medizinischer Daten in Forschung und Versorgung" der TMF, der KGNW und der GMDS. Essen, 16.02.2006.

Anhang: Weiterer Zeitplan

Arbeitsschritt	Termin (geplant)
Einholung von schriftlichen Stellungnahmen	Bis 30.05.2006 (vier Wochen nach Veröffentlichung des Vorberichts)
Wissenschaftliche Erörterung	Juni 2006
Weitergabe des Abschlussberichts an den G-BA	Juli 2006
Veröffentlichung des Abschlussberichts	Zwei Monate nach Weitergabe an den G-BA