



**Entwicklung und Anwendung von Modellen zur
Berechnung von Schwellenwerten bei
Mindestmengen für die Knie-Totalendoprothese**

- Vorbericht -

[Auftrag B05/01a]

Version 1.0

Stand: 17.10.2005

Thema: Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für die Knie-Totalendoprothese

Auftraggeber: Gemeinsamer Bundesausschuss
Datum des Auftrags: 22.12.2004, konkretisiert am 09.06.2005
Interne Auftragsnummer: B05/01a

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

Bemerkung:

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

Inhaltsverzeichnis

| | Seite |
|--|-----------|
| Inhaltsverzeichnis | 3 |
| Tabellenverzeichnis | 4 |
| Abbildungsverzeichnis..... | 5 |
| Abkürzungsverzeichnis | 6 |
| 1. Ziel der Untersuchung..... | 7 |
| 2. Hintergrund..... | 7 |
| 2.1 Auftrag | 7 |
| 2.2 Mindestmengen für die Knie-Totalendoprothese | 8 |
| 3. Projektablauf einschließlich Konkretisierung | 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..... | 12 |
| 4.3.3 Berechnung von Schwellenwerten..... | 13 |
| 4.3.4 Darstellung der klinischen Relevanz | 13 |
| 4.3.5 Multiples Testen und Signifikanzniveau | 14 |
| 4.3.6 Software | 14 |
| 4.4 Ergänzungen im Vergleich zum Berichtsplan | 14 |
| 5. Ergebnisse..... | 15 |
| 5.1 Deskriptive Statistik..... | 15 |
| 5.2 Ergebnisse für den Indikator "Unbeweglichkeit" | 17 |
| 5.2.1 Zusammenhang zwischen Fallzahl und Risiko (2004)..... | 17 |
| 5.2.2 Zusammenhang zwischen Fallzahl und Risiko (2003)..... | 24 |
| 5.2.3 Untersuchung der Mindestmenge MM=50..... | 27 |
| 5.2.4 Schlussfolgerungen | 27 |
| 5.3 Ergebnisse für den Indikator "Infektion" | 28 |
| 5.3.1 Zusammenhang zwischen Fallzahl und Risiko (2004)..... | 28 |
| 5.3.2 Zusammenhang zwischen Fallzahl und Risiko 2003..... | 32 |
| 5.3.3 Berechnung des Benchmarks VARL..... | 33 |
| 5.3.4 Untersuchung der Mindestmenge MM=50..... | 35 |
| 5.3.5 Schlussfolgerungen | 36 |
| 6. Diskussion | 36 |
| 7. Zusammenfassung | 38 |
| 8. Literatur | 39 |
| Anhang: Weiterer Zeitplan..... | 41 |

Tabellenverzeichnis

| | |
|------------|--|
| Tabelle 1 | Deskriptive Übersicht |
| Tabelle 2 | Rangkorrelation aller Kovariablen |
| Tabelle 3 | Deskriptive Übersicht der Kovariablen bzgl. fehlender Werte bei "Unbeweglichkeit" |
| Tabelle 4 | Anteil fehlender Werte bei "Unbeweglichkeit" in Abhängigkeit von der Fallzahl |
| Tabelle 5 | Logistisches Regressionsmodell mit FP für "Unbeweglichkeit" (Modell I, 2004) |
| Tabelle 6 | Multiples logistisches Regressionsmodell mit FP und Kovariablen für "Unbeweglichkeit" (Modell II, 2004) |
| Tabelle 7 | Ergebnisse des GEE-Ansatzes für "Unbeweglichkeit" (Modell III, 2004) |
| Tabelle 8 | Multiples logistisches Regressionsmodell mit FP und Kovariablen inklusive der Verweildauer für "Unbeweglichkeit" (Modell IV, 2004) |
| Tabelle 9 | Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-IV für "Unbeweglichkeit" (2004) |
| Tabelle 10 | Logistisches Regressionsmodell mit FP für "Unbeweglichkeit" (Modell I, 2003) |
| Tabelle 11 | Multiples logistisches Regressionsmodell mit FP und Kovariablen für "Unbeweglichkeit" (Modell II, 2003) |
| Tabelle 12 | Ergebnisse des GEE-Ansatzes für "Unbeweglichkeit" (Modell III, 2003) |
| Tabelle 13 | Multiples logistisches Regressionsmodell mit FP und Kovariablen inklusive der Verweildauer für "Unbeweglichkeit" (Modell IV, 2003) |
| Tabelle 14 | Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-IV für "Unbeweglichkeit" (2003) |
| Tabelle 15 | Einfaches logistisches Regressionsmodell für "Infektion" (Modell I, 2004) |
| Tabelle 16 | Multiples logistisches Regressionsmodell mit Kovariablen für "Infektion" (Modell II, 2004) |
| Tabelle 17 | Ergebnisse des GEE-Ansatzes für "Infektion" (Modell III, 2004) |
| Tabelle 18 | Einfaches logistisches Regressionsmodell für "Infektion" (Modell I, 2003) |
| Tabelle 19 | Multiples logistisches Regressionsmodell mit Kovariablen für "Infektion" (Modell II, 2003) |
| Tabelle 20 | Ergebnisse des GEE-Ansatzes für "Infektion" (Modell III, 2004) |
| Tabelle 21 | Ergebnisse des VARL-Ansatzes nach Bender [14] für "Infektion" |

Abbildungsverzeichnis

| | |
|--------------|---|
| Abbildung 1 | Risikoplot für "Unbeweglichkeit" vs. Fallzahl (2004) |
| Abbildung 2 | Risikoverlauf für "Unbeweglichkeit" aus Modell I (2004) |
| Abbildung 3 | Adjustierter Risikoverlauf für "Unbeweglichkeit" aus Modell II (2004) |
| Abbildung 4 | Risikoverlauf für "Unbeweglichkeit" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos (2004) |
| Abbildung 5 | Risikoverlauf für "Unbeweglichkeit" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos (2003) |
| Abbildung 6 | Risikoplot für "Infektion" vs. Fallzahl (2004) |
| Abbildung 7 | Risikoverlauf für "Infektion" aus Modell I (2004) |
| Abbildung 8 | Adjustierter Risikoverlauf für "Infektion" aus Modell II (2004) |
| Abbildung 9 | Risikoverlauf für "Infektion" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos bei stark gedehnter Y-Achse (2004) |
| Abbildung 10 | Risikoverlauf für "Infektion" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos bei stark gedehnter Y-Achse (2003) |
| Abbildung 11 | Risikoverlauf für "Infektion" nach Modell I mit Kennzeichnung des berechneten VARL-Wertes mit 95%-KI bei stark gedehnter Y-Achse (2004) |

Abkürzungsverzeichnis

| Abkürzung | Bedeutung |
|------------------|--|
| AG | Arbeitsgruppe |
| ASA | American Association of Anesthesists |
| BQS | Bundesgeschäftsstelle für Qualitätssicherung |
| CDC | Centers for Disease Control |
| FG | Freiheitsgrad |
| FP | Fractional Polynomials |
| G-BA | Gemeinsamer Bundesausschuss |
| GEE | Generalized Estimating Equations |
| HV | High Volume |
| IQWiG | Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen |
| KI | Konfidenzintervall |
| Knie-TEP | Knie-Totalendoprothese |
| LV | Low Volume |
| MM | Mindestmenge |
| OR | Odds Ratio |
| SD | Standard Deviation |
| VARL | Value of an Acceptable Risk Limit |

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 Knie-Totalendoprothese (Knie-TEP) ableitbar sind. Wenn ja, werden für diesen Zweck adäquate Regressionsmodelle entwickelt und diese zur Berechnung von Schwellenwerten bzw. Benchmarks angewendet. Die inhaltlichen und statistischen Eigenschaften der abgeleiteten Werte werden kritisch diskutiert.

2. Hintergrund

2.1 Auftrag

Der Gemeinsame Bundesausschuss (G-BA) hat in seiner Sitzung am 21.12.2004 beschlossen, das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Entwicklung eines Rechenmodells zur Festlegung von Schwellenwerten für Mindestmengen zu beauftragen. Dieses Rechenmodell sollte unter Berücksichtigung deutscher Daten für die Leistungsbereiche 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 09.06.2005 für die Knie-TEP neu formuliert. Der Auftragsgegenstand lautet nun wie folgt: Das IQWiG soll adäquate Rechenmodelle für die Indikation Knie-TEP 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 Knie-Totalendoprothese

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 wurde gemäß § 91 Abs. 7 am 21.09.2004 beschlossen, diesen Mindestmengenkatalog um die Leistungen Knie-TEP und koronarchirurgische Eingriffe zu erweitern. Hierfür wurden jedoch 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 am 09.06.2005 für die Knie-TEP konkretisiert und neu formuliert. Am 16.08.2005 hat das Beschlussgremium des G-BA in seiner für Krankenhausbehandlung zuständigen Besetzung eine verbindliche Mindestmenge von 50 Knie-TEP-Eingriffen pro Krankenhaus und Jahr festgelegt. Diese Regelung tritt ab dem 01.01.2006 in Kraft. Eine Übergangsregelung sieht vor, dass Krankenhäuser, die knapp unter der Menge von 50 liegen (40-49 Eingriffe pro Jahr) und im BQS-Verfahren zur stationären Qualitätssicherung im Jahr 2004 die erforderlichen Kriterien erfüllt haben, diese Leistung auch noch im Jahr 2006 erbringen dürfen. Von diesem Beschluss unberührt bleibt der Auftrag an das IQWiG anhand von deutschen Daten zu prüfen, ob für die Leistung Knie-TEP mit Hilfe eines Berechnungsmodells Schwellenwerte ableitbar sind.

3. Projektablauf einschließlich Konkretisierung

Die Bearbeitung des Auftrags erfolgte nach dem im Berichtsplan dargestellten Schema, das bereits während der Konkretisierung zum Teil durchlaufen war. In Klammern stehen die jeweils verantwortlichen Institutionen. Die Festlegung der Qualitätsindikatoren und Einflussfaktoren sowie die Festlegung der klinisch relevanten Unterschiede wurden von den vom Unterausschuss "*Sonstige stationäre Qualitätssicherung*" benannten Fachexperten für die Knie-TEP 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, wenn verfügbar)
(Unterausschuss "*Sonstige stationäre Qualitätssicherung*")
- 4) Mittleres Risiko der Patienten und Quartile aus 2003er Daten
(IQWiG)
- 5) Klinisch relevante Unterschiede sowie Wichtigkeit der Indikatoren
(ggfs. Rangfolge) festlegen und dokumentieren
(Fachexperten, Konsentierung in AG Mindestmengen)
- 6) Zusammenhang darstellen mit 2003er Daten
(Risikoplot, Regressionsmodell)
(IQWiG)
- 7) Modellwahl
(IQWiG)
- 8) 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) Schwellenwertberechnung für 2004er Daten
(IQWiG)

4. Methoden

Die Methoden zur Erstellung des Berichts wurden im Berichtsplan vom 18.07.2005 vorab festgelegt. Sofern sich im Verlauf diesbezüglich Änderungen ergeben haben, sind diese im Abschnitt 4.4 dargestellt.

4.1 Datenquellen

Es wurden die Daten der BQS aus dem Bundesdatenpool 2003 und 2004 verwendet (siehe Abschnitt 4.3.1). Diese Daten wurden für den vorliegenden Auftrag von der BQS aufbereitet und dem IQWiG am 11.07.2005 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

Die Fachexperten für die Knie-TEP hatten im Rahmen der Konkretisierung am 10.10.2005 folgende Festlegungen bezüglich der Eigenschaften und Effektstärken einer Mindestmengenregelung getroffen. Informationen zu den hier genannten Variablen findet man im Abschnitt 4.3.1.

