

IQWiG-Berichte · Jahr: 2006 Nr. 9

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

Abschlussbericht

Auftrag B05/01b
Version 1.0
Stand: 20.06.2006

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

Thema:

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

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

22.12.2004

Interne Auftragsnummer:

B05/01b

Anschrift des Herausgebers:

IQWiG
Dillenburger Straße 27
51105 Köln

Telefon (02 21) 3 56 85-0

Telefax (02 21) 3 56 85-1

berichte@iqwig.de

www.iqwig.de

ISSN

1864-2500



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

- Abschlussbericht vom 20.06.2006 -

[Auftrag B05/01b]

Version 1.0

Stand: 20.06.2006

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

Externer Reviewer: Prof. Dr. Helmut Küchenhoff (München)

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.

Dieser Bericht sollte wie folgt zitiert werden:

IQWiG. Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für die Koronarchirurgie. Abschlussbericht B05/01b. Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Juni 2006.

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
3.1 Verlauf des Projektes	8
3.2 Änderungen des Vorberichts aufgrund der Stellungnahmen	9
4. Methoden	10
4.1 Datenquellen	10
4.2 Vorgaben der Fachexperten	10
4.3 Auswertungen	11
4.3.1 Datenbeschreibung	11
4.3.2 Modellbildung	13
4.3.3 Berechnung von Schwellenwerten	15
4.3.4 Multiples Testen und Signifikanzniveau	15
4.3.5 Software	15
4.4 Ergänzungen im Vergleich zum Berichtsplan	15
5. Ergebnisse	16
5.1 Deskriptive Statistik	16
5.2 Ergebnisse für den Indikator "Letalität"	20
5.2.1 Zusammenhang zwischen Fallzahl und Risiko (2004)	20
5.2.2 Zusammenhang zwischen Fallzahl und Risiko (2003)	25
5.2.3 Schlussfolgerung	26
5.3 Ergebnisse für den Indikator "Infektion"	27
5.3.1 Zusammenhang zwischen Fallzahl und Risiko (2004)	27
5.3.2 Zusammenhang zwischen Fallzahl und Risiko (2003)	31
5.3.3 Schlussfolgerung	32
5.4 Ergebnisse für den Indikator "Brustwandarterie"	33
5.4.1 Zusammenhang zwischen Fallzahl und Risiko (2004)	33
5.4.2 Zusammenhang zwischen Fallzahl und Risiko (2003)	37
5.4.3 Schlussfolgerung	38
5.5 Ergebnisse für den Indikator "Anastomosen"	39
5.5.1 Zusammenhang zwischen Fallzahl und Risiko (2004)	39
5.5.2 Zusammenhang zwischen Fallzahl und Risiko (2003)	42
5.5.3 Schlussfolgerung	43
6. Diskussion	44
7. Zusammenfassung	45
8. Literatur	47
Anhang: Stellungnahmen	49
A.1 Stellungnahme der DKG	49
A.2 Stellungnahme der BÄK	53
A.3 Stellungnahme der BQS	58
A.4 Antwort auf die Stellungnahmen	66

Tabellenverzeichnis

Tabelle 1	Deskriptive Übersicht
Tabelle 2	Rangkorrelationen der Kovariablen
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
BÄK	Bundesärztekammer
BMI	Body Mass Index
BQS	BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH
CDC	Centers for Disease Control and Prevention
DKG	Deutsche Krankenhausgesellschaft
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

3.1 Verlauf des Projektes

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 potentiellen 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.

Der Vorbericht in der Version 1.0 vom 24.04.2006 wurde am 25.04.2006 dem Gemeinsamen Bundesausschuss und dem Kuratorium des Instituts und übermittelt und am 02.05.2006 im Internet veröffentlicht.

Zu diesem Vorbericht konnten bis einschließlich 30.05.2006 schriftliche Stellungnahmen von allen interessierten Privatpersonen, Patientenvertretern, Fachgesellschaften und Industrieunternehmen im Sinne einer Anhörung abgegeben werden. Der Vorbericht wurde zudem einem externen Review unterzogen. Eine wissenschaftliche Erörterung der schriftlichen Stellungnahmen fand nicht statt. Die schriftlichen Stellungnahmen sowie eine Antwort des IQWiG auf die Stellungnahmen finden sich im Anhang.

3.2 Änderungen des Vorberichts aufgrund der Stellungnahmen

Durch die schriftlichen Stellungnahmen ergaben sich im Abschlussbericht folgende Änderungen im Vergleich zum Vorbericht.

- In Abschnitt 4.3.1 wurde der Text um den Hinweis ergänzt, dass die Daten von 2003 auch zur Festlegung der klinisch relevanten Veränderungen durch die Fachexperten verwendet wurden.
- In den Abschnitten 4.3.2, 5.2.1 und 5.2.2 wurden die Methode bzw. die Ergebnisse des Tests auf Wechselwirkung zwischen Dringlichkeit und Fallzahl ergänzt.
- In Kapitel 6 wurden einige Umformulierungen vorgenommen, um die Äußerungen zu präzisieren sowie missverständliche Interpretationen zu vermeiden.

4. Methoden

Die Methoden zur Erstellung des Berichts wurden im Berichtsplan vom 05.09.2005 vorab festgelegt. Sofern sich in der Bearbeitung des Projekts Änderungen im Vergleich zum Berichtsplan 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 auf der Basis der deskriptiven Ergebnisse der Daten aus dem Jahr 2003 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 sowie zur Festlegung der klinisch relevanten Veränderung (siehe Abschnitt 4.2) 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 Variablen zur Verfügung gestellt:

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

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]. Dabei wurde gemäß der Selektionsprozedur von Royston und Sauerbrei [10] vorgegangen.

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 wurde dieser 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. Zudem wurde die Wechselwirkung zwischen Dringlichkeit und Fallzahl untersucht. Für die stetigen Kovariablen der multifaktoriellen Modelle wurde ebenfalls die FP-Methode [9,10] zur flexiblen Modellierung eingesetzt. 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 Überprüfung der Modellstabilität wurde dann ein erneuter Selektionsprozess zur Identifizierung der bestangepassten FPs für die Fallzahl unter Berücksichtigung der im Endmodell vorhandenen Confounder durchgeführt. Falls eine Modellierung der Fallzahl durch FPs 2. Ordnung vorlag, wurde zur Bewertung des Einflusses der Fallzahl auch ein simultaner Test bezüglich beider FPs durchgeführt.

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, CDC-Klasse, Dringlichkeit).

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)
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 der Kovariablen

Jahr 2003						
	Fallzahl	Dringlichkeit ^a	CDC-Klasse	BMI	Euro-Score	KCH-Score
Fallzahl	1	-0.02	-0.10	0.01	0.02	0.03
Dringlichkeit ^a		1	0.07	-0.03	0.28	0.28
CDC-Klasse			1	0.04	0.11	0.12
BMI				1	-0.10	-0.07
Euro-Score					1	0.92
Jahr 2004						
Fallzahl	1	-0.01	-0.13	0.02	0.01	0.01
Dringlichkeit ^a		1	0.06	-0.03	0.26	0.27
CDC-Klasse			1	0.03	0.12	0.12
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.

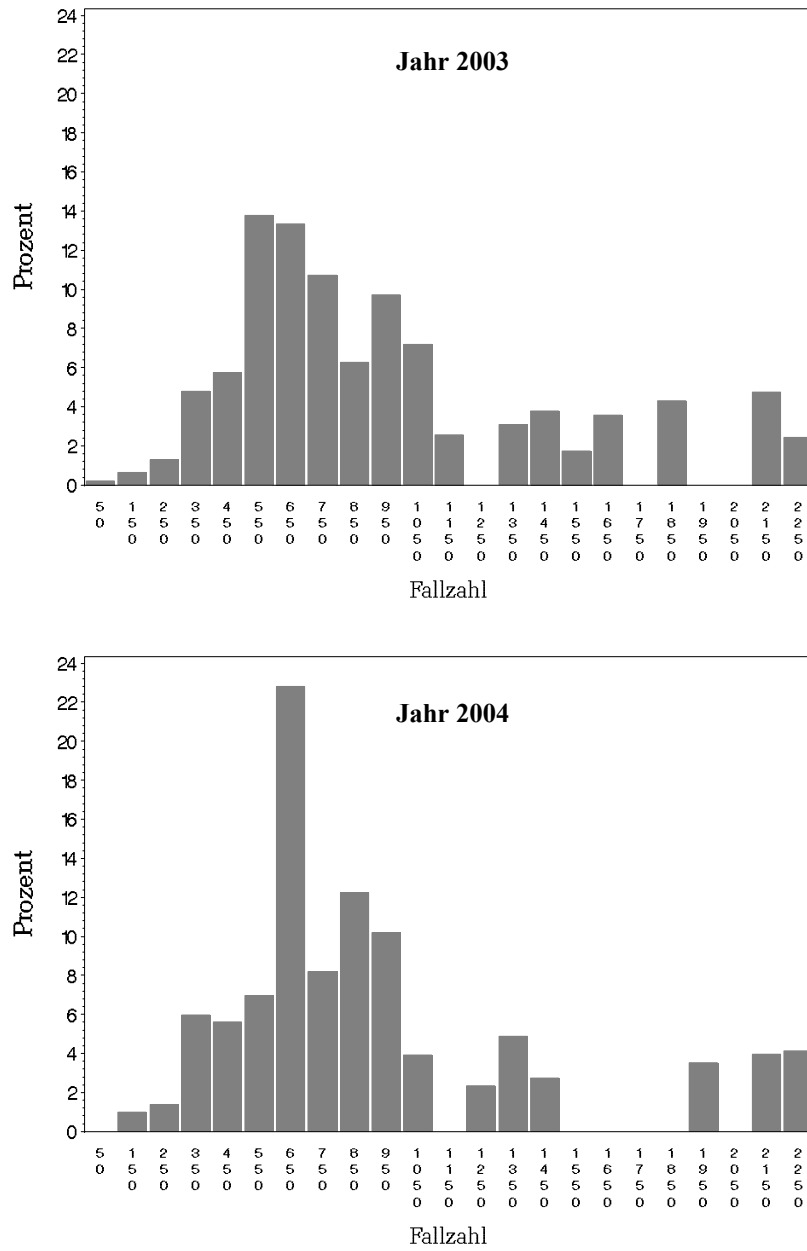


Abb. 1: Verteilung der Fallzahl pro Krankenhaus (2003 oben, 2004 unten)
 (Die Zahlen an der horizontalen Achse geben die Klassenmittelwerte an)

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".

