

IQWiG-Berichte · Jahr: 2005 Nr. 2

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

Abschlussbericht

Auftrag B05/01a
Version 1.0
Stand: 05.12.2005

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

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

Interne Auftragsnummer:

B05/01a

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 Knie-Totalendoprothese**

- Abschlussbericht vom 05.12.2005 -

[Auftrag B05/01a]

Version 1.0

Stand: 05.12.2005

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
Externer Sachverständiger: 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

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	11
4.1 Datenquellen	11
4.2 Vorgaben der Fachexperten	11
4.3 Auswertungen	12
4.3.1 Datenbeschreibung.....	12
4.3.2 Modellbildung.....	14
4.3.3 Berechnung von Schwellenwerten.....	15
4.3.4 Darstellung der klinischen Relevanz	15
4.3.5 Multiples Testen und Signifikanzniveau	17
4.3.6 Software	17
4.4 Ergänzungen im Vergleich zum Berichtsplan	17
5. Ergebnisse	18
5.1 Deskriptive Statistik.....	18
5.2 Ergebnisse für den Indikator "Unbeweglichkeit"	21
5.2.1 Zusammenhang zwischen Fallzahl und Risiko (2004).....	21
5.2.2 Zusammenhang zwischen Fallzahl und Risiko (2003).....	29
5.2.3 Untersuchung der Mindestmenge MM=50.....	32
5.2.4 Schlussfolgerungen	33
5.3 Ergebnisse für den Indikator "Infektion"	34
5.3.1 Zusammenhang zwischen Fallzahl und Risiko (2004).....	34
5.3.2 Zusammenhang zwischen Fallzahl und Risiko 2003.....	38
5.3.3 Berechnung des Benchmarks VARL.....	39
5.3.4 Untersuchung der Mindestmenge MM=50.....	41
5.3.5 Schlussfolgerungen	42
6. Diskussion	43
7. Zusammenfassung	44
Literatur	46
Anhang: Stellungnahmen.....	48
A1. Stellungnahme der BQS.....	48
A2. Stellungnahme von Dr. Schröder	54

Tabellenverzeichnis

Tabelle 1	Deskriptive Übersicht
Tabelle 2	Rangkorrelationen 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, 2003)
Tabelle 21	Ergebnisse des VARL-Ansatzes für "Infektion"

Abbildungsverzeichnis

Abbildung 1	Verteilung der Fallzahl pro Krankenhaus (2003 oben, 2004 unten)
Abbildung 2	Risikoplot für "Unbeweglichkeit" vs. Fallzahl (2004)
Abbildung 3	Risikoverlauf mit 95%-Konfidenzbändern für "Unbeweglichkeit" aus Modell I (2004)
Abbildung 4	Adjustierter Risikoverlauf für "Unbeweglichkeit" aus Modell II (2004)
Abbildung 5	Risikoverlauf für "Unbeweglichkeit" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos (2004)
Abbildung 6	Risikoverlauf für "Unbeweglichkeit" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos (2003)
Abbildung 7	Risikoplot für "Infektion" vs. Fallzahl (2004)
Abbildung 8	Risikoverlauf mit 95%-Konfidenzbändern für "Infektion" aus Modell I (2004)
Abbildung 9	Adjustierter Risikoverlauf für "Infektion" aus Modell II (2004)
Abbildung 10	Risikoverlauf für "Infektion" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos bei stark gedehnter Y-Achse (2004)
Abbildung 11	Risikoverlauf für "Infektion" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos bei stark gedehnter Y-Achse (2003)
Abbildung 12	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	BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH
CDC	Centers for Disease Control and Prevention
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
NNT	Number Needed to Treat
OR	Odds Ratio
PIN	Population Impact Number
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 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-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 BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH (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. Die Schritte 5 und 8 erfolgten hierbei zeitgleich nach Schritt 7, da es zum einen bei der Festlegung der Vorgaben durch die Fachexperten erhebliche Verzögerungen gab und zum anderen aufgrund der Datenlage und der Zwischenergebnisse weitere Festlegungen zur Berechnung anderer Schwellenwerte nicht mehr notwendig waren.