A) Rangfolge der Qualitätsindikatoren bezüglich ihrer Wichtigkeit

(gemeint ist hierbei die Wichtigkeit bezüglich einer Mindestmengenregelung)

- | | |
|--------------------|------------------|
| 1) Unbeweglichkeit | 2) Wundinfektion |
|--------------------|------------------|

B) Klinisch relevante Veränderung des absoluten Risikos

(gemeint ist hierbei die Veränderung bezüglich des mittleren absoluten Risikos aufgrund einer Intervention)

- | | |
|--------------------------------|-------------------------------|
| 1) Unbeweglichkeit: $\pm 10\%$ | 2) Wundinfektion: $\pm 0.4\%$ |
|--------------------------------|-------------------------------|

C) Akzeptables absolutes Risiko

(gemeint ist hierbei die absolute Risikogrenze, die aufgrund einer Intervention unterschritten werden soll)

- | | |
|-------------------------|----------------------|
| 1) Unbeweglichkeit: 10% | 2) Wundinfektion: 1% |
|-------------------------|----------------------|

4.3 Auswertungen

4.3.1 Datenbeschreibung

Die primäre Datenbasis besteht aus Daten von 110349 Patienten mit Knie-Totalendoprothesen-Erstimplantation aus 1016 Krankenhäusern aus dem Bundesdatenpool 2004 der BQS [3]. Für Untersuchungen zur Modellstabilität wurden auch die Daten von 90004 Patienten mit Knie-Totalendoprothesen-Erstimplantation aus 965 Krankenhäusern aus dem Bundesdatenpool 2003 [4] verwendet. Die beiden gewählten Qualitätsindikatoren sind wie folgt definiert.

- 1) Unbeweglichkeit: keine postoperative Beweglichkeit (Extension/Flexion von mindestens 0/0/90 bei Entlassung)
- 2) Infektion: postoperative Wundinfektion nach den Kriterien der Centers for Disease Control (CDC)

Der erste Qualitätsindikator wurde in der Form "Unbeweglichkeit ja/nein" kodiert, damit für beide Qualitätsindikatoren das Risiko für ein unerwünschtes Ereignis in Abhängigkeit der Fallzahl untersucht wird und so die Richtung einer Qualitätsverschiebung gleich 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.

Um eine Risikoadjustierung durchzuführen, wurden die folgenden potentiellen Risikofaktoren zur Verfügung gestellt:

- (a) Alter: in Jahren
- (b) Geschlecht: männlich ja/nein
- (c) ASA-Status: 0 = ASA-Score 1 bis 2 , 1 = ASA-Score 3 bis 5
- (d) Deformität: 0 = keine , 1 = X-Stellung , 2 = O-Stellung
- (e) Arthrose-Score: Score-Werte von 1 bis 8
(der BQS-Arthrose-Score berechnet sich hierbei als Summe der Punktwerte der Variablen Osteophyten, Gelenkspalt, Sklerose und Deformierung)
- (f) Verweildauer: postoperative Verweildauer in Tagen

Für deskriptive Datenauswertungen wurden Standardmethoden der deskriptiven Statistik eingesetzt (absolute und relative Häufigkeiten, Mittelwert, Standardabweichung, Korrelationen).

4.3.2 Modellbildung

Die Modellbildung wurde für die beiden gewählten Qualitätsindikatoren "Unbeweglichkeit" und "Infektion", jeweils getrennt in mehreren aufeinander aufbauenden Schritten, für die Daten des Bundesdatenpools 2004 vorgenommen. Zur visuellen Darstellung des Zusammenhangs zwischen Fallzahl und Risiko wurden zunächst deskriptive Risikoplots erstellt [5]. 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. Diese Risikoplots geben einen ersten Eindruck, in welcher Form die betrachtete erklärende Variable sinnvoll in einem logistischen Regressionsmodell untersucht werden kann.

Im Rahmen eines einfachen logistischen Regressionsmodells [6] wurde dann der rohe Zusammenhang zwischen dem jeweiligen Qualitätsindikator und der Fallzahl analysiert. Dabei wurde die Methode der Fractional Polynomials (FP) eingesetzt, die eine flexible Modellierung auch nicht monotoner Zusammenhänge ermöglicht [7]. Anschließend wurden multiple logistische Regressionsanalysen durchgeführt. Neben der Fallzahl wurden für den Indikator "Unbeweglichkeit" die Risikofaktoren (a) bis (e) und für den Indikator "Infektion" die Risikofaktoren (a) bis (c) betrachtet sowie jeweils alle Wechselwirkungen 1. Ordnung untersucht. Danach wurde mit Hilfe der Methode der Generalized Estimating Equations (GEE) [8] 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 [9,10]. Aufgrund der unerwarteten nicht-monotonen Beziehung zwischen dem Indikator "Unbeweglichkeit" und der Fallzahl wurde bei der Analyse dieses Indikators zudem noch die Verweildauer (f) als zusätzlicher potentieller Risikofaktor betrachtet, um ein mögliches Confounding durch diese Variable auszuschließen.

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 zwei Maße für erklärte Variabilität berechnet: Das adjustierte Likelihood-Ratio R^2 nach Nagelkerke [11] sowie die von Mittelböck & Schemper [12] empfohlene Squared-Pearson-Correlation r^2 . Die Robustheit der Modelle wurde mit regressionsdiagnostischen Verfahren überprüft [13].

Zur Untersuchung der Ergebnisstabilität wurden die gleichen Modelle auch auf die Daten des Bundesdatenpools 2003 angewendet. Da bei der primären Zielvariable "Unbeweglichkeit" bei der primären Analyse der Daten von 2004 sehr viele fehlende Werte auftraten, wurden hier weitere Sensitivitätsuntersuchungen vorgenommen. Dies geschah in der Weise, dass fehlende Werte komplett auf Ereignis bzw. Nicht-Ereignis bezüglich der "Unbeweglichkeit" gesetzt wurden.

4.3.3 Berechnung von Schwellenwerten

Bezüglich des Qualitätsindikators "Unbeweglichkeit" kam keine Schwellenwertberechnung in Frage. Bezüglich des Qualitätsindikators "Infektion" wurde der VARL-Ansatz nach Bender [14] zur Ableitung von Schwellenwerten verwendet. Der entsprechende Benchmark VARL wurde, wie bei Bender [14] beschrieben, aus dem logistischen Regressionsmodell geschätzt und die Schätzunsicherheit mit Hilfe von 95%-Konfidenzintervallen dokumentiert.

4.3.4 Darstellung der klinischen Relevanz

Es wurde untersucht, ob die gefundenen Benchmark-Werte mit einer klinisch relevanten Risikoreduktion bezüglich einer Mindestmengen-Regelung assoziiert sind. Als geeignetes Maß ergibt sich die Differenz (δ) des mittleren Risikos über alle Patienten (π) und dem mittleren Risiko der Patienten aus High-Volume-Krankenhäusern (π_{HV}), d.h. $\delta = \pi - \pi_{HV}$, wobei der Trennwert für die Unterscheidung zwischen High-Volume und Low-Volume-Krankenhäusern gerade durch den Benchmark VARL gegeben ist.

Sei π_{LV} das mittlere Risiko der Patienten aus Low-Volume-Krankenhäusern und ω_{LV} der relative Anteil der Patienten aus Low-Volume-Krankenhäusern, so gilt die Beziehung $\delta = \omega_{LV} \times (\pi_{LV} - \pi_{HV})$. Die Größe $\pi_{LV} - \pi_{HV}$ stellt hierbei die

gewöhnliche Differenz zweier unabhängiger Risiken dar, für die es Standardmethoden zur Berechnung von Konfidenzintervallen gibt. Unter Annahme fester Randhäufigkeiten der entsprechenden Vierfeldertafel lassen sich durch Anwendung obiger Beziehung auch Konfidenzintervalle für δ berechnen. Zur Intervallschätzung für die Größe $\pi_{LV} - \pi_{HV}$ wurde hierbei die Wilson-Score-Methode verwendet [15,16].

Mit den gleichen Methoden wurde auch die klinische Relevanz der am 16.08.2005 festgelegten Mindestmenge von $MM = 50$ untersucht.

4.3.5 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 konfirmatorische Prüfung handelt. Alle hier durchgeführten Signifikanztests haben nur hypothesengenerierenden Charakter.

4.3.6 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 ergeben. Zum einen erschien es sinnvoll, aufgrund des unerwarteten Risikoverlaufs in Abhängigkeit von der Fallzahl für den Indikator "Unbeweglichkeit" auch die Verweildauer als potentiellen Confounder im multiplen Regressionsmodell zu berücksichtigen. Zum anderen wurde untersucht, ob die am 16.08.2005 beschlossene Mindestmenge von 50 Eingriffen pro Krankenhaus pro Jahr mit einer klinisch relevanten Risikoveränderung bezüglich der hier untersuchten Qualitätsindikatoren assoziiert ist.

5. Ergebnisse

5.1 Deskriptive Statistik

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 | - | 90004 | - | 110349 |
| Anzahl Krankenhäuser | - | 965 | - | 1016 |
| Alter (Jahre) | 24 | 70.2 (SD 8.4) | 20 | 70.2 (SD 8.4) |
| Geschlecht (männlich / weiblich) | - | 29.3 / 70.6 | - | 29.5 / 70.5 |
| ASA-Status (<=2 / >2) | - | 66.3 / 33.7 | - | 66.1 / 33.9 |
| Deformität (keine / X-Stellung / O-Stellung) | - | 28.7 / 20.8 / 50.5 | - | 29.8 / 20.5 / 49.7 |
| Arthrose-Score (0-3 / 4-8) | 19154 | 21.0 / 79.0 | - | 17.8 / 82.2 |
| Postoperative Verweildauer (Tage) | 212 | 16.8 (SD 6.9) | 410 | 15.0 (SD 5.7) |
| Fallzahl | - | 208.1 (SD 155.7) | - | 233.4 (SD 170.4) |
| Unbeweglichkeit (ja / nein) | 32177 | 27.22 / 72.78 | 18635 | 26.86 / 73.14 |
| Infektion (ja / nein) | - | 0.92 / 99.08 | - | 0.79 / 99.21 |
| Mittelwert (SD=Standardabweichung) bzw. Häufigkeiten (in Prozent) | | | | |

Tab. 2: Rangkorrelationen aller Kovariablen

| Jahr 2003 | | | | | | | |
|--|----------|-------|------------|------------|-----------------------------|-------------------------|--------------|
| | Fallzahl | Alter | Geschlecht | ASA-Status | Arthrose-Score ^a | Deformität ^b | Verweildauer |
| Fallzahl | 1.00 | -0.04 | 0.000 | -0.03 | 0.02 | -0.001 | -0.10 |
| Alter | | 1.00 | 0.11 | 0.17 | 0.06 | 0.05 | 0.09 |
| Geschlecht | | | 1.00 | 0.001 | -0.02 | 0.07 | 0.05 |
| ASA-Status | | | | 1.00 | 0.10 | 0.04 | 0.06 |
| Arthrose-Score ^a | | | | | 1.00 | 0.18 | 0.04 |
| Deformität ^b | | | | | | 1.00 | 0.03 |
| Jahr 2004 | | | | | | | |
| Fallzahl | 1.00 | -0.05 | 0.004 | -0.008 | 0.06 | 0.03 | -0.08 |
| Alter | | 1.00 | 0.12 | 0.17 | 0.05 | 0.04 | 0.08 |
| Geschlecht | | | 1.00 | 0.004 | -0.01 | 0.07 | 0.06 |
| ASA-Status | | | | 1.00 | 0.10 | 0.03 | 0.05 |
| Arthrose-Score ^a | | | | | 1.00 | 0.19 | 0.04 |
| Deformität ^b | | | | | | 1.00 | 0.03 |
| ^a Arthrose-Score: Werte von 1-8, ^b Deformität: 1=keine, 2=O-Stellung, 3=X-Stellung | | | | | | | |

Es zeigen sich keine gravierenden Unterschiede in der Verteilung der Variablen in den Jahren 2003 und 2004. Alle Variablen, insbesondere auch Fallzahl und die übrigen Kovariablen, sind in beiden Jahren nur schwach miteinander korreliert.

Wegen der erheblichen Anzahl fehlender Werte bei der primären Zielvariablen "Unbeweglichkeit" wurde die Verteilung der Kovariablen zwischen den Patienten mit und ohne Berücksichtigung der fehlenden Werte betrachtet (Tabelle 3). Es zeigten sich keine gravierenden Unterschiede zwischen den Messungen der Patienten mit fehlenden und ohne fehlende Angaben zur "Unbeweglichkeit". Lediglich bei der Fallzahl zeigte sich ein auffällig niedrigerer Mittelwert in der Gruppe mit fehlenden Werten zur "Unbeweglichkeit".