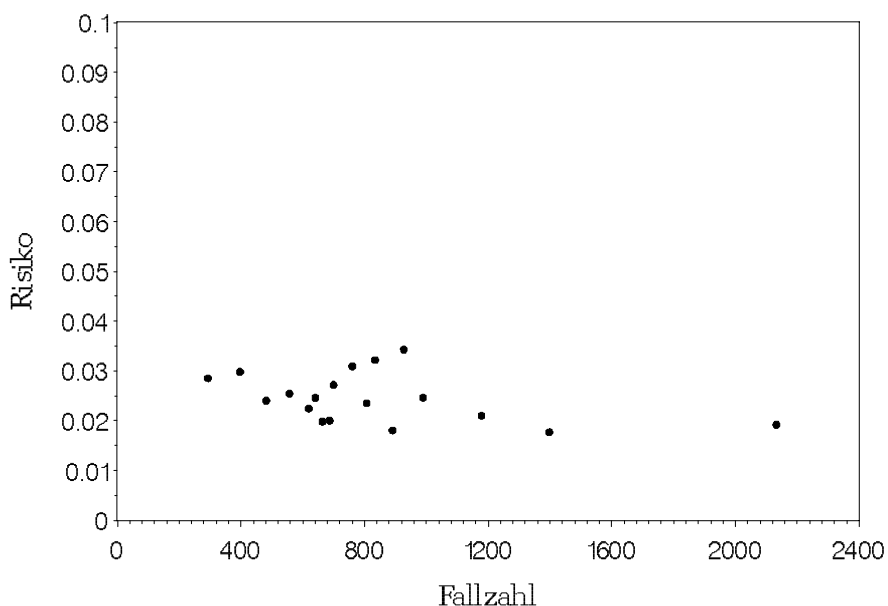


Abb. 2: Risikoplot für "Letalität" vs. Fallzahl (2004)

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.

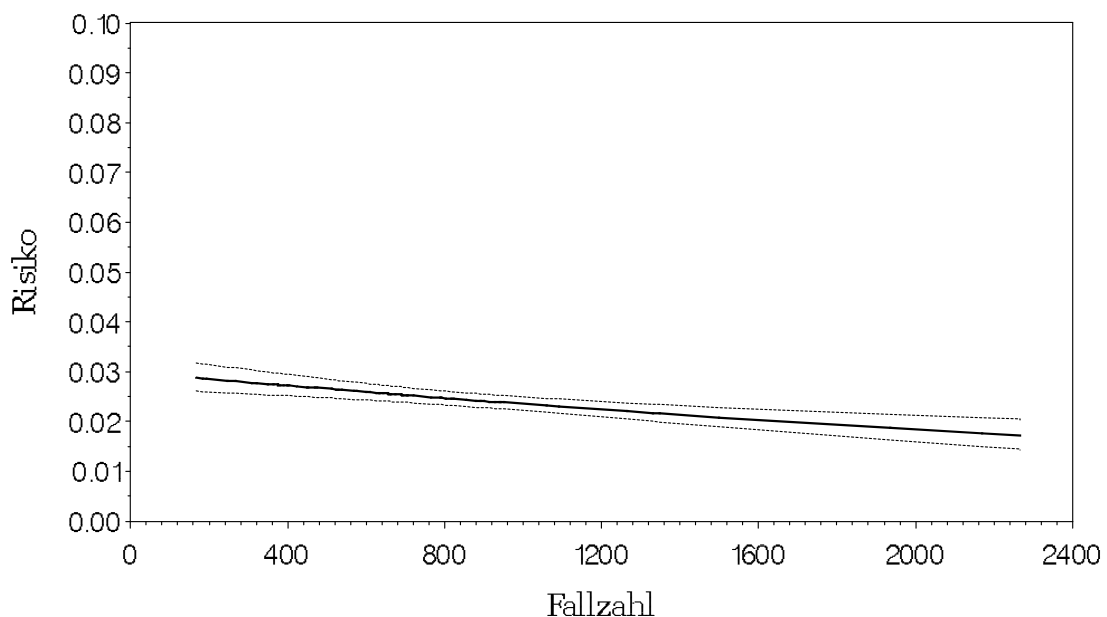


Abb. 3: Rohe Risikokurve mit 95%-Konfidenzband für "Letalität" aus Modell I (2004)

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 multi-faktoriellen 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.

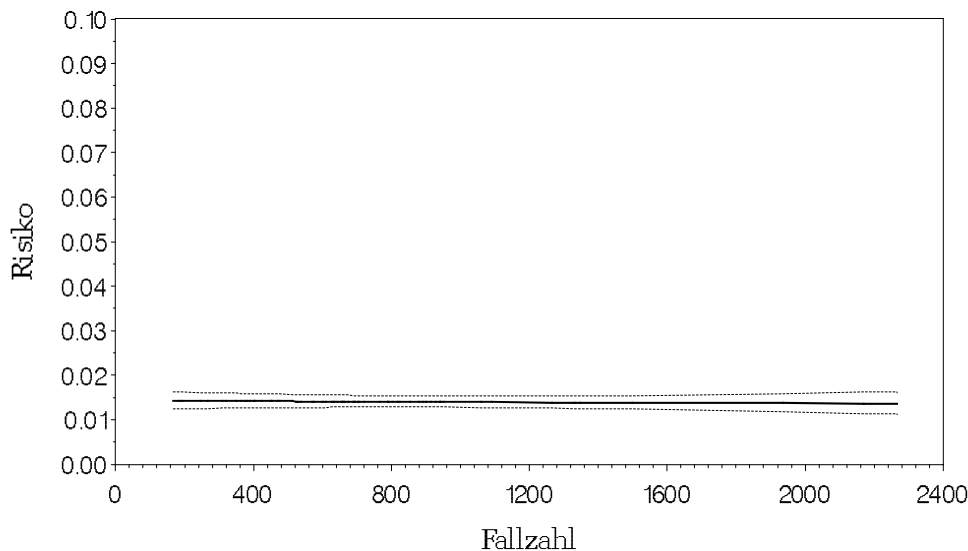


Abb. 4: Adjustierte Risikokurve mit 95%-Konfidenzband für "Letalität" aus Modell III (2004) (Kovariablen auf Mittelwert gesetzt)

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. Die Wechselwirkung zwischen Dringlichkeit und Fallzahl war zum Niveau $\alpha=0.05$ (zweiseitig) nicht statistisch signifikant.

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.

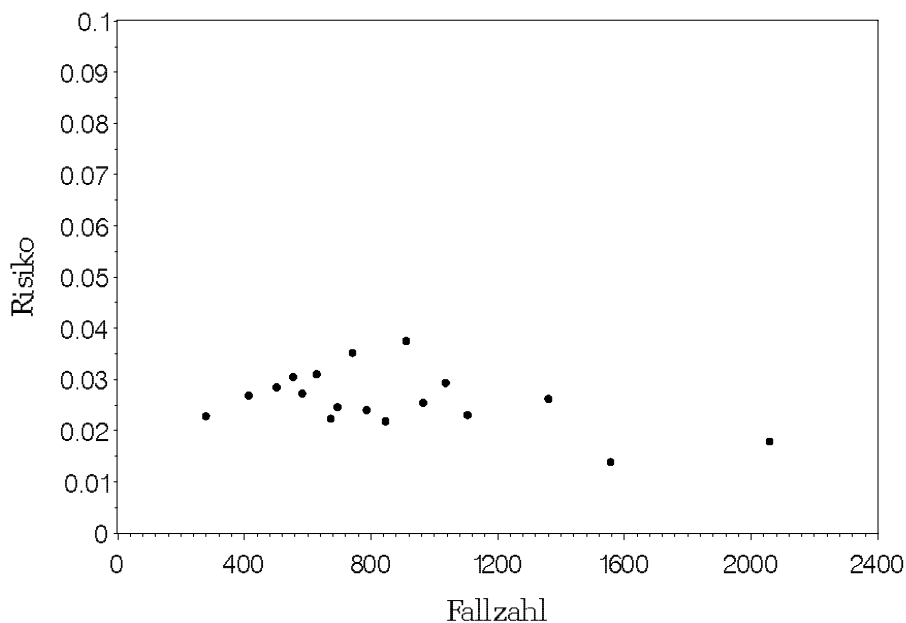


Abb. 5: Risikoplot für "Letalität" vs. Fallzahl (2003)

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.

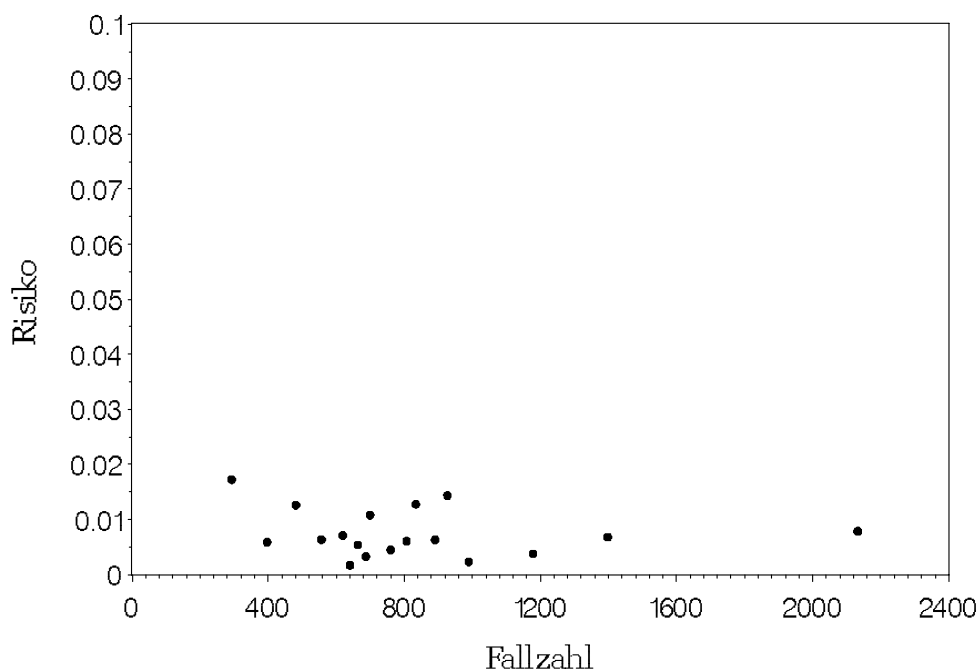


Abb. 6: Risikoplot für "Infektion" vs. Fallzahl (2004)

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 ($p=0.09$).