- 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)

Die erforderlichen Daten für die Bearbeitung des Auftrags wurden von der BQS am 11.07.2005 geliefert. Der Berichtsplan zu diesem Auftrag wurde am 18.07.2005 fertig gestellt und am 25.07.2005 veröffentlicht. Der Vorbericht wurde am 17.10.2005 fertig gestellt und am 25.10.2005 veröffentlicht. Die Frist für Stellungnahmen lief bis zum 15.11.2005. Es gingen zwei Stellungnahmen ein, die diesem Bericht als Anhang beigefügt sind.

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 jeweils auf ein Krankenhaus. Hierbei handelt es sich um die Zahl der für die externe stationäre Qualitätssicherung im Jahr 2003 bzw. 2004 dokumentierten Fälle. Die Daten wurden in pseudonymisierter Form unter Beachtung der amtlichen Datenschutzrichtlinien aufbereitet und statistisch ausgewertet.

4.2 Vorgaben der Fachexperten

Als Fachexperten des G-BA für die Knie-TEP wurden auf Vorschlag der Deutschen Krankenhausgesellschaft und Gesetzlichen Krankenversicherung folgende Personen benannt:

- Dr. Peter Schröder (Essen)
- Prof. Dr. Karl-Dieter Heller (Braunschweig)

Im Rahmen der Auftragskonkretisierung wurden von diesen Fachexperten die Qualitätsindikatoren (Unbeweglichkeit, Wundinfektion) sowie die für eine Risikoadjustierung zu betrachtenden Einflussgrößen festgelegt (Alter, Geschlecht, ASA Score, präoperativer Status). Des Weiteren wurden von den Fachexperten am 10.10.2005 die nachfolgenden Festlegungen bezüglich der Eigenschaften und Effektstärken einer Mindestmengenregelung getroffen. Informationen zu den hier genannten Variablen findet man im Abschnitt 4.3.1. Hierzu wurden den Fachexperten am 27.07.2005 Ergebnisse zu den mittleren Risiken für die beiden

Qualitätsindikatoren aus dem Jahr 2003 in grafisch aufbereiteter Form zur Verfügung gestellt. Darüber hinaus gehende Festlegungen für andere Schwellenwertberechnungen waren nicht notwendig, da die entsprechenden Modelle aufgrund der gefundenen Datenlage nicht berechnet wurden.

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%

Diese Festlegungen, insbesondere diejenigen zu den klinisch relevanten Veränderungen bezüglich des mittleren Risikos aller Patienten aufgrund einer Intervention und dem akzeptablen absoluten Risiko, sind nicht als eindeutig feststehende Grenzen zu interpretieren. Auch andere Festlegungen wären denkbar gewesen. Die oben angegebenen Grenzen sollen bei der Interpretation der nachfolgenden Ergebnisse als Hilfestellung dienen, um die Größenordnungen der präsentierten Risikoverschiebungen besser beurteilen zu können.

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 ausreichende postoperative Beweglichkeit (keine Extension/Flexion von mindestens 0/0/90 bei Entlassung)
- 2) Infektion: postoperative Wundinfektion nach den Definitionen der Centers for Disease Control and Prevention (CDC) [5]

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. Es sei ausdrücklich darauf hingewiesen, dass "Unbeweglichkeit" hier nicht etwa "keine Beweglichkeit" sondern "keine ausreichende Beweglichkeit" bedeutet. Die Beweglichkeit wurde als ausreichend beurteilt, wenn kein Streckdefizit und eine Beugefähigkeit von mind. 90° vorhanden ist (Extension/Flexion von mindestens 0/0/90 bei Entlassung).