Tab. 3: Deskriptive Übersicht der Kovariablen bzgl. fehlender Werte bei "Unbeweglichkeit"

| | Jahr 2003 | | Jahr 2004 | |
|--|--------------------|--------------------|--------------------|--------------------|
| | fehlende Werte | vorhandene Werte | fehlende Werte | vorhandene Werte |
| Anzahl | 32177 | 57827 | 18635 | 91714 |
| Alter (Jahre) | 70.3 (SD 8.6) | 70.1 (SD 8.3) | 70.5 (SD 8.7) | 70.1 (SD 8.4) |
| Geschlecht (männlich / weiblich) | 29.5 / 70.5 | 29.3 / 70.7 | 29.3 / 70.7 | 29.5 / 70.5 |
| ASA-Status (<=2 / >2) | 66.8 / 33.2 | 66.0 / 34.0 | 69.0 / 31.0 | 65.6 / 34.4 |
| Deformität (keine / X-Stellung / O-Stellung) | 33.2 / 20.9 / 45.9 | 26.2 / 20.7 / 53.1 | 35.9 / 22.0 / 42.1 | 28.6 / 20.2 / 51.2 |
| Arthrose-Score (0-3 / 4-8 / k.A.) | 11.2 / 29.3 / 59.5 | 19.5 / 80.5 / 0.0 | 24.1 / 75.9 / 0.0 | 16.5 / 83.5 / 0.0 |
| Postoperative Verweildauer (Tage) | 16.9 (SD 6.9) | 16.7 (SD 6.8) | 15.1 (SD 6.7) | 14.9 (SD 5.5) |
| Fallzahl | 188.8 (SD 152.1) | 218.8 (SD 156.7) | 166.3 (SD 124.3) | 247.1 (SD 175.2) |
| Infektion (ja / nein) | 99.12 / 0.88 | 99.06 / 0.94 | 98.70 / 1.30 | 99.32 / 0.68 |
| k.A.=keine Angabe, Mittelwert (SD=Standardabweichung) bzw. Häufigkeiten (in Prozent) | | | | |

Zur näheren Untersuchung des Zusammenhangs zwischen den fehlenden Werten und der Fallzahl pro Krankenhaus und Jahr wurde der Datensatz basierend auf den empirisch ermittelten Quantilen in 5 etwa gleich große Gruppen unterteilt und der Anteil fehlender Werte pro Gruppe berechnet (Tabelle 4). Es zeigt sich, dass der Anteil fehlender Werte mit steigender Fallzahl tendenziell abnimmt, d.h. bei Low-Volume-Krankenhäusern treten mehr fehlende Werte auf als bei High-Volume-Krankenhäusern. Daher ist ein Bias durch fehlende Werte möglich.

Tab. 4: Anteil fehlender Werte bei "Unbeweglichkeit" in Abhängigkeit von der Fallzahl

| Jahr 2003 | | | Jahr 2004 | | |
|-----------|-------|-------------------------------|-----------|-------|-------------------------------|
| Fallzahl | N | Anteil fehlender Werte (in %) | Fallzahl | N | Anteil fehlender Werte (in %) |
| 1-77 | 18043 | 44.4 | 1-88 | 22121 | 24.7 |
| 78-134 | 17968 | 39.9 | 89-150 | 22413 | 23.8 |
| 135-220 | 18322 | 32.9 | 151-239 | 22010 | 17.9 |
| 221-313 | 17702 | 29.4 | 240-358 | 22023 | 8.7 |
| 314-941 | 17969 | 32.1 | 359-874 | 21782 | 9.1 |

Fallzahl-Intervalle sind anhand von Quantilen gebildet

5.2 Ergebnisse für den Indikator "Unbeweglichkeit"

5.2.1 Zusammenhang zwischen Fallzahl und Risiko (2004)

Es wurde zunächst eine Complete-Case-Analyse durchgeführt. Von 91714 Patienten mit vorhandenen Angaben weisen 24631 (26.9%) eine Unbeweglichkeit auf. Abbildung 1 zeigt den Risikoplot in Abhängigkeit von der Fallzahl. Der Verlauf ist nicht monoton und weist einen klar U-förmigen Zusammenhang zwischen Fallzahl und Risiko auf.

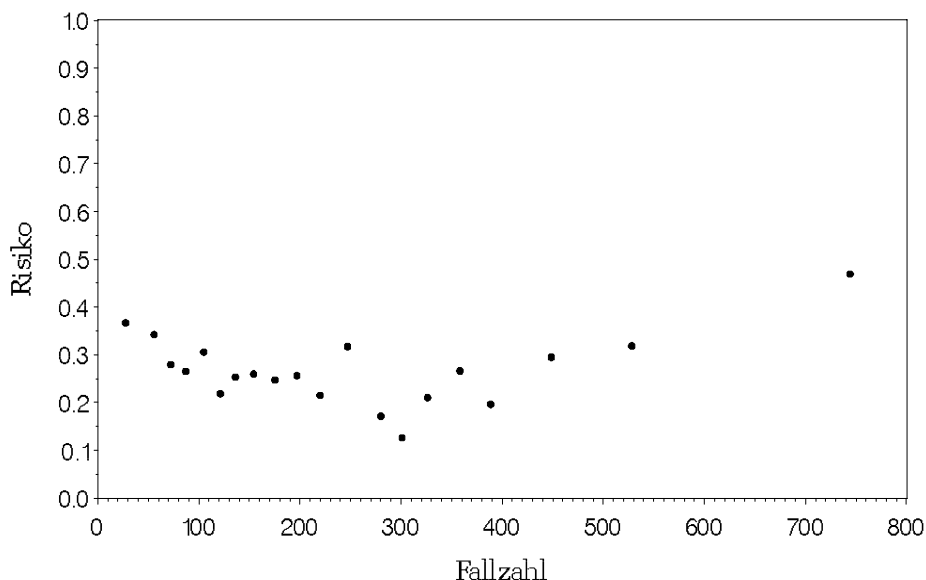


Abb. 1: Risikoplot für "Unbeweglichkeit" vs. Fallzahl (2004)

Da sich der nicht-monotone Risikoverlauf im Rahmen eines logistischen Regressionsmodells nicht mittels eines einzelnen, linearen Terms adäquat modellieren lässt, wurde die FP-Methode verwendet. Gemäß dem Ansatz von Royston et al. [7] wurde aus einer definierten Anzahl von Polynomen 1. und 2. Grades das Modell mit der besten Anpassung ermittelt. Die besten Ergebnisse lieferten die Polynompaare mit (i) linearer + kubischer Term bzw. (ii) Wurzel + kubischer Term. Für die weitere Modellierung wurden linearer und kubischer Term verwendet (Modell I, Tabelle 5). Abbildung 2 zeigt den aus der logistischen Regression geschätzten Risikoverlauf bei Verwendung dieser FPs.

Tab. 5: Logistisches Regressionsmodell mit FP für "Unbeweglichkeit" (Modell I, 2004)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|--|----|--------------------|----------------|----------|--------------|--------|
| Intercept | 1 | -0.6315 | 0.0162 | < 0.0001 | - | - |
| Fallzahl (linear) | 1 | -0.0028 | 0.0001 | < 0.0001 | siehe Tab. 7 | |
| Fallzahl (kubisch) | 1 | 7.074E-9 | 1.45E-10 | < 0.0001 | | |
| Hosmer-Lemeshow-Anpassungstest: 198.7 ($p < 0.0001$) | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.037$; Adjusted-Likelihood-Ratio $R^2 = 0.046$ | | | | | | |

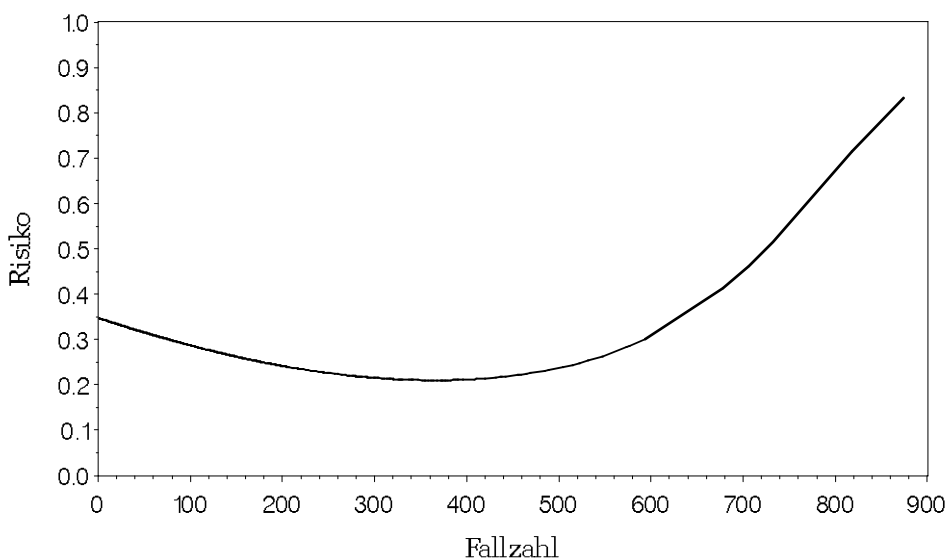


Abb. 2: Risikoverlauf für "Unbeweglichkeit" aus Modell I (2004)

Die Kurve zeigt einen klar U-förmigen Verlauf, d.h. nach zunächst fallendem Risiko mit wachsender Fallzahl steigt das Risiko mit weiter wachsender Fallzahl wieder an.

Um zu untersuchen, ob dieser nicht-monotone Risikoverlauf durch Confounding verursacht wird, wurde in einem nächsten Schritt eine Adjustierung bezüglich der vorgegebenen potentiellen Risikofaktoren mit Hilfe eines multiplen logistischen Regressionsmodells vorgenommen. Dabei wurden zunächst nur Haupteffekte betrachtet. Im Rahmen des multiplen Modells wurden zuerst die FPs mit der besten Anpassung ermittelt, wobei wiederum linearer plus kubischer Term die beste Anpassungsgüte für die Fallzahl lieferten. Die Ergebnisse des multiplen Modells sind in Tabelle 6 zusammengefasst (Modell II).

Tab. 6: Multiples logistisches Regressionsmodell mit FP und Kovariablen für "Unbeweglichkeit" (Modell II, 2004)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|--|----|--------------------|----------------|----------|--------------------|---------------|
| Intercept | 1 | -0.1696 | 0.0711 | 0.0170 | - | - |
| Fallzahl (linear) | 1 | -0.0029 | 0.0001 | < 0.0001 | siehe Tab. 7 | |
| Fallzahl (kubisch) | 1 | 7.111E-9 | 1.45E-10 | < 0.0001 | | |
| Alter | 1 | -0.0052 | 0.0009 | < 0.0001 | 0.975 ^a | 0.966 – 0.985 |
| Geschlecht (männlich) | 1 | -0.0108 | 0.0170 | 0.5232 | 0.989 | 0.957 – 1.023 |
| ASA-Status (1-2) | 1 | -0.1650 | 0.0161 | < 0.0001 | 0.848 | 0.822 – 0.875 |
| Arthrose-Score (0-3) | 1 | 0.0114 | 0.0208 | 0.5831 | 1.011 | 0.971 – 1.054 |
| Deformität (O-Stellung) | 1 | 0.0281 | 0.0181 | 0.1209 | 1.028 | 0.993 – 1.066 |
| Deformität (X-Stellung) | 1 | -0.0132 | 0.0225 | 0.5581 | 0.987 | 0.944 – 1.031 |
| Hosmer-Lemeshow-Anpassungstest: 97.4 ($p < 0.0001$) | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.048$; Adjusted-Likelihood-Ratio $R^2 = 0.038$ | | | | | | |
| ^a bezogen auf 5 Jahre | | | | | | |

Die Modellanpassungsgüte, gemessen mit dem Hosmer-Lemeshow-Goodness-of-Fit Test, konnte deutlich verbessert werden. Dennoch zeigt der Test immer noch ein signifikantes Ergebnis. Dies kann mehrere Ursachen haben. Es kann ein Hinweis darauf sein, dass es weitere wichtige Risikofaktoren oder Wechselwirkungen gibt, die noch nicht im Modell enthalten sind. Es können einzelne Ausreißer vorhanden sein, die die globale Modellgüte stören. Beides wurde daher in weiteren Modellbildungsschritten genauer untersucht (siehe unten). Zu bedenken ist aber auch der sehr hohe Stichprobenumfang, der auch bei kleinen unbedeuteten Modellabweichungen zu signifikanten Anpassungstests führen kann. Der Anteil der erklärten Variabilität ist im Vergleich zum Modell ohne Kovariablen leicht höher. Die Parameterschätzungen für die Fallzahl bleiben

jedoch nahezu unverändert, so dass kein Confounding durch die betrachteten Einflussfaktoren vorliegt.