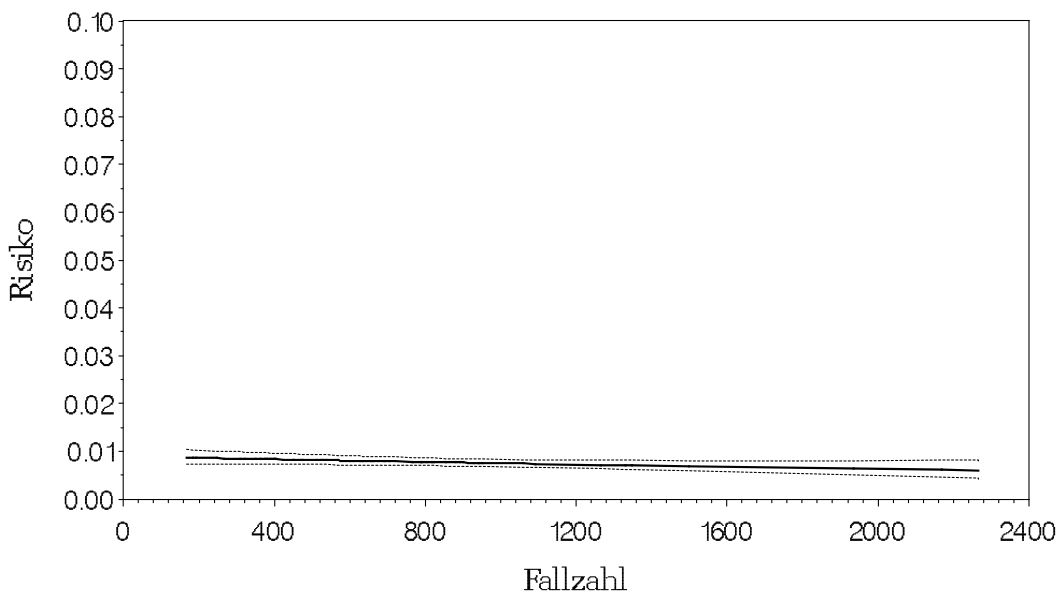


Abb. 7: Rohe Risikokurve mit 95%-Konfidenzband für "Infektion" aus Modell I (2004)

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.

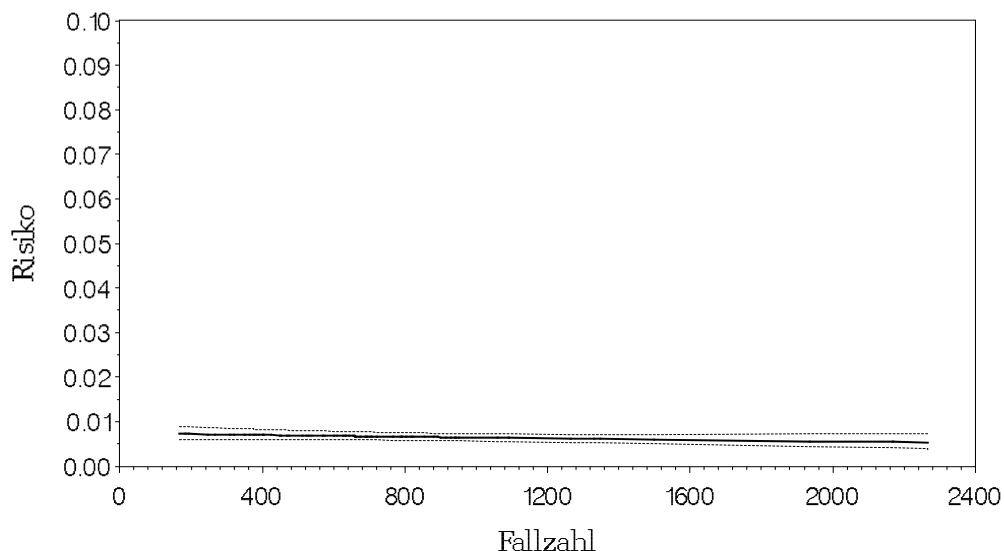


Abb. 8: Adjustierte Risikokurve mit 95%-Konfidenzband für "Infektion" aus Modell III (2004) (Kovariablen auf Mittelwert gesetzt)

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 Resultate entsprachen denen 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. Die Wechselwirkung zwischen Dringlichkeit und Fallzahl war zum Niveau $\alpha=0.05$ (zweiseitig) nicht statistisch signifikant.

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.

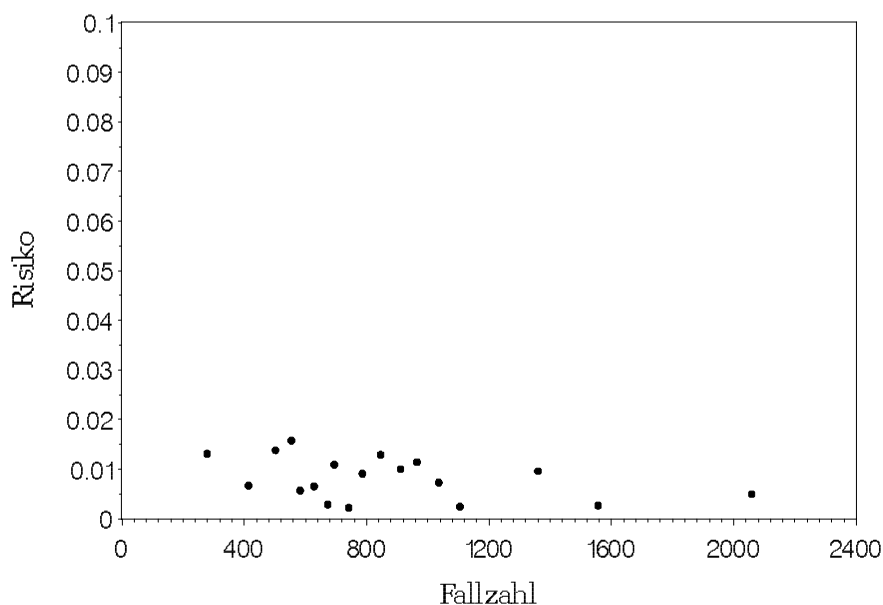


Abb. 9: Risikoplot für "Infektion" vs. Fallzahl (2003)

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.

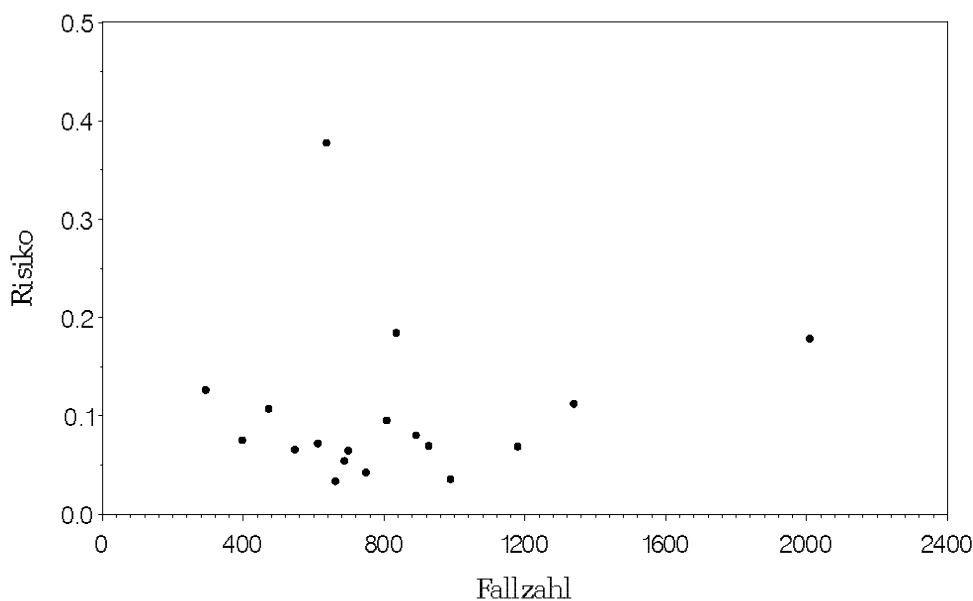


Abb. 10: Risikoplot für "Brustwandarterie" vs. Fallzahl (2004)

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] ^a	1	-3.6E-7	9.25E-7	0.6937	siehe Tab. 17	
Fallzahl ³ [FP2] ^a	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 $r^2=0.04$ ^a Simultaner Test der FPs für die Fallzahl: $p=0.8276$						

Die Berücksichtigung des Cluster-Effekts führt dazu, dass der Einfluss der Fallzahl nicht länger statistisch signifikant ist ($p=0.8276$). 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. Zunächst wurde eine erneute Selektionsprozedur zur Ermittlung der optimalen FPs unter Berücksichtigung der in Modell III eingeschlossenen Kovariablen durchgeführt. Dies resultierte in anderen FPs, die sich jedoch hinsichtlich der Modellanpassung und der Form der Risikokurve nur minimal von den verwendeten FPs unterschieden und die Gesamtergebnisse nicht veränderten. In einem nächsten Schritt wurden 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 II) 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 III zu einer niedrigeren Korrelation ($r=0.06$) und einer Reduktion des simultanen p -Werts für die FPs der Fallzahl ($p=0.3641$) sowie zu einer Verkleinerung der Konfidenzintervallbreite ohne jedoch 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.

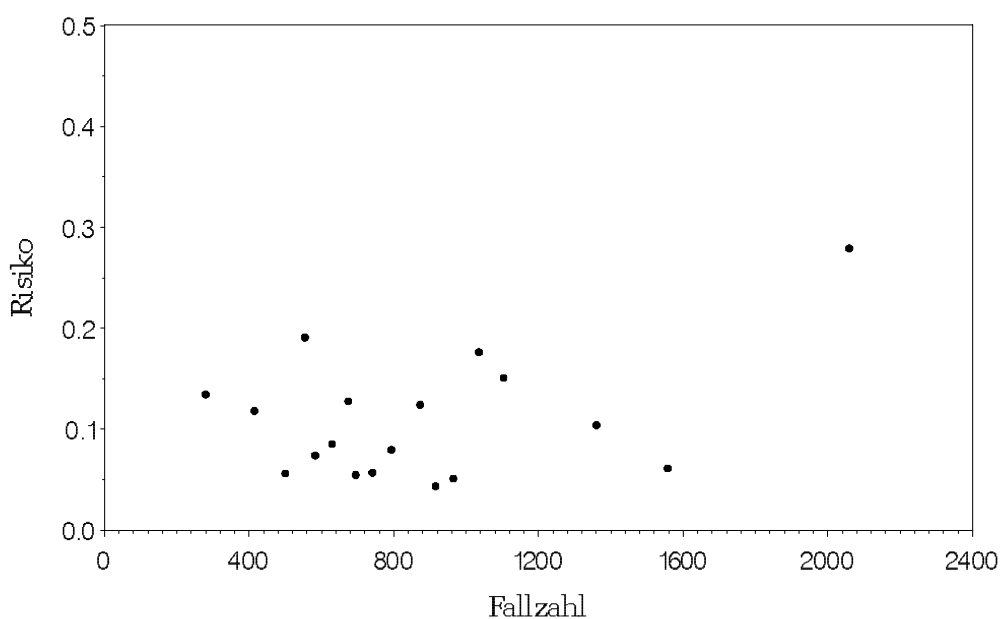


Abb. 11: Risikoplot für "Brustwandarterie" vs. Fallzahl (2003)

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] ^a	1	-9.09E-7	9.85E-7	0.3562	siehe Tab. 19	
Fallzahl ³ [FP2] ^a	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 $r^2=0.056$ ^a Simultaner Test der FPs für die Fallzahl: $p=0.6215$						

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.