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 0 bis 8
(der BQS-Arthrose-Score berechnet sich hierbei als Summe der Punktwerte der Variablen Osteophyten, Gelenkspalt, Sklerose und Deformierung [6])
- (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 [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. 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 [8] 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 [9]. 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) [10] 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 [11,12]. 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. Für den Indikator "Infektion" wurden dabei die ORs für eine Änderung der Fallzahl um 50 Fälle berechnet. Wegen des nicht monotonen

Risikoverlaufs des Indikators "Unbeweglichkeit" ist es nicht möglich, zur Beschreibung des Zusammenhangs zwischen Risiko und Fallzahl ein konstantes OR zu verwenden. Es wurden daher auf Quantilen beruhende Intervalle gebildet und deren Mittelpunkte als Vergleichswerte zur Berechnung von ORs herangezogen [9]. Zur Beschreibung des Erklärungswertes der betrachteten Risikofaktoren wurden zwei Maße für erklärte Variabilität berechnet: Das adjustierte Likelihood-Ratio R^2 nach Nagelkerke [13] sowie die von Mittelböck & Schemper [14] empfohlene Squared-Pearson-Correlation r^2 . Die Robustheit der Modelle wurde mit regressionsdiagnostischen Verfahren überprüft [15].

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 aufgrund des nicht monotonen Risikoverlaufs keine Schwellenwertberechnung in Frage. Bezüglich des Qualitätsindikators "Infektion" wurde der VARL-Ansatz nach Bender [16] zur Ableitung von Schwellenwerten verwendet. Der entsprechende Benchmark VARL wurde, wie bei Bender [16] 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 im Rahmen einer Mindestmengenregelung mit einem klinisch relevanten Effekt assoziiert sind. Hierfür kommen verschiedene Effektmaße in Frage. Die Vorgaben der Fachexperten beziehen sich hierbei auf die Differenz (δ) des mittleren Risikos über alle Patienten (π) und dem mittleren Risiko der Patienten aus High-Volume-Kranken-

häusern (π_{HV}), d.h. $\delta = \pi - \pi_{HV}$, wobei der Trennwert für die Unterscheidung zwischen High-Volume und Low-Volume-Krankenhäusern durch die jeweils betrachtete Mindestmenge gegeben ist. Dieses Maß beurteilt den Effekt einer Intervention mit Blick auf die Grundgesamtheit der Patienten. Der Kehrwert von δ wird neuerdings als Population Impact Number (PIN) bezeichnet [17].

Zur Berechnung von Konfidenzintervallen für δ ist folgende Darstellung hilfreich. 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 $\delta = \omega_{LV} \times \gamma$ mit $\gamma = (\pi_{LV} - \pi_{HV})$. Die Größe γ 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 γ wurde hierbei die Wilson-Score-Methode verwendet [18,19].

Die gewöhnliche Differenz γ der Risiken zwischen Low-Volume- und High-Volume-Krankenhäusern stellt hier zudem ein anderes mögliches Effektmaß dar. Die Differenz γ vergleicht direkt die beiden Risiken π_{LV} und π_{HV} und stellt somit den Effekt einer Maßnahme aus der Sichtweise der Patienten aus Low-Volume-Krankenhäusern dar, allerdings ohne Berücksichtigung des populationsbezogenen Gesamteffekts. Der Kehrwert von γ wird als Number Needed to Treat (NNT) bezeichnet [20]. Sowohl δ als auch γ haben Vor- und Nachteile. Die Vorteile beider Maße liegen in der einfachen Berechenbarkeit. Der Nachteil von δ ist, dass der Effekt einer Intervention niemals größer sein kann als der relative Anteil der Patienten aus Low-Volume-Krankenhäusern. Der Nachteil von γ ist, dass beim Vergleich von zwei bevölkerungsbezogenen Interventionen A und B (z.B. zwei Regelungen mit unterschiedlichen Mindestmengendefinitionen) der mit Hilfe von γ gemessene Effekt von Intervention A größer sein kann als der von B, obwohl die erwartete Ergebnisqualität nach Intervention A geringer ist als die nach Intervention B. Dies kann bei monoton fallenden Risikokurven beim Effektmaß δ nicht vorkommen.