Abbildung 3 zeigt den aus Modell II berechneten adjustierten Risikoverlauf, wobei für die Kovariablen jeweils der Mittelwert eingesetzt wurde. Der Risikoverlauf ist der nicht adjustierten Analyse sehr ähnlich.

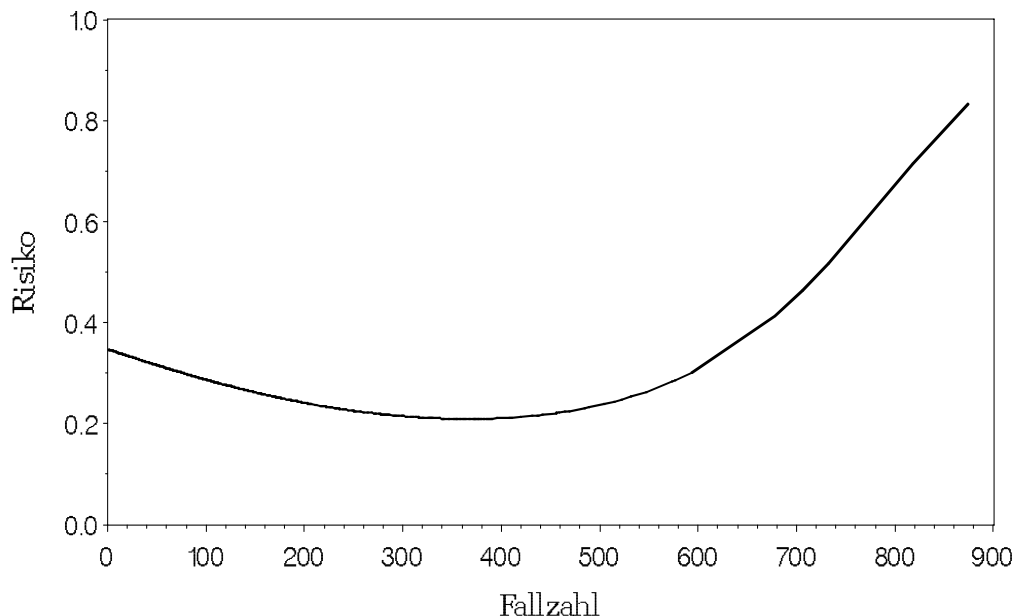


Abb. 3: Adjustierter Risikoverlauf für "Unbeweglichkeit" aus Modell II (2004)

Zur Überprüfung auf signifikante Wechselwirkungen wurde in einem nächsten Schritt Modell II um Wechselwirkungen 1. Ordnung erweitert und dann ein Backward-Eliminationsverfahren mit Einschlussgrenze $p=0.01$ durchlaufen. Keine der betrachteten Wechselwirkungen war signifikant.

Um mögliche Clustereffekte durch Abhängigkeiten von Werten innerhalb eines Krankenhauses zu überprüfen, wurde schließlich die GEE-Methode basierend auf dem logistischen Regressionsmodell mit Haupteffekten eingesetzt. Tabelle 7 zeigt die Ergebnisse des GEE-Ansatzes (Modell III). Die Parameterschätzer der Fallzahl verändern sich nur minimal, d.h. die Abhängigkeit des Risikoverlaufs von der Fallzahl ändert sich auch bei Berücksichtigung eines möglichen Clustereffektes nicht.

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

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|--|----|--------------------|----------------|----------|--------------------|---------------|
| Intercept | 1 | 0.0978 | 0.0889 | 0.2711 | - | - |
| Fallzahl (linear) | 1 | -0.0034 | 0.0005 | < 0.0001 | siehe Tab. 7 | |
| Fallzahl (kubisch) | 1 | <0.0000 | <0.0000 | < 0.0001 | | |
| Alter | 1 | -0.0071 | 0.0009 | < 0.0001 | 0.965 ^a | 0.957 – 0.974 |
| Geschlecht (männlich) | 1 | -0.0181 | 0.0186 | 0.3307 | 0.982 | 0.947 – 1.019 |
| ASA-Status (1-2) | 1 | -0.1206 | 0.0177 | < 0.0001 | 0.886 | 0.856 – 0.918 |
| Arthrose-Score (0-3) | 1 | -0.0179 | 0.0286 | 0.5317 | 0.982 | 0.929 – 1.039 |
| Deformität (O-Stellung) | 1 | -0.0166 | 0.0236 | 0.4800 | 0.984 | 0.939 – 1.030 |
| Deformität (X-Stellung) | 1 | -0.0260 | 0.0245 | 0.2878 | 0.974 | 0.929 – 1.022 |
| Exchangeable Kovarianzstruktur: $r = 0.25$ | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.059$ | | | | | | |
| ^a bezogen auf 5 Jahre | | | | | | |

In einem letzten Schritt wurde schließlich noch ein möglicher Einfluss der postoperativen Verweildauer überprüft. Hierzu wurde das Modell II um die Verweildauer erweitert (Modell IV). Die Ergebnisse dieses Modells sind in Tabelle 8 dargestellt. Es zeigen sich nur geringe Unterschiede in den Parameterschätzungen im Vergleich zum Modell ohne postoperative Verweildauer, insbesondere die Parameterschätzungen für die Fallzahl sind nahezu identisch.

Tab. 8: Multiples logistisches Regressionsmodell mit FP und Kovariablen inklusive der Verweildauer für "Unbeweglichkeit" (Modell IV, 2004)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|--|----|--------------------|----------------|----------|--------------------|---------------|
| Intercept | 1 | 0.0925 | 0.0750 | 0.2173 | - | - |
| Fallzahl (linear) | 1 | -0.0029 | 0.0001 | < 0.0001 | siehe Tab. 7 | |
| Fallzahl (kubisch) | 1 | 7.249E-9 | 1.48E-10 | < 0.0001 | | |
| Alter | 1 | -0.0046 | 0.0009 | < 0.0001 | 0.975 ^a | 0.970 – 0.985 |
| Geschlecht (männlich) | 1 | -0.0149 | 0.0170 | 0.3815 | 0.985 | 0.953 – 1.019 |
| ASA-Status (1-2) | 1 | -0.1757 | 0.0161 | < 0.0001 | 0.839 | 0.813 – 0.866 |
| Arthrose-Score (0-3) | 1 | 0.0044 | 0.0209 | 0.8336 | 1.004 | 0.964 – 1.046 |
| Deformität (O-Stellung) | 1 | 0.0250 | 0.0181 | 0.1688 | 1.025 | 0.989 – 1.062 |
| Deformität (X-Stellung) | 1 | -0.0085 | 0.0225 | 0.7057 | 0.992 | 0.949 – 1.036 |
| Postop. Verweildauer | 1 | -0.0187 | 0.0017 | < 0.0001 | 0.913 ^b | 0.895 – 0.927 |
| Hosmer-Lemeshow-Anpassungstest: 48.0 ($p < 0.0001$) | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.040$; Adjusted-Likelihood-Ratio $R^2 = 0.051$ | | | | | | |
| ^a bezogen auf 5 Jahre, ^b bezogen auf 5 Tage | | | | | | |

Die Berechnung eines einzelnen ORs für die Fallzahl ist bei dem vorliegenden nicht-monotonen Risikoverlauf nicht möglich. Zur Berechnung von ORs für die Fallzahl wurden die Daten basierend auf den Quartilen in Intervalle unterteilt und die mittleren beiden Quartile zusammengefasst. Als Vergleichswerte für die Berechnung von ORs wurden die Mittelwerte der drei Intervalle definiert, wobei das mittlere Intervall als Referenzgruppe gewählt wurde. Die resultierenden ORs für die Fallzahl, basierend auf den verschiedenen Modellen, sind in Tabelle 9 aufgelistet. Die geschätzten ORs verändern sich von Modell zu Modell nur unwesentlich.

Tab. 9: Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-IV für "Unbeweglichkeit" (2004)

| Intervall | Referenzwert | Modell | | | |
|-----------|--------------|---------------------|---------------------|---------------------|---------------------|
| | | I | II | III | IV |
| 1 – 112 | 56 | 1.499 (1.463–1.536) | 1.502 (1.466-1.539) | 1.520 (1.483-1.558) | 1.634 (1.422–1.879) |
| 113 – 341 | 227 | 1.000 | 1.000 | 1.000 | 1.000 |
| 342 – 874 | 608 | 1.528 (1.478-1.579) | 1.531 (1.481-1.582) | 1.530 (1.480-1.582) | 1.533 (1.098-2.141) |

Abschließend wurde die Robustheit der Ergebnisse und insbesondere der nicht-monotone Zusammenhang zwischen Risiko und Fallzahl im Rahmen einer Sensitivitätsanalyse überprüft. Hierzu wurden zunächst regressionsdiagnostische Verfahren zur Identifikation extremer Werte eingesetzt, die stark vom angepassten Modell abweichen und eine neue Modellanpassung ohne die identifizierten Extremwerte durchgeführt. Ferner wurden getrennte Analysen für Teilmengen der Daten (d.h. Weglassen von Kliniken mit sehr niedrigen bzw. sehr hohen Fallzahlen) durchgeführt. Die nicht-monotone Form des Risikoverlaufs blieb in allen Analysen erhalten, so dass wir trotz des signifikanten Hosmer-Lemeshow-Tests von diesem Risikoverlauf ausgehen.

Um eine mögliche Verzerrung des Zusammenhangs zwischen Risiko und der Fallzahl aufgrund des hohen Anteils fehlender Werte zu untersuchen, wurde eine "Best-Case- / Worst-Case-Analyse" durchgeführt, d.h. die fehlenden Werte der postoperativen "Unbeweglichkeit" wurden komplett durch Ereignis bzw. Nicht-Ereignis ersetzt. Anschließend wurde jeweils eine erneute Modellanpassung für

den dann vollständigen Datensatz vorgenommen. In beiden extremen Fällen blieb der U-förmige Risikoverlauf in Abhängigkeit von der Fallzahl bestehen, so dass wir die Ergebnisse der vorher dargestellten Complete-Case-Analyse für die weitere Bearbeitung nutzen.

Die Auswertungen zeigen, dass sich der Zusammenhang zwischen Fallzahl und Risiko einer "Unbeweglichkeit" mit den vorhandenen Daten mit Hilfe des Modells I beschreiben lässt. Abbildung 4 zeigt den Risikoverlauf in Abhängigkeit von der Fallzahl basierend auf Modell I. Eingezeichnet sind mittleres Risiko (26.9%) sowie der von den Fachexperten festgelegte Irrelevanzbereich (16.9% – 26.9%) und das akzeptable absolute Risiko (10%).

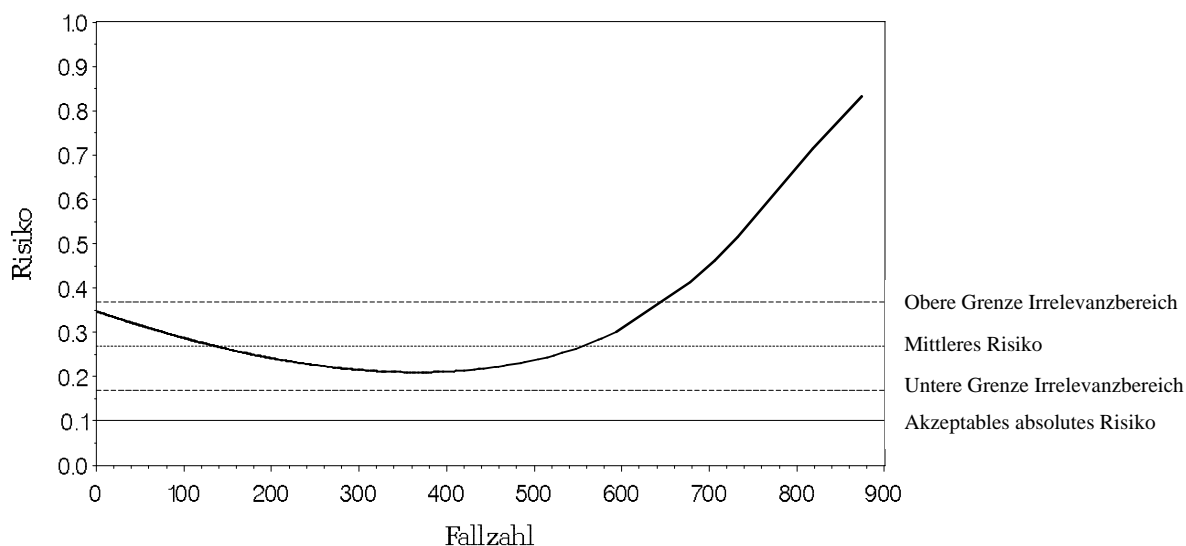


Abb. 4: Risikoverlauf für "Unbeweglichkeit" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos (2004)

Die Kurve liegt für Fallzahlen unterhalb von etwa 650 innerhalb des Irrelevanzbereichs und für höhere Fallzahlen oberhalb der oberen Irrelevanz-Grenze. Die gesamte Kurve verläuft deutlich oberhalb des von den Fachexperten angegebenen akzeptablen absoluten Risikos.