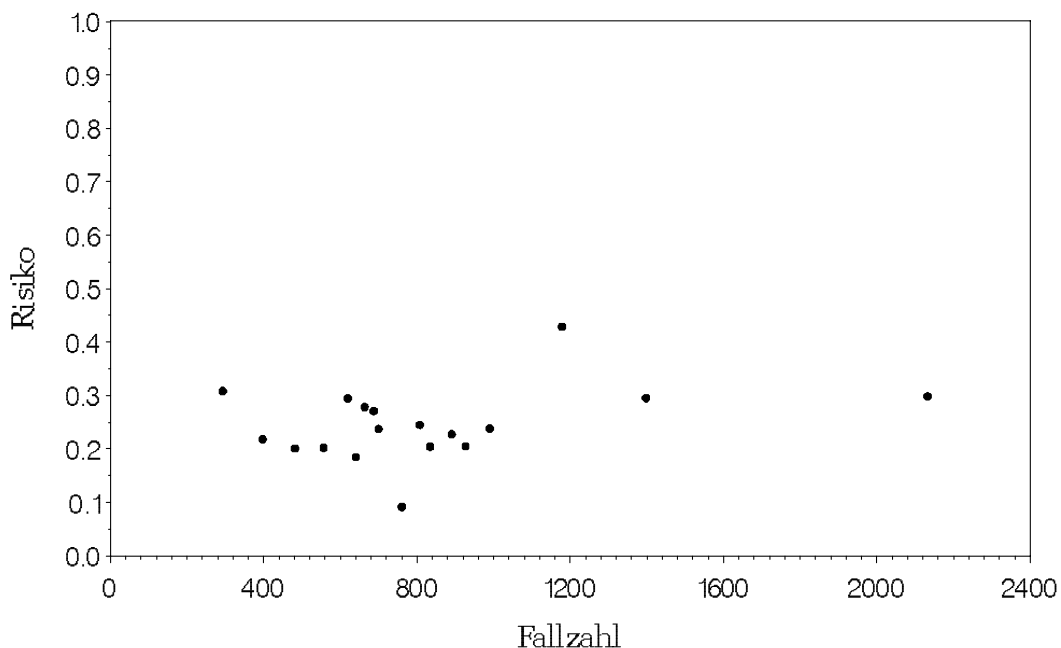


Abb. 12: Risikoplot für "Anastomosen" vs. Fallzahl (2004)

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-Lemeshow-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 $R^2=0.01$, $c=0.58$						

Im 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 $r=0.07$ 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] ^a	1	-142.03	1390.9	0.9187	siehe Tab. 23	
(1/Fallzahl)×log(Fallzahl) [FP2] ^a	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 $r^2=0.009$ ^a Simultaner Test der FPs für die Fallzahl: $p=0.4262$						

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 im hohen p -Wert für die Fallzahl in diesem Modell ($p=0.4262$) 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. Zunächst wurde eine erneute Selektionsprozedur zur Ermittlung der optimalen FPs unter Berücksichtigung der im endgültigen Modell eingeschlossenen Kovariablen durchgeführt. Dies resultierte in anderen FPs, die jedoch zu einer sehr ähnlichen Modellanpassung und Form der Risikokurve führten und die Gesamtergebnisse nicht veränderten. 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.

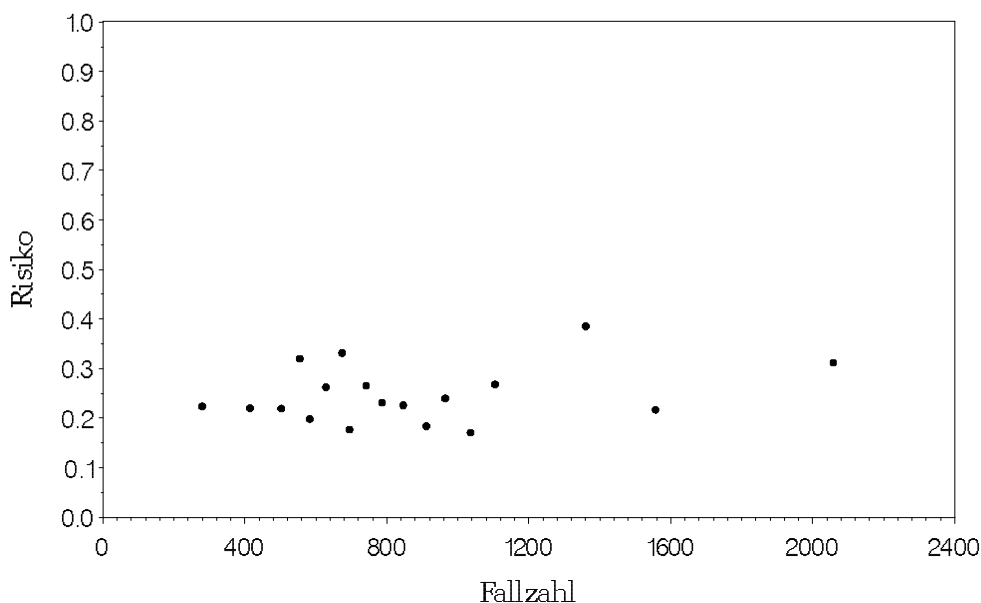


Abb. 13: Risikoplot für "Anastomosen" vs. Fallzahl (2003)

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 Modelle 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] ^a	1	-475.90	459.84	0.3007	siehe Tab. 25	
(1/Fallzahl)×log(Fallzahl) [FP2] ^b	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 $r^2=0.01$ ^a Simultaner Test der FPs für die Fallzahl: $p=0.5532$						

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 Hinweis auf eine fragwürdige Qualität der Berichterstattung der Krankenhäuser liefert ein von der BQS durchgeführter Datenabgleich von BQS-Daten mit den zugehörigen Krankenakten im Rahmen eines Probetriebs zur Datenvalidierung [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. Diese Angaben gelten speziell für alle behandlungsbedürftigen, chirurgischen und allgemeinen Komplikationen. Die Übereinstimmung bezüglich anderer Items war deutlich höher [18]. Man kann vermuten, dass die Datenqualität bezüglich harter Endpunkte wie Letalität deutlich besser ist als die der behandlungsbedürftigen, chirurgischen und allgemeinen Komplikationen in der BQS-Stichprobe, aber letztlich ist die Berichtsqualität der Krankenhäuser speziell für die in diesem Bericht ausgewerteten Daten 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 die Auswertungsergebnisse bezüglich des Zusammenhangs zwischen Leistungsmenge und Komplikationsrisiko einen deutlichen Einfluss haben. Aus diesem Grund sind die Ergebnisse dieses Berichts entsprechend vorsichtig zu interpretieren. Es ist aber unplausibel, dass die möglichen Verzerrungen so große Effekte haben, dass ein in Wahrheit vorhandener Zusammenhang zwischen Leistungsmenge und Komplikationsrisiko von besonderem Ausmaß durch diese Verzerrungen gänzlich überdeckt wird.

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 der Ergebnisqualität 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 (Bundesdatenpool 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. Auf der Basis dieser Ergebnisse dürfte eine Qualitätskontrolle allein über die Leistungsmenge kaum möglich sein. Angebracht erscheint eher eine individuelle Qualitätskontrolle der einzelnen Einrichtungen.

Aus den vorliegenden Daten lassen sich somit 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: Stellungnahmen

A.1 Stellungnahme der DKG



Formblatt zur Anhörung (schriftlichen Stellungnahme) zum Vorbericht

Auftrag des IQWiG/Projektbezeichnung mit Auftragsnummer:
„Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für die Koronarchirurgie“, B05/01b

Name, Vorname; Titel und Funktion des/der Stellungnehmenden:
<ol style="list-style-type: none"> Grüning, Thilo, Dr. med., Referent, Dezernat I, Deutsche Krankenhausgesellschaft Schreck, Jens-Uwe, Dr. med., stellvertretender Geschäftsführer, Dezernat I, Deutsche Krankenhausgesellschaft

Kontakt für weitere Nachfragen:
Institution: Deutsche Krankenhausgesellschaft e.V. Dezernat I – Personal und Organisation Anschrift: Wegelystraße 3, 10623 Berlin Tel./-Fax-Nr.: +49 (0)30-39801-1125 / +49 (0)30-39801-3110 E-Mail-Adresse: t.gruening@dkgev.de

1. Folgende Originalstudien fehlen im Vorbericht

2. Bewertungen von Originalstudien im Vorbericht

3. Anmerkungen zur projektspezifischen Methodik (bei Änderungsvorschlägen einschließlich Begründung unter Angabe entsprechender wissenschaftlicher Belege)

Der Vorbericht ist sehr ausführlich, differenziert und mit großer Sorgfalt und wissenschaftlicher Präzision erstellt worden. Nach dem Bericht zu Mindestmengen bei Knie-TEP wird hier für einen zweiten Leistungsbereich sehr deutlich gezeigt, dass die Hypothese eines Zusammenhangs zwischen der erbrachten Leistungsmenge und der Ergebnisqualität anhand deutscher Daten nicht bestätigt werden kann. Vielmehr wird im Vorbericht für die Koronarchirurgie nachgewiesen, dass ein initial scheinbarer Zusammenhang praktisch vollständig durch Confounding erklärt werden kann und der ohnehin niedrige Erklärungswert der Fallzahl verschwindet. Ein Einfluss der erbrachten Leistungsmenge auf entscheidende Qualitätsparameter wie Letalität und tiefe sternale Infektion oder auf die untersuchte Prozessqualität kann anhand deutscher Daten nicht nachgewiesen werden.

Zur Diskussion:

Hier werden von den Autoren zwei Aspekte zu den zugrunde liegenden Daten kritisch angemerkt, die aus unserer Sicht unzureichend bewertet werden:

1) Die Daten der BQS seien nicht für den Zweck der vorliegenden Untersuchung erhoben worden; dieses schränke die Aussagefähigkeit ein.

Während wir der im ersten Halbsatz getroffenen Einschätzung der Autoren zustimmen, halten wir die Beurteilung dieses Sachverhaltes im Kontext der zurzeit im IQWiG untersuchten Mindestmengen für unzureichend: Im Vergleich mit in anderen Untersuchungen verwendeten Daten aus der Literatur (zum Teil Abrechnungsdaten anderer Länder, Medicare-Daten etc.) sind BQS-Daten zum Zwecke der Qualitätssicherung erhoben worden (ebenfalls Gegenstand des Vorberichts). Sie vermeiden darüber hinaus das Problem der fragwürdigen Übertragung von Ergebnissen aus anderen Ländern auf deutsche Verhältnisse. Insofern haben die BQS-Daten eine viel höhere Aussagekraft.

2) Die Daten seien von „fragwürdiger Qualität“; als Hinweis wird ein stichprobenartiger Datenabgleich mit 685 Krankenakten aus 40 (von insgesamt 2200) Krankenhäusern angegeben.