Wir haben bei der Darstellung der klinischen Relevanz der gefundenen Benchmark-Werte die Ergebnisse sowohl für δ als auch für γ sowie deren Kehrwerte angegeben. Bei der Interpretation der Ergebnisse muss berücksichtigt werden, dass δ und γ unterschiedliche Sichtweisen bei der Darstellung von Interventionseffekten beleuchten (mit und ohne Bezug zur Grundgesamtheit). Beide Maße und deren Kehrwerte werden auch verwendet um die klinische Relevanz der am 16.08.2005 festgelegten Mindestmenge von $MM=50$ zu untersuchen.

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 confirmatorische 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. Wie in Kapitel 3 beschrieben, gab es zeitliche Verschiebungen bei der Festlegung der klinisch relevanten Unterschiede. Des Weiteren 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. Ausserdem 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 0-8, ^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. Abbildung 1 zeigt die patientenbezogene Verteilung der Fallzahl pro Krankenhaus in den Jahren 2003 und 2004.

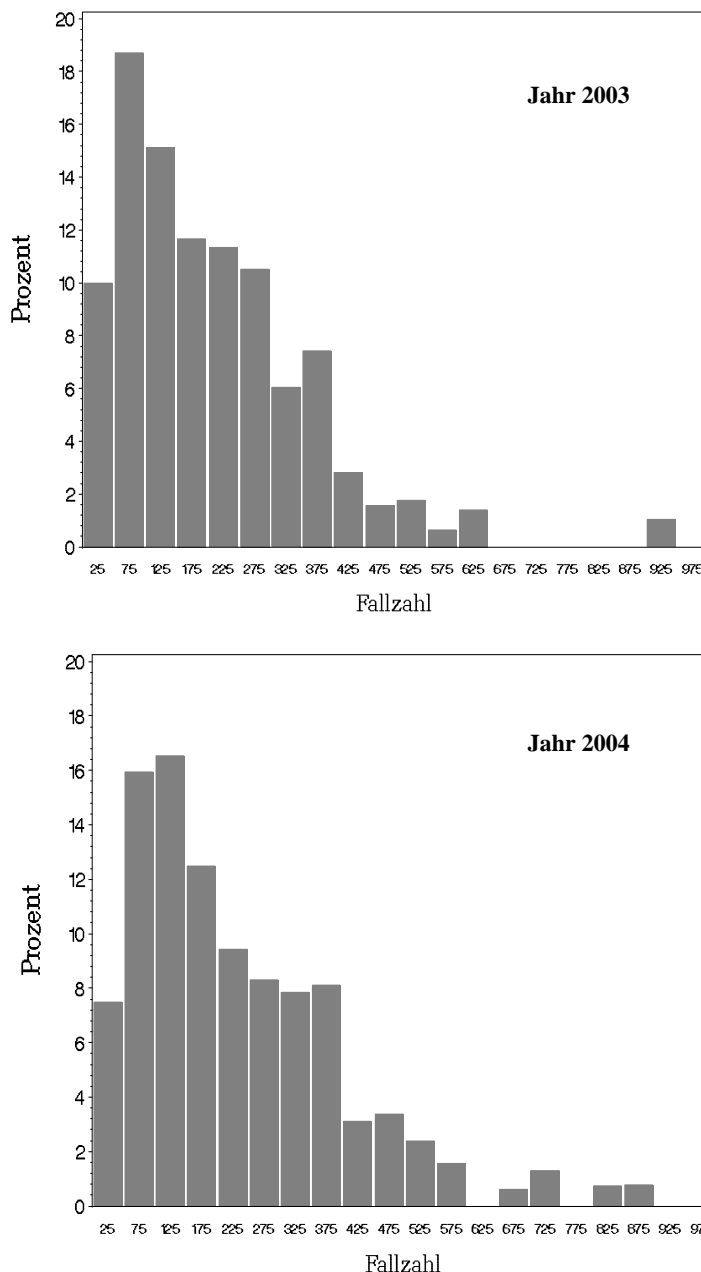


Abb. 1: Verteilung der Fallzahl pro Krankenhaus (2003 oben, 2004 unten)
(Die Zahlen an der horizontalen Achse geben die Klassenmittelpunkte an)