Durch die U-förmige Beziehung zwischen dem Risiko für "Unbeweglichkeit" und der Fallzahl ist es nicht sinnvoll, für die primäre Zielvariable mit Hilfe der im Berichtsplan vorgesehenen Methoden Schwellenwerte oder Benchmarks abzuleiten, die als Mindestmengen diskutiert werden können.

5.2.2 Zusammenhang zwischen Fallzahl und Risiko (2003)

Zur Untersuchung der Ergebnisstabilität wurden die Daten aus dem Jahr 2003 in analoger Form zu den Daten von 2004 ausgewertet. Die Ergebnisse in den beiden Jahren sind sehr ähnlich. Das mittlere Risiko der "Unbeweglichkeit" beträgt 27.2% (15738 von 57827 Patienten mit vorhandenen Angaben). Auch bei den Daten aus dem Jahr 2003 zeigte sich ein nicht-monotoner, U-förmiger Risikoverlauf in Abhängigkeit von der Fallzahl. Den besten Fit bei der Anpassung mit der FP-Methode lieferten linearer plus quadratischer Term (Daten aus 2004: linearer plus kubischer Term). Bei der Untersuchung von Wechselwirkungen war die Interaktion zwischen Arthrose-Score und Deformität statistisch signifikant zum Niveau 1%. Im Vergleich zum Modell mit Kovariablen (Modell II) gab es jedoch keine Veränderungen der Parameterschätzer für die Fallzahl. Auch die Anpassungsgüte (Hosmer-Lemeshow-Test) und die erklärte Variabilität (r^2 und R^2) verbesserten sich nicht. Die Ergebnisse der Modellanpassung der einzelnen Modellvarianten sind in den folgenden Tabellen 10 bis 14 zusammengefasst.

Tab. 10: Logistisches Regressionsmodell mit FP für "Unbeweglichkeit" (Modell I, 2003)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|--|----|--------------------|----------------|----------|---------------|--------|
| Intercept | 1 | -0.5200 | 0.0222 | < 0.0001 | - | - |
| Fallzahl (linear) | 1 | -0.0047 | 0.0002 | < 0.0001 | siehe Tab. 12 | |
| Fallzahl (quadratisch) | 1 | 7.763E-6 | 2.399E-7 | < 0.0001 | | |
| Hosmer-Lemeshow-Anpassungstest: 449.4 ($p < 0.0001$) | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.027$; Adjusted-Likelihood-Ratio $R^2 = 0.035$ | | | | | | |

Tab. 11: Multiples logistisches Regressionsmodell mit FP und Kovariablen für "Unbeweglichkeit" (Modell II, 2003)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|--|----|--------------------|----------------|----------|--------------------|---------------|
| Intercept | 1 | -0.0889 | 0.0901 | 0.3239 | - | - |
| Fallzahl (linear) | 1 | -0.0048 | 0.0002 | < 0.0001 | siehe Tab. 12 | |
| Fallzahl (quadratisch) | 1 | 7.792E-6 | 2.404E-7 | < 0.0001 | | |
| Alter | 1 | -0.0059 | 0.0012 | < 0.0001 | 0.970 ^a | 0.961 – 0.980 |
| Geschlecht (männlich) | 1 | 0.0095 | 0.0212 | 0.6544 | 1.010 | 0.969 – 1.052 |
| ASA-Status (1-2) | 1 | -0.1666 | 0.0201 | < 0.0001 | 0.847 | 0.814 – 0.881 |
| Arthrose-Score (0-3) | 1 | 0.0823 | 0.0244 | 0.0007 | 1.086 | 1.035 – 1.139 |
| Deformität (O-Stellung) | 1 | 0.1192 | 0.0235 | < 0.0001 | 1.127 | 1.074 – 1.180 |
| Deformität (X-Stellung) | 1 | 0.0820 | 0.0285 | 0.0041 | 1.085 | 1.026 – 1.148 |
| Hosmer-Lemeshow-Anpassungstest: 55.1 ($p < 0.0001$) | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.038$; Adjusted-Likelihood-Ratio $R^2 = 0.029$ | | | | | | |
| ^a bezogen auf 5 Jahre | | | | | | |

Tab. 12: Ergebnisse des GEE-Ansatzes für "Unbeweglichkeit" (Modell III, 2003)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|--|----|--------------------|----------------|----------|--------------------|---------------|
| Intercept | 1 | 0.1108 | 0.1062 | 0.2967 | - | - |
| Fallzahl (linear) | 1 | -0.0053 | 0.0008 | < 0.0001 | siehe Tab. 12 | |
| Fallzahl (quadratisch) | 1 | <0.0000 | <0.0000 | < 0.0001 | | |
| Alter | 1 | -0.0073 | 0.0011 | < 0.0001 | 0.964 ^a | 0.954 – 0.975 |
| Geschlecht (männlich) | 1 | 0.0213 | 0.0236 | 0.3679 | 1.022 | 0.975 – 1.070 |
| ASA-Status (1-2) | 1 | -0.0860 | 0.0214 | < 0.0001 | 0.918 | 0.880 – 0.957 |
| Arthrose-Score (0-3) | 1 | -0.0146 | 0.0288 | 0.6131 | 0.986 | 0.931 – 1.043 |
| Deformität (O-Stellung) | 1 | 0.0309 | 0.0266 | 0.2455 | 1.031 | 0.979 – 1.087 |
| Deformität (X-Stellung) | 1 | 0.0495 | 0.0305 | 0.1045 | 1.051 | 0.990 – 1.116 |
| Exchangeable Kovarianzstruktur: $r = 0.25$ Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.038$ | | | | | | |
| ^a bezogen auf 5 Jahre | | | | | | |

Tab. 13: Multiples logistisches Regressionsmodell mit FP und Kovariablen inklusive der Verweildauer für "Unbeweglichkeit" (Modell IV, 2003)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|---|----|--------------------|----------------|----------|--------------------|---------------|
| Intercept | 1 | 0.2375 | 0.0949 | 0.0123 | - | - |
| Fallzahl (linear) | 1 | -0.0048 | 0.0002 | < 0.0001 | siehe Tab. 12 | |
| Fallzahl (quadratisch) | 1 | 7.818E-6 | 2.425E-7 | < 0.0001 | | |
| Alter | 1 | -0.00527 | 0.0012 | < 0.0001 | 0.975 ^a | 0.961 – 0.985 |
| Geschlecht (männlich) | 1 | 0.0046 | 0.0212 | 0.8290 | 1.005 | 0.964 – 1.047 |
| ASA-Status (1-2) | 1 | -0.1793 | 0.0202 | < 0.0001 | 0.836 | 0.803 – 0.870 |
| Arthrose-Score (0-3) | 1 | 0.0755 | 0.0245 | 0.0020 | 1.078 | 1.028 – 1.131 |
| Deformität (O-Stellung) | 1 | 0.1153 | 0.0236 | < 0.0001 | 1.122 | 1.072 – 1.175 |
| Deformität (X-Stellung) | 1 | 0.0846 | 0.0286 | 0.0031 | 1.088 | 1.029 – 1.151 |
| Postop. Verweildauer | 1 | -0.0212 | 0.0018 | < 0.0001 | 0.899 ^b | 0.886 – 0.918 |
| Hosmer-Lemeshow-Anpassungstest: 73.7 ($p < 0.0001$) Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.030$; Adjusted-Likelihood-Ratio $R^2 = 0.042$ | | | | | | |
| ^a bezogen auf 5 Jahre, ^b bezogen auf 5 Tage | | | | | | |

Tab. 14: Odds Ratios mit 95%-Konfidenzintervallen für die Fallzahl basierend auf den Modellen I-IV für "Unbeweglichkeit" (2003)

| Intervall | Referenzwert | Modell | | | |
|-----------|--------------|---------------------|---------------------|---------------------|---------------------|
| | | I | II | III | IV |
| 1 – 98 | 49 | 1.518 (1.470-1.567) | 1.524 (1.476-1.574) | 1.537 (1.489-1.588) | 1.612 (1.409-1.845) |
| 99 – 293 | 196 | 1.000 | 1.000 | 1.000 | 1.000 |
| 294 - 941 | 617 | 1.938 (1.821-2.063) | 1.927 (1.810-2.051) | 1.892 (1.777-2.015) | 1.877 (1.066-3.306) |

Ähnlich wie bei den Daten aus dem Jahr 2004 lässt sich auch bei den Daten des Jahres 2003 der Zusammenhang des Risikos einer postoperativen "Unbeweglichkeit" und der Anzahl behandelter Patienten mit Hilfe eines logistischen Regressionsmodells mit FP (Modell I) adäquat beschreiben. Die Adjustierung nach potentiellen Risikofaktoren und die Berücksichtigung von Clustereffekten führte zu keiner relevanten Veränderung des Zusammenhangs zwischen Risiko und Fallzahl.

In Abbildung 5 ist der Risikoverlauf in Abhängigkeit von der Fallzahl basierend auf Modell I dargestellt. Eingezeichnet sind mittleres Risiko sowie der von den Fachexperten festgelegte Irrelevanzbereich und das akzeptable absolute Risiko.

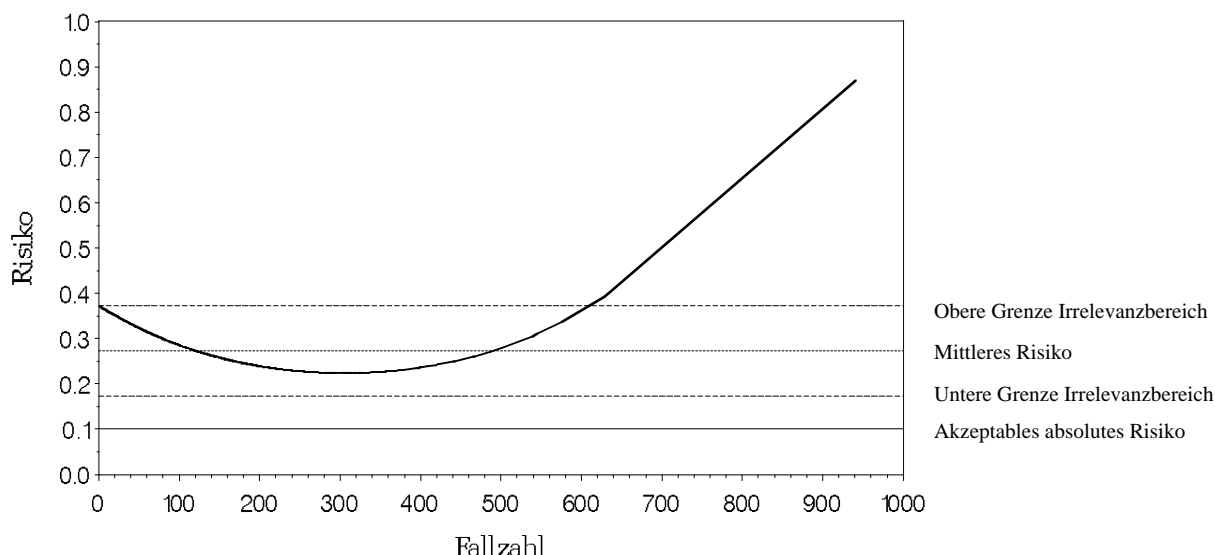


Abb. 5: Risikoverlauf für "Unbeweglichkeit" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos (2003)

Der Risikoverlauf ähnelt dem aus dem Jahr 2004. Für Fallzahlen unterhalb von etwa 620 liegt die Kurve innerhalb des Irrelevanzbereichs und für höhere Fallzahlen oberhalb der oberen Irrelevanz-Grenze. Die gesamte Kurve verläuft deutlich oberhalb des von den Fachexperten angegebenen, akzeptablen absoluten Risikos. Diese Ergebnisse bestätigen, dass es nicht möglich ist, aus den vorhandenen Daten der BQS bezüglich des Qualitätsindikators "Unbeweglichkeit" Schwellenwerte abzuleiten, die sich sinnvoll als Mindestmengen eignen würden.