Während wir hier die Einschätzung der Autoren grundsätzlich teilen, dass eine unklare Validität der Daten zur Unsicherheit der Ergebnisse des Vorberichts beitragen kann, so halten wir jedoch die Interpretation der Autoren für unzureichend. Insbesondere die hier als Hinweis angeführte Stichprobenvalidierung ist nach wissenschaftlichen Kriterien ungenügend beurteilt und entsprechend überbewertet. Die angeführte Stichprobe wurde im Rahmen eines Probetriebs zur Datenvalidierung im BQS-Verfahren durchgeführt. Entsprechend ist die Stichprobe mit 40 Krankenhäusern (von etwa 2200 Krankenhäusern bundesweit) aus drei Bundesländern nicht repräsentativ. Darüber hinaus schränkt die geringe Zahl von 685 Akten im Vergleich zu beispielsweise 54335 Patienten, die 2004, allein im Bereich Koronarchirurgie in den BQS-Datenpool eingegangen sind, die Aussagefähigkeit dieser Stichprobenvalidierung weiter ein. Unabhängig davon findet die Stichprobenuntersuchung hinsichtlich der Komplikationsdaten neben der von den Autoren angeführten geringen „Sensitivität“ und dem niedrigen „positiven prädiktiven Wert“ auch eine von den Autoren jedoch nicht angeführte hohe „Spezifität“ von 98,1% und einen hohen „negativen prädiktiven Wert“ von 98,3%. Schließlich ist eine Vergleichbarkeit der in der Stichprobe untersuchten „behandlungsbedürftigen chirurgischen und allgemeinen Komplikationen“, die auch zahlreiche kleinere Komplikationen umfassen dürften, und der im Vorbericht untersuchten schwersten Komplikationen der tiefen sternalen Infektion und der Letalität nicht gegeben. Man muss davon ausgehen, dass die Übereinstimmung von Krankenakte und BQS-Daten beispielsweise im Falle des Todes eines Patienten unvergleichlich besser ist als für die in der Stichprobe berichteten Komplikationen. Insofern wird der Datenvalidität durch die Stichprobe im Probetrieb ein zu schlechtes Ergebnis ausgestellt. Die Argumentationsweise und Bewertung der Autoren hinsichtlich der Validität der untersuchten BQS-Daten entspricht nicht dem hohen wissenschaftlichen Standard des Vorberichts.

In der Diskussion fehlt eine Würdigung der vorgenommenen Sensitivitätsanalysen, die zur Erhöhung der Aussagekraft der Ergebnisse des Vorberichts beitragen.

Zusammenfassend ist jedoch aus Sicht der DKG die hohe wissenschaftliche Qualität des Vorberichts hervorzuheben.

Wir stimmen vor diesem Hintergrund der Schlussfolgerung der Autoren zu, dass sich aus den Ergebnissen der vorliegenden Untersuchung kein Zusammenhang zwischen der erbrachten Leistungsmenge und der Ergebnisqualität in Deutschland für Patienten mit koronarchirurgischen Erst-Operationen und sich insbesondere keine Hinweise auf eine Ergebnisqualitätsverbesserung durch Einführung einer Mindestmenge ableiten lässt.




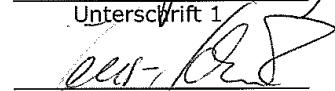
Substanzielle Stellungnahmen, die der Form genügen und rechtzeitig eingereicht werden, finden Berücksichtigung. Anhand folgender Checkliste können Sie prüfen, ob Ihre Stellungnahme die formalen Kriterien erfüllt. Bitte beachten Sie darüber hinaus die Hinweise im „Leitfaden zur Abgabe von Stellungnahmen“.

- Der Umfang der Stellungnahme umfasst maximal **6 DIN A4-Seiten**.
- Für alle Zitate sind **Originalpublikationen (Volltext)** beigelegt.
- Das Formblatt zur Stellungnahme ist vollständig ausgefüllt und **von allen Stellungnehmenden** unterschrieben.
- Das **Formblatt zur Darlegung potenzieller Interessenskonflikte** liegt **für alle Stellungnehmenden** einzeln **ausgefüllt** vor.
(Anm.: Dies gilt auch für Stellungnahmen von Institutionen / Gesellschaften)

Mir / uns ist bekannt, dass alle Stellungnahmen im Internet veröffentlicht werden können. Einer solchen Veröffentlichung stimme ich / stimmen wir mit meiner / unseren Unterschrift(en) ausdrücklich zu.

Berlin, den 29.05.2006
Ort/Datum


Unterschrift 1


Unterschrift 2

Unterschrift 3

Vielen Dank für Ihre Stellungnahme.

A.2 Stellungnahme der BÄK



Formblatt Stellungnahme zum Vorbericht

Auftrag des IQWiG/Projektbezeichnung mit Auftragsnummer:

Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für die Koronarchirurgie – Vorbericht – Auftrag B05-01b, Version 1.0

Name, Vorname; Titel und Funktion des/der Stellungnehmenden:

1. Klakow-Franck, Regina, Dr.med. M.A., Dezentin, Dezernat Qualitätssicherung
2. Wetzel, Hermann, Dr.med. M.Sc., Referent, Dezernat Gebührenordnung
- 3.

Kontakt für weitere Nachfragen:

Institution: Bundesärztekammer
Anschrift: Herbert-Lewin-Platz 1, 10623 Berlin
Tel./Fax-Nr.: 030 / 400456 – 435 oder – 443
E-Mail-Adresse: Regina.Klakow-Franck@baek.de oder Hermann.Wetzel@baek.de

1. Folgende Originalstudien fehlen im Vorbericht

Der G-BA hat am 21.09.2004 beschlossen, koronarchirurgische Eingriffe ohne konkrete Fallzahlgrenzwerte in den Mindestmengenkatalog gemäß § 137 Abs. 1 Nr. 3 SGB V aufzunehmen. Die Bundesärztekammer hat sich in den vorangegangenen G-BA-Beratungen gegen diese Einschätzung gewandt, da vor dem Hintergrund einer inkonsistenten Ergebnislage in der internationalen Literatur bei suffizienter Datenanalyse ein klinisch relevanter Zusammenhang zwischen Fallzahl und Ergebnisqualität evidenzbasiert nicht bestätigt werden konnte, kontextabhängige Ergebnisse insbesondere aus angloamerikanischen Ländern schwerlich auf die Versorgung in Deutschland übertragen werden können, und die Vorhersagekraft von Mindestmengen als Surrogat-Indikatoren für Ergebnisqualität als unzureichend angesehen werden muss.

Es wurde hierbei insbesondere auf die bis dahin methodisch beste Untersuchung von Peterson ED, Coombs LP, DeLong ER, Haan CK, Ferguson MD (2004) Procedural volume as a marker of quality for CABG surgery. JAMA 291: 195–201, verwiesen. In dieser Arbeit wurde zwar eine statistisch signifikante Assoziation zwischen Leistungsmenge je Krankenhaus und risikoadjustierter Letalität gefunden, jedoch von nur sehr geringem Ausmaß (OR: 0,98; 95%KI: 0,96; 0,99; Letalitätsabnahme von 0,07% je 100 zusätzliche Eingriffe) und lediglich für Patienten mit höherem Alter bzw. höherem Risiko. Eine hinreichend genaue Prädiktion unterschiedlicher Letalitätsraten war anhand von Leistungsmengen aber nicht möglich, so dass die Fallzahl nicht als geeigneter, klinisch relevanter Qualitätsindikator angesehen werden konnte.

Nach der vorliegenden Auswertung durch das IQWiG kann nun anhand deutscher Daten aus dem BQS-Bundesdatenpool weder für den primären Qualitätsindikator "Letalität" noch für den sekundären Qualitätsindikator "sternale Infektion" nach Berücksichtigung relevanter Confounder ein signifikanter statistischer Zusammenhang mit der Leistungsmenge nachgewiesen werden; eine Schwellenwertberechnung wurde daher für diese Ergebnis-Qualitätsindikatoren gar nicht erst durchgeführt. Für beide ausgewertete Prozess-Qualitätsindikatoren "Verwendung der Brustwandarterie" und "Anzahl peripherer Anastomosen" ergab sich nach Berücksichtigung relevanter Confounder und Cluster-Effekte ebenfalls kein signifikanter statistischer Zusammenhang mit der Leistungsmenge. Die vorliegende IQWiG-Analyse bestätigt – nunmehr erstmals auf der Grundlage deutscher Daten – nachdrücklich die Zweifel an der Richtigkeit der Entscheidung des G-BA, koronarchirurgische Eingriffe in den Mindestmengenkatalog aufzunehmen.

Auch wenn eine aktualisierte Literatur-Auswertung mit diesem Bericht nicht beabsichtigt sein dürfte, sollte u.E. im Rahmen des Diskussions-Abschnitts unbedingt darauf hingewiesen werden, dass die vorliegenden Ergebnisse, wonach für koronarchirurgische Eingriffen anhand deutscher Daten keine signifikante Assoziation zwischen Leistungsmenge und Ergebnisqualität belegt werden konnte, keineswegs im Widerspruch zur internationalen Literatur stehen. Auch in den folgenden, im letzten Jahr erschienenen Publikationen konnte beispielsweise eine Mengen-Qualitäts-Beziehung bei koronarchirurgischen Eingriffen nicht nachgewiesen werden:

- Plomondon ME, Casebeer AW, Schooley LM, Wagner BD, Grunwald GK, McDonald GO, Grover FL, Shroyer AL (2006) Exploring the volume-outcome relationship for off-pump coronary artery bypass graft procedures. Ann Thorac Surg 81:547-553
- Welke KF, Barnett MJ, Sarrazin MS, Rosenthal GE (2005) Limitations of hospital volume as a measure of quality of care for coronary artery bypass graft surgery. Ann Thorac Surg 80:2114-2119
- Zacharias A, Schwann TA, Riordan CJ, Durham SJ, Shah A, Papadimos TJ, Engoren M, Habib RH (2005) Is hospital procedure volume a reliable marker of quality for coronary artery bypass surgery? A comparison of risk and propensity adjusted operative and midterm outcomes. Ann Thorac Surg 79:1961-1969

Version 2, Stand: 11/2005

Welke et al. (2005) fassen ihre Ergebnisse wie folgt prägnant zusammen: "Volume alone, as a discriminator of mortality, is only slightly better than a coin flip."

Die o.g. Publikationen könnten in der Diskussion der Ergebnisse Berücksichtigung finden.