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 in der Verteilung der Daten zwischen den Patienten mit fehlenden und ohne fehlende Angaben zur "Unbeweglichkeit". Lediglich bei der Fallzahl ergab 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 zeigte 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 bei der Analyse des Zusammenhangs zwischen dem Risiko für Unbeweglichkeit und Fallzahl 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%) keine ausreichende Beweglichkeit auf. Abbildung 2 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.

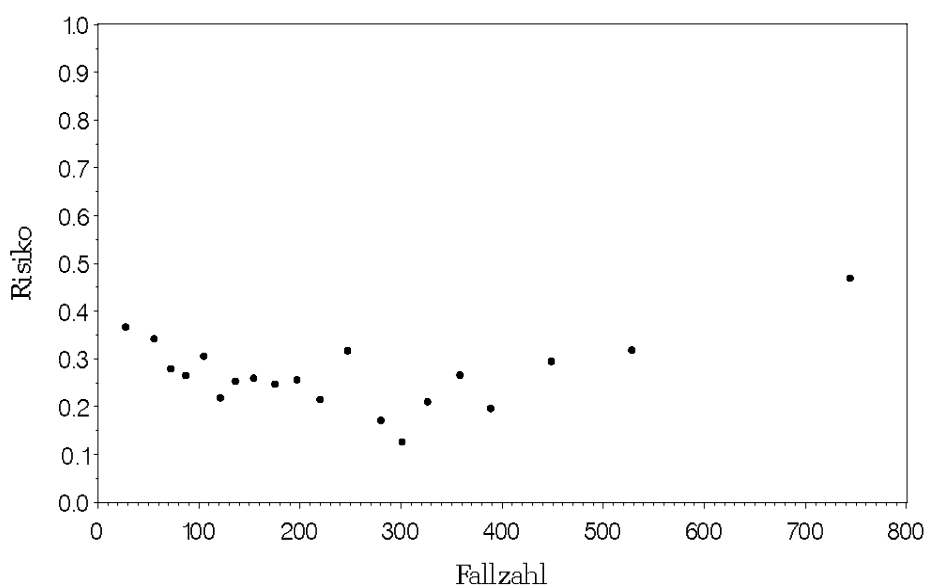


Abb. 2: Risikoplot für "Unbeweglichkeit" vs. Fallzahl (2004)

Zu beachten ist, dass der letzte Punkt rechts in Abbildung 2 nicht auf ein einzelnes Krankenhaus zurückzuführen ist, das möglicherweise einen Ausreißer darstellt. Vielmehr basiert dieser Punkt auf den Ergebnissen von sechs Krankenhäusern mit hohen Fallzahlen, von denen drei ein mittleres und drei ein hohes Risiko aufweisen. Eine entsprechende Grafik ist an dieser Stelle aus datenschutzrechtlichen Gründen leider nicht möglich, da keine einzelnen Krankenhäuser identifizierbar sein dürfen.

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. [9] 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).

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. 9	
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 $R^2 = 0.046$						

Abbildung 3 zeigt den aus der logistischen Regression geschätzten Risikoverlauf mit 95%-Konfidenzbändern bei Verwendung dieser FPs. Die Kurve zeigt einen klar U-förmigen Verlauf, d.h. nach zunächst fallendem Risiko steigt das Risiko mit weiter wachsender Fallzahl wieder an.