5.2.3 Untersuchung der Mindestmenge MM=50

Verwendet man zur Definition von Low- und High-Volume-Krankenhäusern die eingeführte Mindestmenge von MM=50, so ergibt sich aus den Daten von 2004 ein geschätztes mittleres Risiko über alle Patienten von $\pi=26.9\%$ und für die Patienten aus High-Volume-Krankenhäusern $\pi_{HV}=26.2\%$. Damit ergibt sich eine geschätzte Risikoreduktion von $\delta = 0.7\%$ (**95%-KI: 0.6 bis 0.8%**). Aus den Daten von 2003 erhält man $\pi=27.2\%$, $\pi_{HV}=26.4\%$ und $\delta = 0.8\%$ (95%-KI: 0.7 bis 0.9%). Die mit der Mindestmenge MM=50 assoziierte Risikoreduktion von 0.7% für 2004 bzw. 0.8% für 2003 ist damit um ein Vielfaches kleiner als die geforderte Risikoreduktion von 10%, die vorab als klinisch relevant festgelegt wurde.

5.2.4 Schlussfolgerungen

Es ist nicht möglich, aus den vorhandenen Daten für den Qualitätsindikator "Unbeweglichkeit" sinnvolle Schwellenwerte abzuleiten, die als Mindestmengen für die Knie-TEP diskutiert werden können. Hierfür gibt es folgende Gründe. Erstens ist der Risikoverlauf in Abhängigkeit von der Fallzahl U-förmig. Alle im Berichtsplan vorgesehenen Methoden zur Ableitung von Schwellenwerten für Mindestmengen setzen einen monotonen Kurvenverlauf voraus. Zweitens sind die Risikoreduktionen, die mit einer festgelegten Mindestmenge assoziiert sind, nicht klinisch relevant. Die geschätzten Werte für die Mindestmenge MM=50 liegen um mehr als den Faktor 10 unterhalb der Relevanzgrenze von 10%. Ausgehend von dem gefundenen Risikoverlauf mit dem starken Anstieg ab einer Fallzahl von 500, wären vermutlich eher "Maximalmengen" mit einer deutlicheren Risikoreduktion assoziiert. Eine solche Berechnung war jedoch nicht Gegenstand des Auftrags und wurde daher nicht weiter verfolgt.

5.3 Ergebnisse für den Indikator "Infektion"

5.3.1 Zusammenhang zwischen Fallzahl und Risiko (2004)

Das mittlere Risiko einer postoperativen "Infektion" lag bei 0.79% (870 von 110349 Patienten). In Abbildung 6 ist der Risikoplot in Abhängigkeit von der Fallzahl dargestellt. Die Punktwolke weist einen leicht monoton fallenden Trend auf.

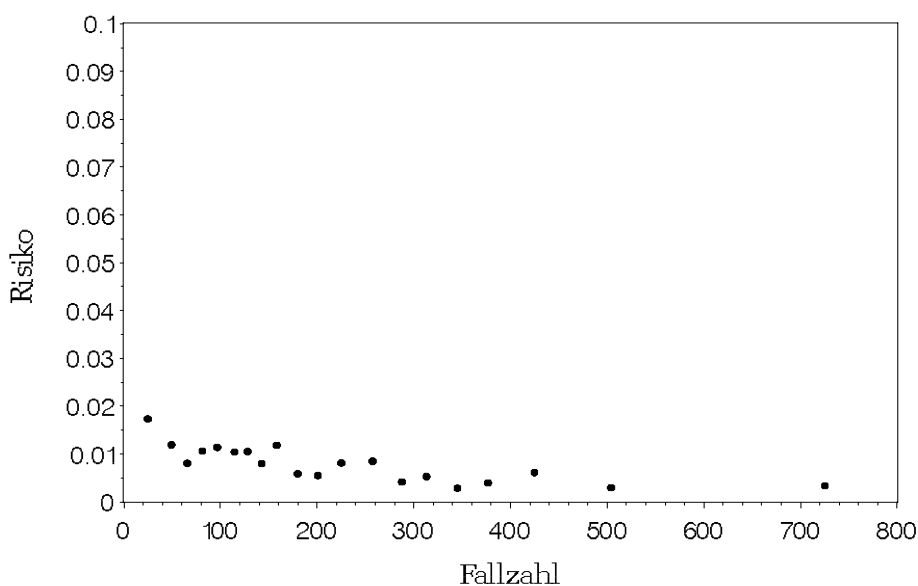


Abb. 6: Risikoplot für "Infektion" vs. Fallzahl (2004)

Es wurde zunächst ein einfaches logistisches Regressionsmodell mit der Fallzahl als Einflussfaktor berechnet. Die Ergebnisse sind in Tabelle 15 dargestellt (Modell I). Zwar ist die Assoziation zwischen Fallzahl und Infektionsrisiko deutlich statistisch signifikant ($p < 0.0001$), aber die Fallzahl hat nur einen sehr niedrigen Erklärungswert für das Infektionsrisiko ($r^2 = 0.001$, $R^2 = 0.01$).

Tab. 15: Einfaches logistisches Regressionsmodell für "Infektion" (Modell I, 2004)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|---|----|--------------------|----------------|----------|--------------------|---------------|
| Intercept | 1 | -4.2649 | 0.0570 | < 0.0001 | - | - |
| Fallzahl | 1 | -0.0029 | 0.0003 | < 0.0001 | 0.861 ^a | 0.861 – 0.905 |
| Hosmer-Lemeshow-Anpassungstest: 14.4 ($p = 0.07$) | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.001$; Adjusted-Likelihood-Ratio $R^2 = 0.01$ | | | | | | |
| ^a bezogen auf 50 Fälle | | | | | | |

Abbildung 7 zeigt den aus dem Modell geschätzten Risikoverlauf. Das Risiko fällt monoton mit steigender Fallzahl, allerdings ist die absolute Risikoreduktion gering. Dies deckt sich mit den Ergebnissen eines zwar statistisch signifikanten Zusammenhangs zwischen Fallzahl und Risiko, aber einem sehr geringen Erklärungswert der Fallzahl.

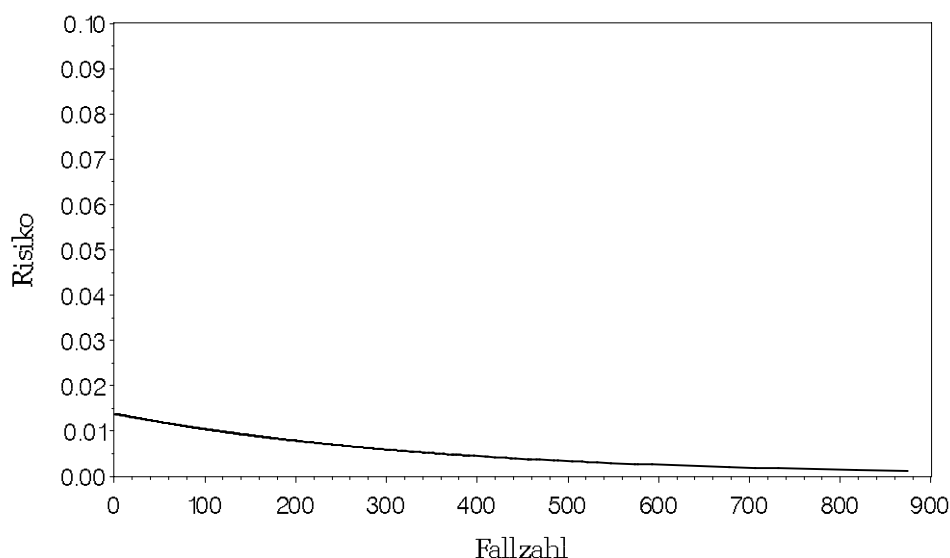


Abb. 7: Risikoverlauf für "Infektion" aus Modell I (2004)

Zur Untersuchung eines möglichen Einflusses potentieller Risikofaktoren wurde in einem nächsten Schritt eine entsprechende Adjustierung, im Rahmen eines multiplen logistischen Regressionsmodells, vorgenommen. Die Ergebnisse dieses Modells sind in Tabelle 16 dargestellt (Modell II).

Durch Berücksichtigung der Risikofaktoren verbessern sich die Modellanpassung und erklärte Variabilität geringfügig. Der Hosmer-Lemeshow-Test ($p=0.12$) zeigt keine Hinweise auf wesentliche Abweichungen des Modells von den Daten. Parameterschätzer und Standardfehler für die Fallzahl sind nahezu identisch im Vergleich zu Modell I. Ein Confounding durch die betrachteten Einflussfaktoren liegt daher nicht vor, was auch durch die in Abbildung 8 gezeigte adjustierte Risikokurve veranschaulicht wird.

Tab. 16: Multiples logistisches Regressionsmodell mit Kovariablen für "Infektion"
(Modell II, 2004)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|--|----|--------------------|----------------|----------|--------------------|---------------|
| Intercept | 1 | -5.1837 | 0.3078 | < 0.0001 | - | - |
| Fallzahl | 1 | -0.0028 | 0.0003 | < 0.0001 | 0.861 ^a | 0.861 – 0.905 |
| Alter | 1 | 0.0115 | 0.0043 | 0.0075 | 1.061 ^b | 1.015 – 1.104 |
| Geschlecht (männlich) | 1 | -0.2364 | 0.0724 | 0.0011 | 0.789 | 0.685 – 0.910 |
| ASA-Status (1-2) | 1 | 0.6056 | 0.0692 | < 0.0001 | 1.832 | 1.600 – 2.099 |
| Hosmer-Lemeshow-Anpassungstest: 12.8 ($p = 0.12$) | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.002$; Adjusted-Likelihood-Ratio $R^2 = 0.025$ | | | | | | |
| ^a bezogen auf 50 Fälle, ^b bezogen auf 5 Jahre | | | | | | |

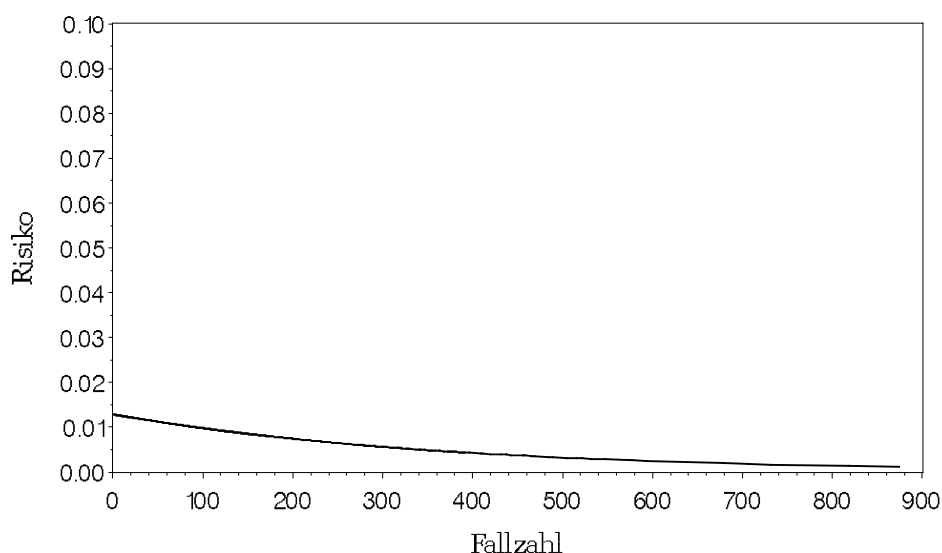


Abb. 8: Adjustierter Risikoverlauf für "Infektion" aus Modell II (2004)

Analog zum Vorgehen bei der Untersuchung der "Unbeweglichkeit" wurde in einem nächsten Schritt das Haupteffektmodell 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 II angepasst. Tabelle 17 zeigt die Ergebnisse des GEE-Ansatzes (Modell III). Die Korrelation innerhalb der Krankenhäuser ist sehr gering ($r = 0.005$), die Parameterschätzer verändern sich kaum im Vergleich zu Modell II.