2. Folgende Bewertungen (von Originalstudien) im Vorbericht sind nicht korrekt (inkl. Begründung):

Im Diskussions-Abschnitt wurde darauf verwiesen, dass die ausgewerteten BQS-Daten nicht für den Zweck des Berichts, nämlich die Untersuchung von Volume-Outcome-Beziehungen bei der Koronarchirurgie erhoben wurden, und dass die Qualität der Daten angesichts mangelnder Überprüfbarkeit ihrer Validität als unklar angesehen werden müsse. Vor dem Hintergrund eines BQS-Vortrags auf einem Workshop im Februar 2006 wird dargelegt, anhand einer Stichprobe von 685 Krankenakten habe sich gezeigt, dass in den BQS-Daten ca. 34% der Komplikationen aus den Krankenakten fehlten und ca. 37% der an die BQS berichteten Komplikationen in den Krankenakten nicht dokumentiert worden seien. U.E. ist der Verweis auf die Bedeutung von Datenqualität und Datenvalidierung prinzipiell berechtigt, die nachfolgenden Ausführungen zur Datenberichtsqualität in der Koronarchirurgie und mögliche Auswirkungen auf die Ergebnisse des vorliegenden Berichts sind u.E. jedoch spekulativ. Nach unserer Einschätzung dürften die o.g. ungenügenden Konkordanzraten nicht ohne weiteres auf den hier in Frage stehenden Bereich der Koronarchirurgie zu übertragen sein; vielmehr erscheint uns in der Koronarchirurgie generell und insbesondere hinsichtlich der in der vorliegenden Analyse ausgewerteten Ergebnis- und Prozessqualitäts-Indikatoren eine unzureichende Berichtsqualität wie angegeben sehr unwahrscheinlich. Im speziellen ist nicht ersichtlich, dass im Hinblick auf den primären Qualitätsindikator "Letalität" (oder auch den sekundären Qualitätsindikator "tiefe sternale Infektion") mögliche Mängel in der Berichtsqualität zu einer systematischen Verzerrung von derart gravierendem Ausmaß hätten führen können, dass die Schlussfolgerungen des Berichts eines fehlenden Zusammenhangs zwischen Leistungsmenge und Ergebnisqualität dadurch invalidiert würden.

Sollten hinsichtlich der Güte der Datenberichterstattung der in die Auswertung einbezogenen koronarchirurgischen Abteilungen und Kliniken tatsächlich substantielle Zweifel bestehen, so wäre eine gesonderte Auswertung des Datenabgleichs aus dem genannten BQS-Vortrag hinsichtlich der Untergruppe der koronarchirurgischen Zentren anzuregen, oder die BQS sollte um einen Zwischenbericht zum derzeit durchgeführten Datenvalidierungsverfahren zur isolierten Koronarchirurgie gebeten werden.

U.E. wäre es eine im Rahmen der evidenzbasierten Medizin unzulässige erkenntnistheoretische Argumentation, wenn auf der Basis des vorliegenden IQWiG-Berichts von Befürwortern von Mindestmengen bei koronarchirurgischen Eingriffen die irrige Behauptung aufgestellt würde, aufgrund eines vorgeblich nicht ausreichenden Datenmaterials habe eine tatsächlich existierende Mindestmenge nicht verifiziert werden können, der negative

Version 2, Stand: 11/2005

Befund des vorliegenden IQWiG-Berichts einer fehlenden Mengen-Qualitäts-Beziehung erlaube folglich keine gesicherte Aussage darüber, ob nicht doch auch in deutschen Daten eine Mindestmenge verborgen sei, und daher könne man an der eingeführten Mindestmengenregelung für koronarchirurgische Eingriffe auch weiter festhalten. Gegenüber einer solcherart kontrafaktischen Argumentation ist festzuhalten, dass sich mittels einer qualitativ hochwertigen statistischen Methodik auf der Basis verfügbarer deutscher Daten – also nach der derzeit bestverfügbaren Evidenz – weder eine signifikante Assoziation zwischen Leistungsmenge und Ergebnisqualität nachweisen noch gar ein Schwellenwert i.S.e. Mindestmenge bei koronarchirurgischen Eingriffen ableiten ließ.

Wir stimmen im übrigen der Einschätzung nachdrücklich zu, dass man, selbst wenn in der vorliegenden Auswertung ein signifikanter Zusammenhang zwischen Fallzahl und Ergebnisqualität gefunden worden und darüber hinaus ein Schwellenwert ableitbar gewesen wäre, nicht davon ausgehen könnte, die Verwendung eines solchen Schwellenwerts als Mindestmenge würde zu der erwünschten Qualitätsverbesserung führen. Vielmehr kann – wie die Autoren des IQWiG-Berichts sehr zu Recht betonen – ein solcher Beleg zuverlässig nur über eine kontrollierte Interventionsstudie und nicht über den Nachweis einer statistischen Assoziation aus Routinedaten geführt werden.

3. Anmerkungen zur projektspezifischen Methodik (bei Änderungsvorschlägen einschließlich Begründung unter Angabe entsprechender wissenschaftlicher Belege):

Nach unserer Auffassung handelt es sich bei dem vorliegenden Bericht aufgrund der methodischen Vorgehensweise hinsichtlich Faktoren wie Risiko- und Case-Mix-Adjustierung, Validierung bzw. Überprüfung der Güte von Regressionsmodellen, Berücksichtigung möglicher Clustereffekte, ergänzende Sensitivitätsanalysen, transparente und nachvollziehbare Berichterstattung der Ergebnisse etc. um eine statistische Analyse von hoher Qualität, welche in dieser Hinsicht durchaus mit der o.g. JAMA-Arbeit von Peterson et al. (2004) auf eine Stufe zu stellen ist, und die daher auch international publiziert werden sollte.

Ergänzend wäre zu diskutieren, ob nicht hinsichtlich der Unterscheidung von planbaren versus dringlichen Eingriffen eine stratifizierte bzw. Subgruppenanalyse durchgeführt werden sollte, so dass eine Darstellung der Ergebnisqualität für planbare Eingriffe ohne Vermischung mit anderen Dringlichkeitsgraden erfolgen kann. Auch bei Berücksichtigung des Dringlichkeitsgrads eines Eingriffs in einer Risikoadjustierung bzw. als zusätzliche Kovariable würden sich die Ergebnisse auf die Gesamtheit der Prozeduren und nicht nur auf die Unterkategorie der planbaren Eingriffe beziehen.



Substantielle Stellungnahmen, die der Form genügen und rechtzeitig eingereicht werden, finden Berücksichtigung. Anhand folgender Checkliste können Sie prüfen, ob Ihre Stellungnahme die formalen Kriterien erfüllt. Bitte beachten Sie darüber hinaus die Hinweise im „Leitfaden zur Abgabe von Stellungnahmen“.

Der Umfang der Stellungnahme umfasst maximal **6 DIN A4-Seiten**.


Für alle Zitate sind **Originalpublikationen (Volltext)** beigelegt.

Das Formblatt zur Stellungnahme ist vollständig ausgefüllt und **von allen Stellungnehmenden** unterschrieben.

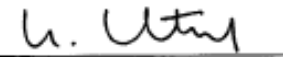
Das **Formblatt zur Darlegung potenzieller Interessenskonflikte** liegt **für alle Stellungnehmenden** einzeln **ausgefüllt** vor.
 (Anm.: Dies gilt auch für Stellungnahmen von Institutionen / Gesellschaften)

Mir / uns ist bekannt, dass alle Stellungnahmen im Internet veröffentlicht werden können. Einer solchen Veröffentlichung stimme ich / stimmen wir mit meiner / unseren Unterschrift(en) ausdrücklich zu.

Berlin, den 30.05.06
 Ort/Datum



Unterschrift 1



Unterschrift 2

 Unterschrift 3

Vielen Dank für Ihre Stellungnahme.

A.3 Stellungnahme der BQS



Eingangsstempel

Formblatt zur Anhörung (schriftlichen Stellungnahme) zum Vorbericht

Auftrag des IQWiG/Projektbezeichnung mit Auftragsnummer:

Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei
Mindestmengen für die Koronarchirurgie (Version 1.0 vom 24.04.2006 – B05/01b)

Name, Vorname; Titel und Funktion des/der Stellungnehmenden:

1. Fischer, Burkhard
Diplom-Informatiker der Medizin
Abteilungsleiter Medizinische Biometrie
2. Kötting, Joachim
Diplom-Statistiker
Projektleiter Medizinische Biometrie
3. Thölen, Frank
Dr. med., Facharzt für Anästhesiologie
Projektleiter Medizin und Pflege
4. Mohr, Volker D.
Dr. med., Facharzt für Chirurgie – Viszeralchirurgie und Intensivmedizin;
Diplom-Gesundheitsökonom
Geschäftsführer

Kontakt für weitere Nachfragen:

Name: Fischer, Burkhard
 Institution: BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH
 Anschrift: Kanzlerstr. 4, D-40472 Düsseldorf
 Tel./Fax-Nr.:
 +49 211 280729 0
 +49 211 280729 99
 E-Mail-Adresse:
 info@bqs-online.de



Generelle Stellungnahme

In Kenntnis der US-amerikanischen Publikationen zur Fallzahlabhängigkeit von Ergebnisqualität und der deutschen Krankenhauslandschaft, die vergleichsweise höhere Fallzahlen pro Krankenhaus im Leistungsbereich Koronarchirurgie aufweist, waren die Ergebnisse des Vorberichts zu erwarten. Im Verfahrensjahr 2004 wurden in jeder deutschen koronarchirurgischen Abteilung über 150 isolierte koronarchirurgische Eingriffe vorgenommen.

Die Risikoadjustierung erscheint nicht sachgerecht. Der EuroSCORE und auch der KCH-Score sind jeweils für die Krankenhaus und 30-Tage-Letalität validiert, jedoch nicht für die Wundinfektion. Er wird aber auch dort verwendet. Der NNIS (National Nosocomial Infection Score; CDC-Score) wurde für den Endpunkt nosokomiale Wundinfektion validiert, jedoch nicht für Letalität. Die Festlegungen der Fachexperten aus Arbeitsschritt 2 "Qualitätsindikatoren und Einflussfaktoren festlegen durch Fachexperten mit BQS" und IQWiG im Berichtsplan [IQWiG 2005] sahen eine spezifische Risikoadjustierung vor. Hierzu waren für den jeweiligen Endpunkt validierte Risikoscores oder –faktoren vorgesehen (Tabelle 1 und Anlage 1). Den Stellungnehmenden ist nicht bekannt, inwieweit das im Bericht beschriebene geänderte Vorgehen auf geänderte Festlegungen der Fachexperten begründet ist.

Tabelle 1: Ausschnitt aus den Festlegungen der Fachexperten (vgl. Anlage 1: Festlegungen der Fachexperten vom 12.08.2006 in der Fassung vom 29.08.2006)

Endpunkt	Qualitätsindikator	Risikoadjustierung	Weitere diskutierte Einflussfaktoren zur Risikoadjustierung, die nicht berücksichtigt werden
Letalität	In-Hospital-Letalität	a) EuroSCORE b) KCH-Score	Einzelfaktoren beider Scores
tiefe sternale Infektion	Postoperative Mediastinitis	ohne Notfälle (Risikostandardisierte Fallkonstellation) und nur Patienten mit CDC-NNIS-Risikoklasse 0 (Additiver Score)	-
Verwendung der inneren Brustwandarterie	Verwendung der linksseitigen Arteria mammaria interna	Risikostandardisierte Fallkonstellation: ohne Notfälle und ohne vorausgegangene Koronarchirurgie	ausgeprägte Adipositas
Anzahl der peripheren Anastomosen	Anzahl peripherer Anastomosen	Keine Risikoadjustierung	ausgeprägte Adipositas

Kritisch anzumerken ist, dass nur die Eignung der Datenbasis und der gewählten Qualitätsindikatoren diskutiert wird, nicht jedoch die Eignung der gewählten Methoden. Wünschenswert wäre auch eine weitergehende inhaltliche Auseinandersetzung mit den resultierenden Modellen im Bericht.