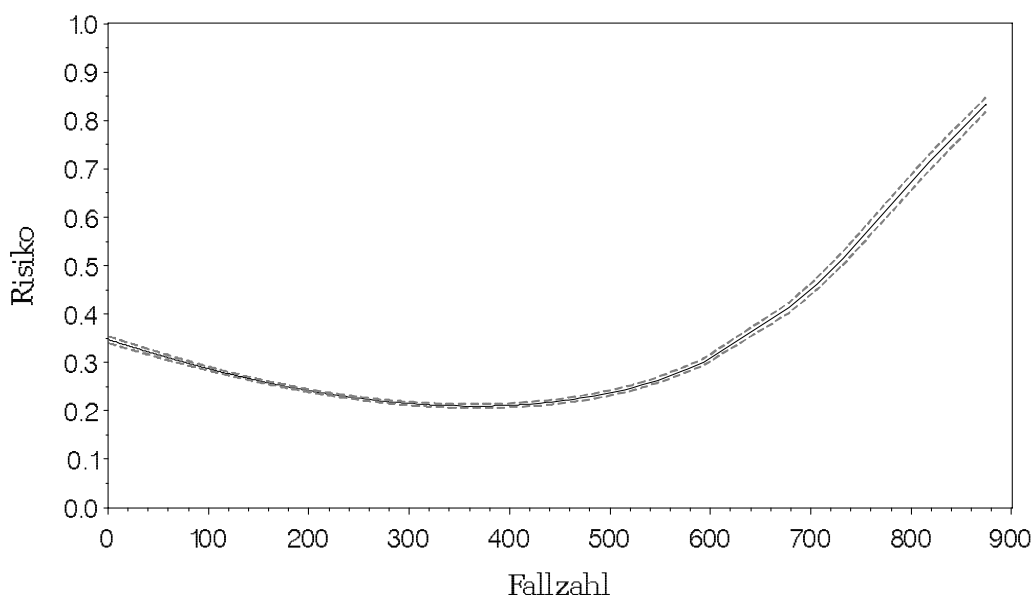


Abb. 3: Risikoverlauf mit 95%-Konfidenzbändern für "Unbeweglichkeit" aus Modell I (2004)

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).

Die Modellanpassungsgüte, gemessen mit dem Hosmer-Lemeshow-Goodness-of-Fit-Test, konnte im Vergleich zu Modell I 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 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 beachtenswertes Confounding durch die betrachteten Einflussfaktoren vorliegt.

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. 9	
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						

Abbildung 4 zeigt den aus Modell II berechneten adjustierten Risikoverlauf, wobei für die Kovariablen jeweils der Mittelwert eingesetzt wurde. Der Risikoverlauf ist dem der nicht adjustierten Analyse (Abb. 3) nahezu identisch. Dies unterstützt obige Aussage, dass die betrachteten Einflussfaktoren keine bedeutsamen Confounder darstellen.