Tab. 17: Ergebnisse des GEE-Ansatzes für "Infektion" (Modell III, 2004)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|---|----|--------------------|----------------|----------|--------------------|----------------|
| Intercept | 1 | -4.7786 | 0.3477 | < 0.0001 | - | - |
| Fallzahl | 1 | -0.0030 | 0.0004 | < 0.0001 | 0.861 ^a | 0.828 – 0.895 |
| Alter | 1 | 0.0116 | 0.0046 | 0.0110 | 1.060 ^b | 1.0130 – 1.109 |
| Geschlecht (männlich) | 1 | 0.2343 | 0.0741 | 0.0016 | 1.264 | 1.093 – 1.462 |
| ASA-Status (1-2) | 1 | -0.6141 | 0.0720 | < 0.0001 | 0.541 | 0.470 – 0.623 |
| Exchangeable Kovarianzstruktur: $r = 0.005$ | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.002$ | | | | | | |
| ^a bezogen auf 50 Fälle, ^b bezogen auf 5 Jahre | | | | | | |

Abbildung 9 zeigt den Risikoverlauf in Abhängigkeit von der Fallzahl basierend auf dem einfachen Modell I, das den Zusammenhang zwischen Fallzahl und Risiko einer postoperativen "Infektion" in adäquater Weise beschreibt (vgl. Abb. 8). Eingezeichnet sind außerdem das mittlere Risiko sowie der von den Fachexperten festgelegte Irrelevanzbereich und das akzeptable absolute Risiko. Zur deutlicheren Lesbarkeit ist die Y-Achse im Vergleich zu den Abbildungen 7 und 8 stark gedehnt dargestellt, was bei der Interpretation beachtet werden muss. Der dargestellte Bereich umfasst nur 3% des gesamten Wertebereichs eines Risikos.

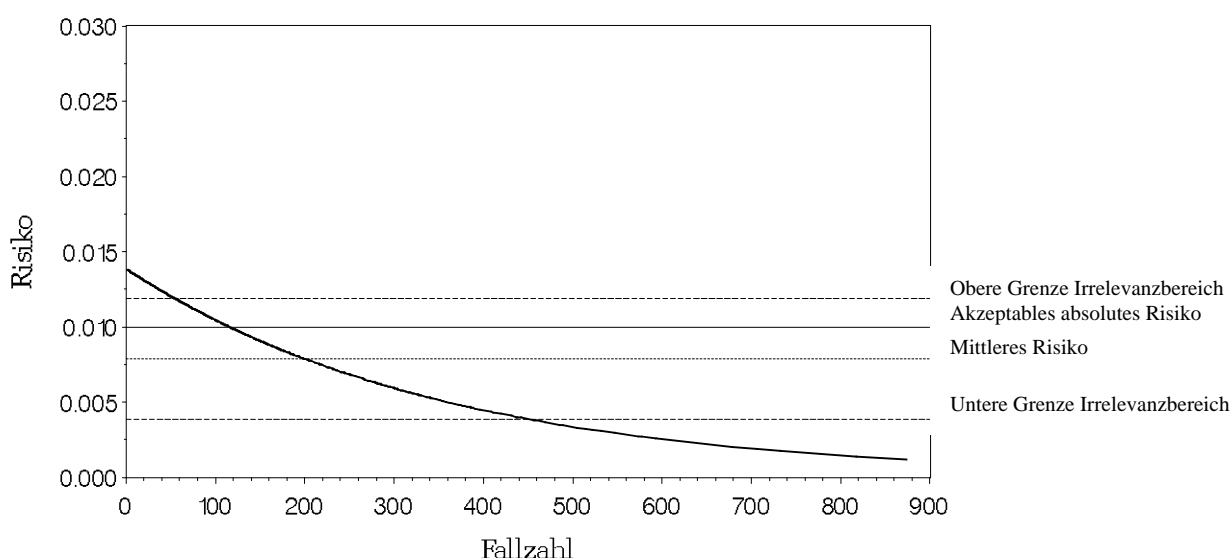


Abb. 9: Risikoverlauf für "Infektion" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos bei stark gedehnter Y-Achse (2004)

5.3.2 Zusammenhang zwischen Fallzahl und Risiko 2003

Zur Untersuchung der Robustheit der Ergebnisse aus dem Jahr 2004 wurden die Daten aus dem Jahr 2003 in analoger Form ausgewertet. Bei 827 von 90004 Patienten (0.9%) trat eine postoperative "Infektion" auf. Die Ergebnisse der angepassten Modelle sind in den Tabellen 18 bis 20 aufgeführt.

Tab. 18: Einfaches logistisches Regressionsmodell für "Infektion" (Modell I, 2003)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|--|----|--------------------|----------------|----------|--------------------|---------------|
| Intercept | 1 | -4.3687 | 0.0586 | < 0.0001 | - | - |
| Fallzahl | 1 | -0.0016 | 0.0003 | < 0.0001 | 0.905 ^a | 0.905 – 0.951 |
| Hosmer-Lemeshow-Anpassungstest: 11.5 ($p = 0.18$) | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.0005$; Adjusted-Likelihood-Ratio $R^2 = 0.0047$ | | | | | | |
| ^a bezogen auf 50 Fälle | | | | | | |

Tab. 19: Multiples logistisches Regressionsmodell mit Kovariablen für "Infektion" (Modell II, 2003)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|--|----|--------------------|----------------|----------|--------------------|---------------|
| Intercept | 1 | -4.4981 | 0.3049 | < 0.0001 | - | - |
| Fallzahl | 1 | -0.0016 | 0.0003 | < 0.0001 | 0.990 ^a | 0.990 – 0.995 |
| Alter | 1 | 0.0003 | 0.0043 | 0.9513 | 1.000 ^a | 0.961 – 1.046 |
| Geschlecht (männl.) | 1 | -0.1281 | 0.0754 | 0.0895 | 0.880 | 0.759 – 1.020 |
| ASA-Status (1-2) | 1 | 0.4975 | 0.0713 | < 0.0001 | 1.645 | 1.430 – 1.891 |
| Hosmer-Lemeshow-Anpassungstest: 12.4 ($p = 0.14$) | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.0013$; Adjusted-Likelihood-Ratio $R^2 = 0.0105$ | | | | | | |
| ^a bezogen auf 50 Fälle, ^b bezogen auf 5 Jahre | | | | | | |

Tab. 20: Ergebnisse des GEE-Ansatzes für "Infektion" (Modell III, 2003)

| Parameter | FG | Parameterschätzung | Standardfehler | p-Wert | OR | 95%-KI |
|---|----|--------------------|----------------|----------|--------------------|---------------|
| Intercept | 1 | -4.0177 | 0.3542 | < 0.0001 | - | - |
| Fallzahl | 1 | -0.0021 | 0.0004 | < 0.0001 | 0.990 ^a | 0.986 – 0.993 |
| Alter | 1 | 0.0001 | 0.0046 | 0.9855 | 1.001 ^b | 0.956 – 1.047 |
| Geschlecht (männl.) | 1 | 0.1320 | 0.0828 | 0.1111 | 1.141 | 0.970 – 1.342 |
| ASA-Status (1-2) | 1 | -0.5239 | 0.0736 | < 0.0001 | 0.592 | 0.513 – 0.684 |
| Exchangeable Kovarianzstruktur: $r = 0.005$ | | | | | | |
| Erklärte Variabilität: Squared-Pearson-Correlation $r^2 = 0.0013$ | | | | | | |
| ^a bezogen auf 50 Fälle, ^b bezogen auf 5 Jahre | | | | | | |

Bezüglich der Parameterschätzungen für die Fallzahl zeigen sich erneut keine gravierenden Unterschiede zwischen den Modellen, so dass – ähnlich wie bei den

Daten aus dem Jahr 2004 – der Zusammenhang des Risikos einer postoperativen "Infektion" und der Anzahl behandelter Patienten mit Hilfe eines einfachen logistischen Regressionsmodells (Modell I) darstellbar ist. Abbildung 10 zeigt den Risikoverlauf in Abhängigkeit von der Fallzahl basierend auf Modell I mit stark gedehnter Y-Achse. Eingezeichnet sind mittleres Risiko sowie der von den Fachexperten festgelegte Irrelevanzbereich und das akzeptable absolute Risiko.

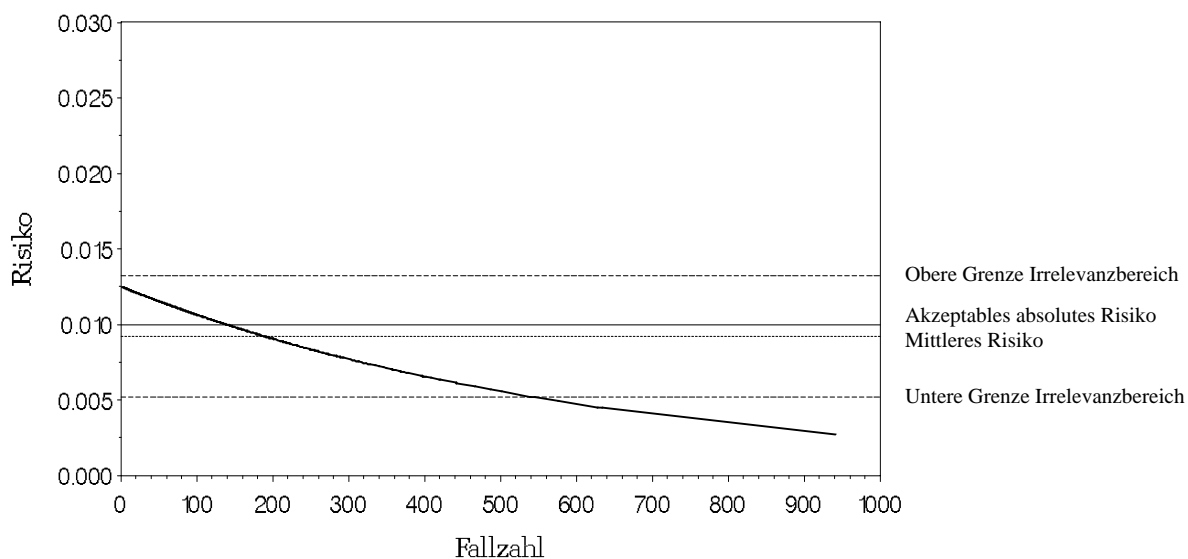


Abb. 10: Risikoverlauf für "Infektion" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos bei stark gedehnter Y-Achse (2003)

5.3.3 Berechnung des Benchmarks VARL

Bei der gefundenen sehr flachen Risikokurve ist mit den vorhandenen Festlegungen der Fachexperten zunächst der VARL-Ansatz nach Bender [14] prinzipiell anwendbar. Hierfür wurde als akzeptables absolutes Risiko der Wert 1% vorgegeben (siehe Abschnitt 4.2). Die geschätzten zugehörigen Benchmarks VARL mit den zugehörigen 95%-Konfidenzintervallen für die Jahre 2004 und 2003 sind in der folgenden Tabelle 21 dargestellt.

Tab. 21: Ergebnisse des VARL-Ansatzes nach Bender [14] für "Infektion"

| Jahr | Vorgegebenes Risiko: 1% | | |
|------|-------------------------|----------------|------------------------|
| | VARL-Wert | Standardfehler | 95%-Konfidenzintervall |
| 2004 | 116 | 13.0 | 90 - 141 |
| 2003 | 139 | 22.2 | 95 - 182 |

Für die Daten aus dem Jahr 2004 ist der errechnete Benchmark VARL in Abbildung 11 grafisch veranschaulicht, wobei die Y-Achse wieder stark gedehnt dargestellt ist.