Fragen an die Autoren des Vorberichts und Hinweise auf Erklärungsbedarf

In folgenden Punkten haben sich bei der Durchsicht des Vorberichts Fragen ergeben, die der Text des Vorberichtes nicht oder nicht ausreichend beantwortet. Es werden Vorschläge für den abschließenden Berichtes gemacht.

Wo?	Was?	Vorschlag
Seite 9	Die Vorgaben der Fachexperten vom 12.08.2005 werden im Berichtsplan und Vorbericht nicht dargestellt (Anlage 1).	Im Bericht sollten die Festlegungen vom 12.08.2005 in der Fassung vom 29.08.2005 dargestellt und das geänderte Vorgehen begründet werden.
Seite 10	„Für Untersuchungen zur Modellstabilität wurden auch die Daten ... aus dem Bundesdatenpool 2003 verwendet.“ 2003er-Daten waren in der ursprünglichen Planung nicht nur zur Untersuchung der Modellstabilität vorgesehen. Vielmehr sollten diese Daten ursprünglich zur Festlegung der Vorgaben zur Schwellenwertbestimmung (relevante Unterschiede und Modellwahl) durch die Fachexperten herangezogen werden (Punkte 4 – 8 des Projektablaufs). Es bleibt unklar ob und wie die 2003er Daten hierzu verwendet wurden.	Im Bericht erläutern, ob und wie die 2003er Daten zur Festlegung der der Vorgaben für die Schwellenwertbestimmung (relevante Unterschiede und Modellwahl) durch die Fachexperten herangezogen wurden.
Seite 13	„Die Ergebnisse 2003 haben jedoch aufgrund der vielen fehlenden Werte beim EuroSCORE nur geringe Aussagekraft.“ Die ca. 20.000 verwendeten Datensätze haben nach Maßstäben in der Statistik keinesfalls eine geringe Aussagekraft. Eine durch fehlende Werte bedingte Ergebnisverzerrung ist nicht offensichtlich.	Diese Aussage sollte gestrichen werden.



Wo?	Was?	Vorschlag
Seite 14	<p>4.4 „geringfügige Änderungen bzw. Ergänzungen“</p> <p>Es bleibt unklar, inwieweit Änderungen gegenüber dem Berichtsplan mit den Fachexperten oder dem Auftraggeber abgestimmt wurden? Mit der Organisation der Stellungnehmenden, die die Fachexperten bei der Auswahl der Qualitätsindikatoren und Risikofaktoren unterstützte, sind keine Abstimmungen nach dem 12.08.2005 erfolgt.</p> <p>Die zusätzliche Aufnahme von (für die Endpunkte nicht validierten) Risikofaktoren stellt alles andere als eine geringfügige Änderung dar. Dadurch ändert sich auch die inhaltliche Bedeutung und Interpretation der Modelle.</p>	<p>Änderungen gegenüber den ursprünglichen Festlegungen und dem Berichtsplan sollten dargestellt und begründet werden.</p> <p>Es sollten Aussagen gemacht werden, warum eine Erweiterung der Risikomodellierung vorgenommen wurde und welche Auswirkungen dies auf die Modelle und ihre Interpretation hat.</p>
Seite 20	CDC-Klasse war nicht als Risikofaktor für die Letalität nicht vorgesehen.	
Seite 21	Der EuroSCORE ist das Ergebnis einer logistischen Regression. Er wird ohne weitere Diskussion als Einflussfaktor einer erneuten logistischen Regression verwendet.	Im Bericht sollte dieser Punkt erläutert werden und dargestellt sein, warum die Autoren dieses Vorgehen für angemessen halten, bzw. welche möglichen Auswirkungen oder Einschränkungen sich für die Interpretation der Ergebnisse daraus ableiten.
Seite 27	Abweichung von den ursprünglichen Festlegungen: erwartete Versterbenswahrscheinlichkeit als Risikofaktor für die Infektionsrate verwendet. Die Versterbenswahrscheinlichkeit hängt aber eher von der Infektionsrate ab als umgekehrt.	Entsprechend der ursprünglichen Festlegungen rechnen oder abweichendes Vorgehen begründen.
Seite 33	Abweichung von den ursprünglichen Festlegungen: CDC-Klasse und EuroSCORE als Risikofaktor für Prozessindikator Brustwandarterie eher ungeeignet.	Entsprechend der ursprünglichen Festlegungen rechnen oder abweichendes Vorgehen begründen.



Wo?	Was?	Vorschlag
Seite 38	Abweichung von den ursprünglichen Festlegungen: Anastomosen sollten ursprünglich ohne Risikoadjustierung gerechnet werden.	Entsprechend der ursprünglichen Festlegungen rechnen oder abweichendes Vorgehen begründen.
Seite 43	Die Autoren diskutieren die Datenbasis. Hierbei zitieren sie Ergebnisse des Probebetriebs für den Leistungsbereich Hüft-TEP-Erstimplantation. Leider jedoch selektiv, sodass der falsche Eindruck entsteht, dass die Übereinstimmung zwischen Patientenakte und QS-Dokumentation lediglich ca. 65% beträgt. Auch werden keine Vergleichszahlen aus der Literatur genannt, die dem Leser eine Bewertung der Zahlen ermöglichen würde.	<p>Im Rahmen der Erprobung eines Datenvalidierungsverfahrens wurde die Übereinstimmung zwischen Patientenakte und QS-Dokumentation untersucht [Fischer et al. 2006].</p> <p>Die Untersuchung ergab eine Übereinstimmung von 95,7% (17.536 von 18.317 geprüften Items). Bei Komplikationsitems* betrug diese Übereinstimmung 96,6% (2.629 von 2.722).</p> <p>Betrachtet man weiter nur die Komplikationen, die in mindestens einer der beiden Dokumentation zu finden waren, so fanden sich 65,6% (84 von 128) der Komplikationen* in der Patientenakte auch in der QS-Dokumentation und 63,8% (84 von 133) der Komplikationen* in der QS-Dokumentation in der Patientenakte.</p> <p>* behandlungsbedürftige chirurgische und allgemeine Komplikationen, Wundinfektion nach CDC, Reintervention aufgrund Komplikationen</p>
Seite 44	<p>Es erscheint nicht sachgerecht, bei Nichterfüllung von Maximalbedingungen (Interventionsstudie) eine Entscheidungsfindung auszuschließen.</p> <p>Es bleibt zudem offen, wie das Design einer kontrollierten Interventionsstudie aussehen würde, die zur Berechnung von Schwellenwerten verwendet werden kann.</p>	Das Design einer möglichen Interventionsstudie beschreiben und die Möglichkeiten der Umsetzung (Aufwand und Ethik) kurz diskutieren. Vor- und Nachteile des Ansatzes gegenüber einer Beobachtungsstudie mit Quasi-Vollerhebung diskutieren.



Referenzen

[BQS 2005a] BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH: Qualitätsindikator "Verwendung der linksseitigen Arteria mammaria interna" (Koronarchirurgie, isoliert, Verfahrensjahr 2004). Online verfügbar unter: http://www.bqs-qualitaetsindikatoren.de/2004/ergebnisse/leistungsbereiche/HCH_KCH/indikatoren/1/
[Recherchedatum: 30.05.2006]

[BQS 2005b] BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH: Qualitätsindikator "Letalität" (Koronarchirurgie, isoliert, Verfahrensjahr 2004). Online verfügbar unter: http://www.bqs-qualitaetsindikatoren.de/2004/ergebnisse/leistungsbereiche/HCH_KCH/indikatoren/2/
[Recherchedatum: 30.05.2006]

[BQS2005c] BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH: Qualitätsindikator "Postoperative Beweglichkeit" (Koronarchirurgie, isoliert, Verfahrensjahr 2004). Online verfügbar unter: http://www.bqs-qualitaetsindikatoren.de/2004/ergebnisse/leistungsbereiche/HCH_KCH/indikatoren/5/
[Recherchedatum: 30.05.2006]

[Fischer et al. 2006] Fischer B, Projektgruppe Datenvalidierung: Datenvalidierung im BQS-Verfahren: Ergebnisse des Probebetriebs. Vortrags im Rahmen des GMDS-Workshops „Qualität medizinischer Daten in Forschung und Versorgung“ vom 16. - 17.02.2006 in Essen. Online verfügbar unter: http://www.ekmed.de/routinedaten/download/workshop_2006/fischer.pdf [Recherchedatum: 30.05.2006]

[IQWiG 2005] Berichtsplan zum Auftrag "Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für die Koronarchirurgie" [Auftrag B05/01b] Version 1.0 vom 05.09.2005. Köln: Institut für Qualität und Wirtschaftlichkeit in der Medizin.



Anlage 1

Endpunkte für Qualität und ihre Risikoadjustierung für Mindestmengen in der Koronarchirurgie: Konkretisierung der Rechenregeln und der zu verwendenden Daten (Datenbasis: BQS-Bundesdatenpool 2003 6.0 und 2004)

2005-08-12 BQS/Fischer
2005-06-29 BQS/Fischer (mit Korrektur der Rechenregel für die tiefe sternale Infektion)

Fassung nach Abstimmung mit Dr. Spitzenfell (2005-08-11) und Prof. Hammel (2005-08-12).

Endpunkt	Indikatortyp	Rangfolge (Klinische Relevanz)	Qualitätsindikator	Risikoadjustierung	Weitere deklarierte Einflussfaktoren zur Risikoadjustierung, die nicht berücksichtigt werden*	Rechenregel	Maßzeitpunkt
Letalität	Ergebnisqualität	1	In-Hospital-Letalität	c) EuroSCORE d) KCH-Score	Einzelaktoren beider Scores	BQS-Qualitätsindikatoren 2004: Kennzahl 66772 BQS-Qualitätsindikatoren 2004: Kennzahl 66781	Postoperativ bis zur Krankenhausentlassung
Tiefe sternale Infektion	Ergebnisqualität	2	Postoperative Mediastinitis	ohne Norfalle (Risikostratifizierte Faltskategorien) und nur Patienten mit CDC-NNIS-Risikoklasse 0 (Additiver Score)	-	Zähler: Patienten mit Mediastinitis oder tiefer Wundinfektion des Thorax als postoperativer Wundinfektion, Sternumrevision oder Re thorakotomie wegen Mediastinitis Grundgesamtheit: Alle Patienten mit CDC-Risikoklasse 0, die in ihrer ersten Operation elektiv oder dringlich isoliert koronarchirurgisch versorgt wurden ohne präoperative Mediastinitis (modifiziert nach BQS-Qualitätsindikatoren 2004)	Postoperativ bis zur Krankenhausentlassung
Verwendung der inneren Brustwandarterie	Prozessqualität	-	Verwendung der linksseitigen Arteria mamma interna	Risikostratifizierte Faltskategorien: ohne Norfalle und ohne vorausgegangene Koronarchirurgie	ausgeprägte Adipositas	Zähler: Verwendung der linksseitigen Arteria mamma interna als Graf (ITA links = Ja) Grundgesamtheit: Alle Patienten, die in ihrer ersten Operation elektiv oder dringlich isoliert koronarchirurgisch versorgt wurden mit Verwendung eines Graf's, ohne Patienten mit vorausgegangener Koronarchirurgie (modifiziert nach BQS-Qualitätsindikatoren 2004)	entfällt
Anzahl der peripheren Anastomosen	Prozessqualität	-	Anzahl peripherer Anastomosen	Keine Risikoadjustierung	ausgeprägte Adipositas	(modifiziert nach BQS-Qualitätsindikatoren 2004) periphere Anastomosenanteil = verofs In Klassen: 1, 2, > 2 Grundgesamtheit: alle Patienten	entfällt

* Quellen:
 • Hammel, Spitzenfell: Expertenvorschlag Mindestmengen Koronarchirurgie (19.04.2005)
 • BQS-Qualitätsindikatoren 2004: www.bqs-qualitaetsindikatoren.de
 • Beratung in der AG Mindestmengen



Substanzielle Stellungnahmen, die der Form genügen und rechtzeitig eingereicht werden, finden Berücksichtigung. Anhand folgender Checkliste können Sie prüfen, ob Ihre Stellungnahme die formalen Kriterien erfüllt. Bitte beachten Sie darüber hinaus die Hinweise im „Leitfaden zur Abgabe von Stellungnahmen“.