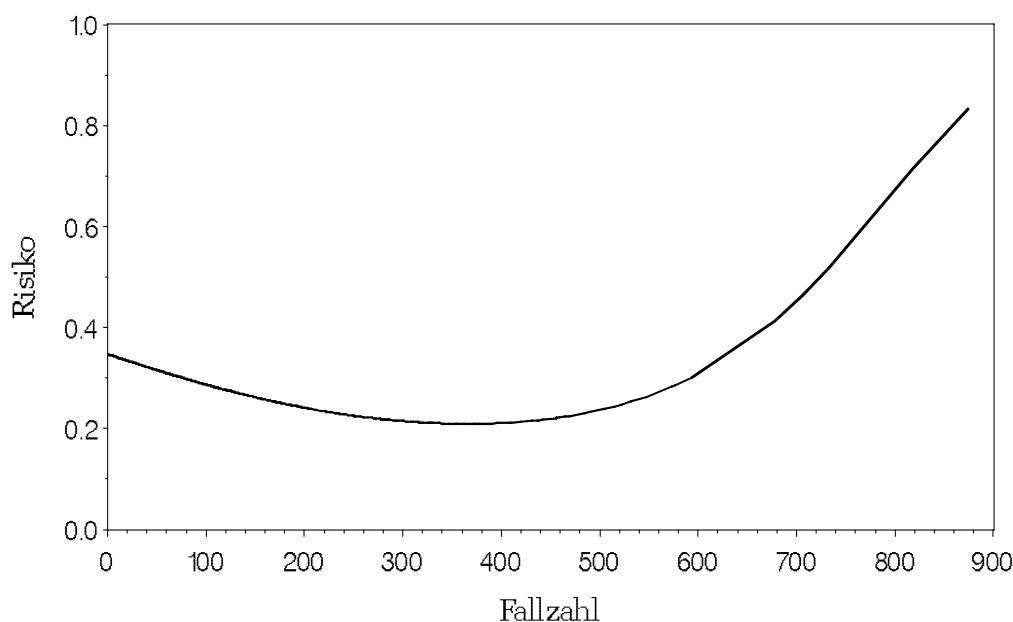


Abb. 4: Adjustierter Risikoverlauf für "Unbeweglichkeit" aus Modell II (2004)

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. Da vorab keine Hypothesen für Wechselwirkungen vorhanden waren, wurde diese niedrige Signifikanzgrenze gewählt, um nur sehr deutliche Wechselwirkungen im Modell zu berücksichtigen. Keine der betrachteten Wechselwirkungen war jedoch signifikant, so dass alle Interaktionsterme aus dem Modell entfernt wurden. Alle Haupteffekte wurden weiterhin betrachtet, auch wenn sie nicht zum Niveau $\alpha=0.05$ signifikant waren, da die entsprechenden Einflussgrößen in der Risiko-adjustierung berücksichtigt werden sollten.

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. 9	
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. 9	
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^a 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)

^abezogen auf Differenz der Intervallmittelpunkte

Abschließend wurde die Robustheit der Ergebnisse und insbesondere der nicht-monotone Zusammenhang zwischen Risiko und Fallzahl mit Hilfe von Sensitivitätsanalysen überprüft. Zunächst wurden regressionsdiagnostische Verfahren zur Identifikation extremer Werte eingesetzt, die stark vom angepassten Modell abweichen, und eine neue Modellanpassung ohne diese Extremwerte durchgeführt. Ferner wurden Analysen für Teilstichproben (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 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 5 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%).