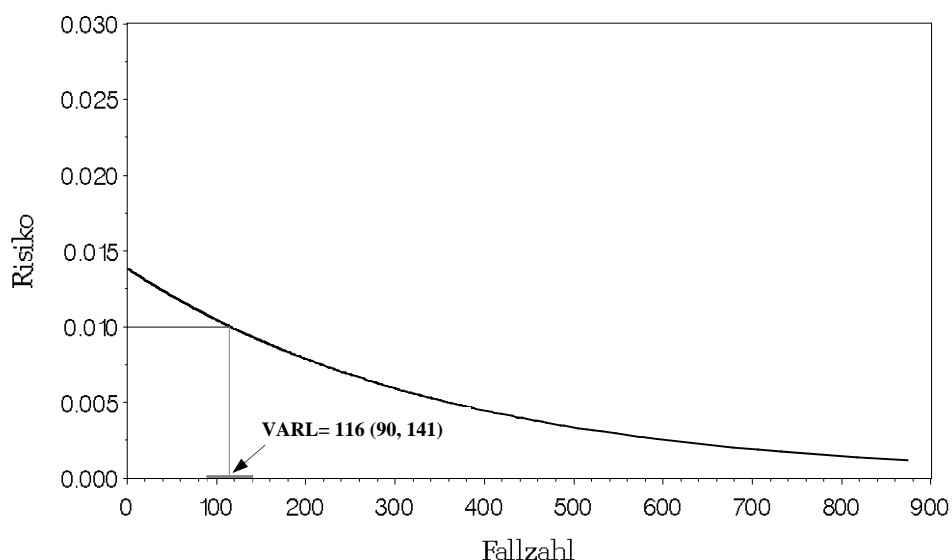


Abb. 11: Risikoverlauf für "Infektion" nach Modell I mit Kennzeichnung des berechneten VARL-Wertes mit 95%-KI bei stark gedehnter Y-Achse (2004)

Verwendet man zur Definition von Low- und High-Volume-Krankenhäusern den geschätzten Benchmark VARL=116, so ergibt sich aus den Daten von 2004 ein geschätztes mittleres Risiko über alle Patienten von $\pi=0.79\%$ und für die Patienten aus High-Volume-Krankenhäusern $\pi_{HV}=0.64\%$. Damit ergibt sich eine geschätzte Risikoreduktion von $\delta = 0.15\%$ (95%-KI: 0.11 bis 0.19%). Aus den Daten des Jahres 2003 erhält man für den Benchmark VARL=139 die Ergebnisse $\pi=0.92\%$, $\pi_{HV}=0.74\%$ und $\delta = 0.18\%$ (95%-KI: 0.13 bis 0.23%).

Die mit diesen Benchmarks assoziierten Risikoreduktionen von 0.15% für 2004 bzw. 0.18% für 2003 sind damit deutlich kleiner als die geforderte Risikoreduktion von 0.4%, die vorab als klinisch relevant festgelegt wurde. Dies gilt auch noch für die oberen Konfidenzgrenzen. Damit stellen diese Benchmark-Werte keine Mindestmengen dar, für die nach den vorliegenden BQS-Daten eine klinisch relevante Qualitätsverbesserung zu erwarten ist.

Eine Mindestmenge, die mit einer relevanten Qualitätsverbesserung bezüglich des Infektionsrisikos assoziiert ist, müsste den Wert **MM=400** oder mehr erreichen (Daten aus 2004). Hierfür würde die Risikoreduktion von $\delta = 0.42\%$ erreicht. Da Schwellenwerte oder Benchmarks unterhalb dieser Schranke keine Größenordnung von genügend hoher klinischer Relevanz aufweisen, wird auf die Anwendung der anderen möglichen Verfahren zur Ableitung von Schwellenwerten oder Benchmarks für den sekundären Zielparameter verzichtet.

5.3.4 Untersuchung der Mindestmenge MM=50

Verwendet man zur Definition von Low- und High-Volume-Krankenhäusern die eingeführte Mindestmenge von MM=50, so ergibt sich aus den Daten von 2004 ein geschätztes mittleres Risiko über alle Patienten von $\pi = 0.79\%$ und für die Patienten aus High-Volume-Krankenhäusern $\pi_{HV} = 0.72\%$. Damit ergibt sich eine geschätzte Risikoreduktion von $\delta = 0.07\%$ (**95%-KI: 0.04 bis 0.09%**). Aus den Daten des Jahres 2003 erhält man $\pi = 0.92\%$, $\pi_{HV} = 0.87\%$ und $\delta = 0.05\%$ (95%-KI: 0.02 bis 0.07%).

Die mit der Mindestmenge MM=50 assoziierten Risikoreduktionen von 0.07% für 2004 bzw. 0.05% für 2003 sind damit deutlich kleiner als die geforderte Risikoreduktion von 0.4%, die vorab als klinisch relevant festgelegt wurde. Dies gilt auch noch für die oberen Konfidenzgrenzen. Damit stellt die eingeführte Mindestmenge von MM=50 keine Maßnahme dar, für die aus den BQS-Daten bezüglich der gewählten Qualitätsindikatoren eine klinisch relevante Qualitätsverbesserung abzuleiten ist.

5.3.5 Schlussfolgerungen

Auch für den sekundären Qualitätsindikator "Infektion" ist aus den Daten des Bundesdatenpools 2004 und 2003 der BQS kein sinnvoller Schwellenwert ableitbar, der als Mindestmenge mit einer klinisch relevanten Qualitätsverbesserung assoziiert ist und sich als hypothetischer Wert zur weiteren Untersuchung in einer Interventionsstudie anbieten würde. Es konnte zwar ein statistisch signifikanter negativer Zusammenhang zwischen der Fallzahl und dem Infektionsrisiko nachgewiesen werden, die entsprechende Risikokurve ist jedoch sehr flach und der Erklärungswert der Fallzahl letztlich zu gering, um einen sinnvollen Schwellenwert abzuleiten, der mit einer klinisch relevanten Qualitätsverbesserung assoziiert ist. Weder der nach dem VARL-Ansatz abgeleitete Wert von 116 Fällen, noch die bereits eingeführte Mindestmenge von 50 Fällen ist mit einer klinisch bedeutsamen Reduktion des Risikos assoziiert.

6. Diskussion

Bei der Bewertung der in diesem Bericht beschriebenen Ergebnisse sind eine Reihe von Aspekten 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 Knie-TEP 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 nicht überprüfbar sind. Für den primären Qualitätsindikator "Unbeweglichkeit" gibt es bei 17% (18635 von 110349) der Patienten aus 2004 fehlende Werte. Dies ist nach Aussagen der BQS nicht alleine auf fehlende Angaben sondern auch auf unplausible Angaben der Krankenhäuser zurückzuführen, was die Qualität der Dokumentation der Krankenhäuser in Frage stellt. Unplausible Angaben zur Beweglichkeit wurden für diese Auswertung auf fehlenden Wert gesetzt und die Stabilität der Berechnungen durch Sensitivitätsuntersuchungen analysiert. Ein weiteres Problem stellt der unklare Messzeitpunkt der "Unbeweglichkeit" dar, der nicht unbedingt

auf das Ende der Verweildauer fällt. Dadurch war es nicht möglich, den Indikator "Unbeweglichkeit" unter Berücksichtigung des Messzeitpunkts sinnvoller zu definieren. Es ist auch nicht auszuschließen, dass der unerwartete U-förmige Verlauf des Risikos für "Unbeweglichkeit" in Abhängigkeit von der Fallzahl durch Confounding mit einer im Datensatz nicht vorhandenen Variablen hervorgerufen wurde. Eine Berücksichtigung der vorhandenen möglichen Confounder (inklusive der Verweildauer) im Rahmen multipler Regressionsmodelle änderte jedoch nichts an dem U-förmigen Verlauf der Risikokurve. Aufgrund der bekannten Einschränkungen der Daten sind jedoch die Ergebnisse entsprechend vorsichtig zu interpretieren.

Auch wenn der sekundäre Qualitätsindikator "Infektion" vollständig vorhanden ist, gelten auch hier ähnliche Einschränkungen. Die grundsätzliche Datenqualität bleibt unklar und es kann ein Confounding durch nicht erhobene Variablen nicht ausgeschlossen werden. Hätte man in der vorliegenden Auswertung in der Beziehung zwischen Infektionsrisiko und Fallzahl deutliche Schwellenwerte gefunden, so wäre dies keinesfalls eine ausreichende Evidenz dafür gewesen, dass die Verwendung dieses Schwellenwerts als Mindestmenge automatisch auch zu einer entsprechenden Qualitätsverbesserung führt. Ein solcher evidenz-basierter Nachweis kann nur über eine kontrollierte Interventionsstudie geführt werden und nicht über den Nachweis einer statistischen Assoziation aus Routinedaten. Die gefundene monoton fallende Beziehung zwischen dem Infektionsrisiko und der Fallzahl war jedoch zu flach und ohne deutliche Schwellenwerte, so dass hier hypothesengenerierend keine Werte ableitbar waren, die sich möglicherweise als Mindestmengen eignen. Dieses Ergebnis aus den verfügbaren deutschen Daten steht in Übereinstimmung mit den Resultaten aus internationalen Daten, dass sich bei der Knie-TEP kein klarer einheitlicher Grenzwert bezüglich der Fallzahl pro Krankenhaus und Jahr für die Unterscheidung zwischen höherer und geringerer Qualität finden lässt [17,18].

7. Zusammenfassung

Mit Hilfe von Daten des Bundesdatenpools der BQS aus den Jahren 2004 und 2003 von Patienten mit Knie-Totalendoprothesen-Erstimplantation wurde der Zusammenhang zwischen dem Risiko für "Unbeweglichkeit" (primärer Qualitätsindikator) bzw. dem Risiko für "Infektion" (sekundärer Qualitätsindikator) und der Fallzahl des betreffenden Krankenhauses pro Jahr mit Hilfe von logistischen Regressionsmodellen untersucht. Für beide Indikatoren konnte ein signifikanter Zusammenhang zwischen der Fallzahl und dem entsprechenden Risiko statistisch nachgewiesen werden. Weitere Risikofaktoren wie z.B. Alter und ASA-Status hatten zwar einen signifikanten Einfluss, stellten aber keine wichtigen Confounder dar. Ein deutlicher Cluster-Effekt, der die Beziehung zwischen Risiko und Fallzahl verändert, konnte ebenso wenig gefunden werden.

Der Zusammenhang zwischen dem primären Qualitätsindikator "Unbeweglichkeit" und der Fallzahl zeigte unerwartet einen U-förmigen Verlauf, der das Konzept einer Mindestmengenregelung für den Leistungsbereich Knie-TEP bezüglich des Risikos für "Unbeweglichkeit" in Frage stellt. Hierfür wurden daher keine Schwellenwerte berechnet.

Die Volume-Outcome-Beziehung für den sekundären Qualitätsindikator "Infektion" zeigte eine sehr flache, mit steigender Fallzahl sehr langsam fallende Risikokurve, die die Hypothese unterstützt, dass High-Volume-Krankenhäuser eine bessere Qualität aufweisen als Low-Volume-Krankenhäuser. Der Erklärungswert der Fallzahl war jedoch zu gering, um aus dieser Beziehung einen Schwellenwert abzuleiten, der mit einer klinisch relevanten Qualitätsverbesserung assoziiert ist.

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] 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.
- [4] 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.
- [5] Hosmer DW, Lemeshow S. *Applied Logistic Regression (2nd Ed.)*. New York: Wiley, 2000.
- [6] Bender R, Ziegler A, Lange S. Logistische Regression. *Dtsch Med Wochenschr* 2002; 127: T11-T13.
- [7] 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.
- [8] Kleinbaum DG, Klein M. *Logistic Regression: A Self-Learning Text (2nd Ed.)*. New York: Springer, 2002.
- [9] 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.
- [10] 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.
- [11] Nagelkerke NJD. A note on the general definition of the coefficient of determination. *Biometrika* 1991; 78: 691-692.

- [12] Mittlböck M, Schemper M. Explained variation for logistic regression. *Stat Med* 1996; 15: 1987-1997.
- [13] Pregibon D. Logistic regression diagnostics. *Ann Stat* 1981; 9: 705-724.
- [14] Bender R. Quantitative risk assessment in epidemiological studies investigating threshold effects. *Biometrical J* 1999; 41: 305-319.
- [15] Newcombe RG. Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Stat Med* 1998; 17: 873-890.
- [16] Bender R. Calculating confidence intervals for the number needed to treat. *Control Clin Trials* 2001; 22: 102-110.
- [17] Klakow-Franck R, Wetzel H. Mindestmengenregelung ohne Mindestmengen. *Dtsch Ärzteztg* 2004; 101: A-2926-2928.
- [18] Schröder P, Rath T. Mindestmengen in der Kniegelenkendoprothetik. *Orthopäde* 2005; 34: 198-209.

Anhang: Weiterer Zeitplan

| Arbeitsschritt | Termin (geplant) |
|--|--|
| Einholung von Stellungnahmen | Bis 15.11.2005 (drei Wochen nach Veröffentlichung des Vorberichts) |
| Wissenschaftliche Anhörung | Dezember 2005 |
| Weitergabe des Abschlussberichts an den G-BA | Dezember 2005 |
| Veröffentlichung des Abschlussberichts | Zwei Monate nach Weitergabe an den G-BA |