- Der Umfang der Stellungnahme umfasst maximal **5 DIN-A4-Seiten**.
- Der Umfang ggf. beigefügter Anhänge zur Stellungnahme umfasst maximal **10 DIN-A4-Seiten**.
- Für alle Zitate sind **Originalpublikationen (Volltext)** beigelegt.
(online verfügbar oder bei dem Autor des Vertriebs)
- Das Formblatt zur Stellungnahme ist vollständig ausgefüllt und **von allen Stellungnehmenden** unterschrieben.
- Das **Formblatt zur Darlegung potenzieller Interessenskonflikte** liegt **für alle Stellungnehmenden** einzeln **ausgefüllt** vor.
(Anm.: Dies gilt auch für Stellungnahmen von Institutionen / Gesellschaften)

Mir / uns ist bekannt, dass alle substanziellen Stellungnahmen im Rahmen des Abschlussberichts im Internet veröffentlicht werden können. Einer solchen Veröffentlichung stimme ich / stimmen wir mit meiner / unseren Unterschrift(en) ausdrücklich zu.

Düsseldorf 30.05.2006
Ort/Datum

B. Engel
Unterschrift 1
J. Schol
Unterschrift 2
J. Köhler
Unterschrift 3
H. ...

Vielen Dank für Ihre Stellungnahme.

A.4 Antwort auf die Stellungnahmen

A.4.1 Antwort auf die Stellungnahme der DKG

Selbstverständlich ist es richtig, dass für die vorliegenden Daten nicht wie bei internationalen Studienergebnissen das Problem einer fragwürdigen Übertragbarkeit auf Deutschland besteht. Dies wurde in unserem Bericht nicht besonders betont, weil es ja gerade ein Bestandteil des Auftrags war, Berechnungen *anhand deutscher Daten* durchzuführen. Wir haben die Diskussion bezüglich der Datenvalidierung im BQS-Verfahren entsprechend umformuliert, um missverständliche Interpretationen zu vermeiden.

A.4.2 Antwort auf die Stellungnahme der BÄK

Die Darstellung der aktuellen internationalen Literatur zum Thema war nicht Bestandteil des Auftrags. Wir haben die Diskussion bezüglich der Datenvalidierung im BQS-Verfahren entsprechend umformuliert, um missverständliche Interpretationen zu vermeiden. Wir haben ergänzend in den Modellen für die Indikatoren der Ergebnisqualität auch die Wechselwirkung zwischen Dringlichkeit und Fallzahl untersucht. Da der entsprechende Wechselwirkungsterm für beide Indikatoren statistisch nicht signifikant war, haben wir keine weiteren Subgruppenanalysen bezüglich planbarer und dringlicher Eingriffe durchgeführt.

A.4.3 Antwort auf die Stellungnahme der BQS

Die Durchführung der Risikoadjustierung wird als nicht sachgerecht bezeichnet. Als Begründung hierfür werden zwei Punkte angesprochen: 1) falsche Verwendung der Confounder in den Regressionsmodellen und 2) Abweichung von der Festlegung der Fachexperten am 12.08.2005 (Anlage zur Stellungnahme der BQS). Wir möchten zunächst auf den 2. Punkt eingehen.

Hier liegt ein grundsätzliches Missverständnis über den geplanten Umfang und die Bedeutung der Festlegungen der Fachexperten vor. Aufgrund des Beschlusses des G-BA vom 21.06.2005 durfte die BQS dem IQWiG nur die Daten liefern, die zur Bearbeitung des Auftrags notwendig sind. Aus diesem Grund mussten – in Absprache mit den Fachexperten – im Vorhinein neben den Qualitätsindikatoren auch die potentiellen Confounder festgelegt werden. Um den entsprechenden

Datensatz zu erzeugen, waren weitere Vorgaben zur Operationalisierung und technischen Umsetzung der beteiligten Datenfelder notwendig. Zu diesem Zweck diente die Absprache der BQS mit den Fachexperten im August 2005. Es war nicht Zweck dieser Absprache, von den Fachexperten eine prädefinierte Modellbildung durchzuführen und es war nie vorgesehen, dass sich das IQWiG an solche Vorgaben zu halten hat. Im Gegenteil: Die Modellbildung stellt einen wesentlichen methodischen Teil des Auftrags an das IQWiG dar und diesen Auftrag hat das IQWiG in unabhängiger Weise durchzuführen. Zudem waren die Vorgaben, so wie sie auf der Anlage der BQS zusammengefasst sind, nicht alle methodisch sinnvoll. Z.B. wurde vorgeschrieben, dass bei der Auswertung zum Qualitätsindikator "Infektion" nur Patienten mit CDC-Risikoklasse 0 berücksichtigt werden sollten. Dies wird fälschlicherweise als Risikoadjustierung bezeichnet, stellt jedoch tatsächlich eine bedeutsame Einschränkung der Stichprobe dar, auf die sich die Berechnung und folglich auch das Ergebnis beziehen. Es ist nicht sinnvoll und inhaltlich in keiner Weise zu rechtfertigen, Berechnungen zu Mindestmengen auf Patienten mit Risikoklasse 0 einzuschränken, obwohl von einer möglichen Mindestmengenregelung doch Patienten aller CDC-Risikoklassen betroffen sind. Von einer möglichen Mindestmengenregelung sind nur Notfälle nicht betroffen. Aus diesem Grund haben wir aus allen Regressionsanalysen Notfälle ausgeschlossen, dies aber für alle Qualitätsindikatoren und nicht nur für einen Teil der Qualitätsindikatoren wie es in der BQS-Anlage vorgesehen war.

Bezüglich des ersten Punkts, die Confounder seien in den Regressionsmodellen nicht korrekt verwendet worden, nehmen wir wie folgt Stellung. Die Argumente der BQS zeigen, dass bei den Stellungnehmenden ein grundsätzliches Missverständnis über den Sinn und Zweck der durchgeführten Regressionsmodelle vorliegt. Z.B. das Argument, ein bestimmter Score sei für einen bestimmten Qualitätsindikator nicht validiert, spielt nur dann eine Rolle, wenn man diesen Score zur optimalen Prognose dieses Indikators einsetzen will. Das ist aber bei den hier durchgeführten Regressionsanalysen gar nicht der Fall. Ziel der im Bericht durchgeführten Regressionsanalysen ist es festzustellen, ob es zwischen der Fallzahl und dem jeweiligen Qualitätsindikator einen Zusammenhang gibt. Da

der Case-Mix der Patienten in den Krankenhäusern unterschiedlich sein kann, ist es notwendig, wichtige Confounder in den Analysen zu berücksichtigen. Als Confounder sind insbesondere solche Variablen von Bedeutung, die das Risikoprofil der Patienten beschreiben. Anstelle vieler einzelner Variablen wie Alter, Geschlecht, pulmonale Erkrankung, arterielle Verschlusskrankheit usw. wurde hier primär der international gebräuchliche Euro-Score verwendet, alternativ dazu der von der BQS entwickelte KCH-Score betrachtet und ergänzend wurden auch andere Variablen untersucht, die im Euro-Score nicht enthalten sind (z.B. BMI) und daher möglicherweise zusätzlich zum Euro-Score als Confounder wichtig sind. Die Tatsache, dass die von uns betrachteten Confounder in allen Regressionsmodellen – mit Ausnahme der CDC-Klasse für den Indikator "Brustwandarterie" – signifikante Effekte zeigten, unterstreicht die Bedeutung und Richtigkeit dieser Vorgehensweise für den vorliegenden Zweck. Die wissenschaftlich korrekte Durchführung sowie die hohe Qualität unserer Auswertungen wurden in der Beurteilung sowohl des externen Reviewers als auch in den anderen Stellungnahmen unterstrichen, so dass wir keine Notwendigkeit gesehen haben, unsere regressionsanalytischen Auswertungen grundlegend zu modifizieren.

Zum Schluss möchten wir noch kurz auf die einzelnen Fragen und Hinweise der BQS eingehen.

Zur Frage bezüglich S. 9: Alle relevanten Vorgaben der Fachexperten wurden im Vorbericht dargestellt (S. 9-10 des Vorberichts). Im Übrigen verweisen wir auf unsere Antwort auf die generelle Stellungnahme oben.

Zur Frage bezüglich S. 10: Der Text wurde um den Hinweis ergänzt, dass die Daten von 2003 auch zur Festlegung der klinisch relevanten Veränderungen durch die Fachexperten verwendet wurden.

Zur Frage bezüglich S. 13: Da der Mechanismus, der zum Fehlen von fast 50% der Euro-Score-Werte geführt hat, unbekannt ist und man nicht ohne Weiteres davon ausgehen kann, dass die Werte völlig zufällig fehlen, haben die Ergebnisse der multiplen Regressionsmodelle für 2003 wie beschrieben nur geringe Aussagekraft. Das Problem ist nicht eine unzureichende statistische Power, sondern die Möglichkeit einer starken Verzerrung durch fehlende Werte.

Zu Fragen bezüglich S. 14-38: Wir verweisen auf unsere Antwort auf die generelle Stellungnahme oben. Speziell für S. 21: Die Forderung, die Verwendung des Euro-Scores als Risikofaktor zu begründen, ist insbesondere deshalb nicht zu verstehen, da in der Anlage der BQS genau dieses Vorgehen für die Letalität vorgesehen ist.

Zur Frage bezüglich S. 43: Wir haben in der Diskussion nur die für diesen Bericht relevanten Ergebnisse des BQS-Datenabgleichs dargestellt. Der betreffende Abschnitt wurde – auch motiviert durch die Stellungnahmen der DKG und der BÄK – umformuliert und präzisiert.

Zur Frage bezüglich S. 44: Wir haben keine Entscheidungsfindung ausgeschlossen. Die Darstellung des Designs einer kontrollierten Interventionsstudie war nicht Ziel des Auftrags.