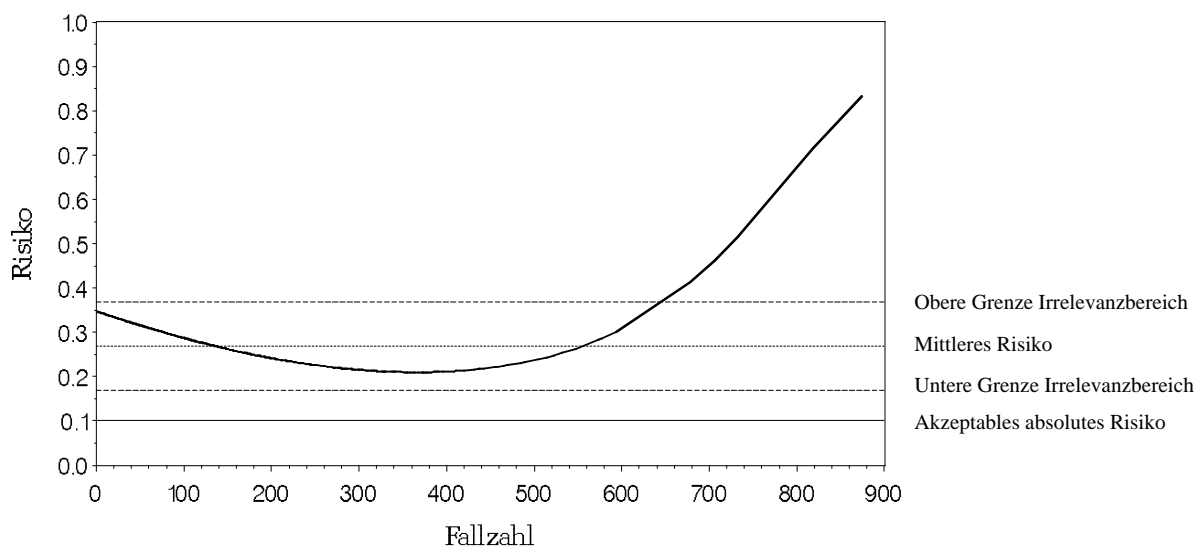


Abb. 5: Risikoverlauf für "Unbeweglichkeit" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos (2004)

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. 14	
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. 14	
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. 14	
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. 14	
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^a 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)
^a bezogen auf Differenz der Intervallmittelpunkte					

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 6 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.

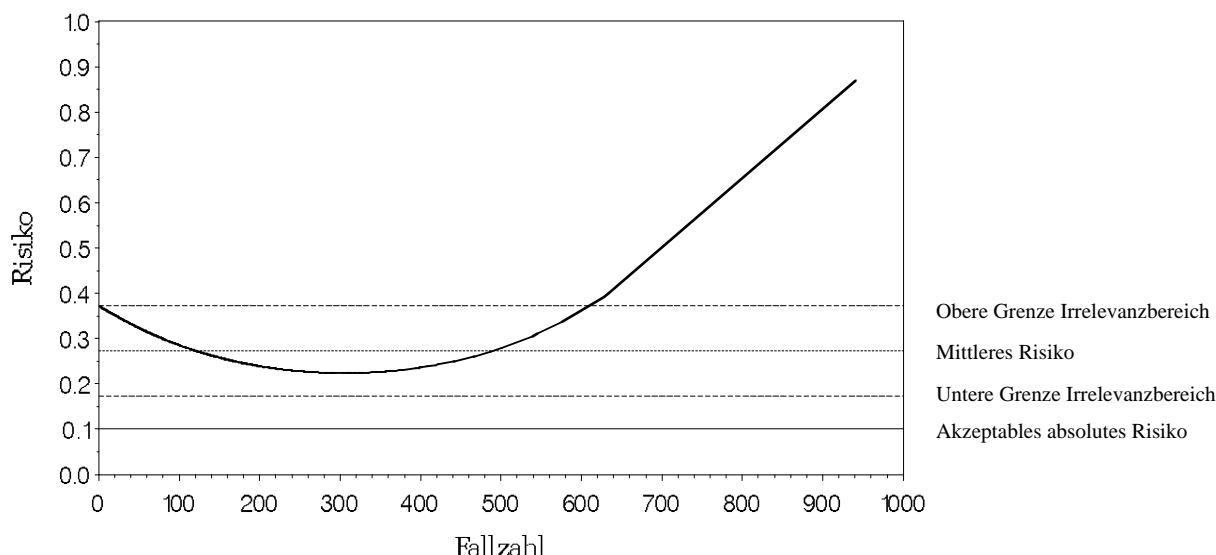


Abb. 6: Risikoverlauf für "Unbeweglichkeit" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos (2003)

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

Um den mit einer Mindestmengenregelung mit dem Wert MM=50 assoziierten Effekt zu beschreiben, wurden die in Abschnitt 4.3.4 beschriebenen Effektmaße eingesetzt. Es wurden sowohl Berechnungen basierend auf einfachen Vierfeldertafeln durchgeführt als auch basierend auf der multiplen logistischen Regression. Da die Unterschiede zwischen den Ergebnissen der rohen und adjustierten Schätzungen vernachlässigbar waren, werden nur die Ergebnisse basierend auf den Vierfeldertafeln dargestellt, da diese einfacher zu interpretieren sind.

Aus den Daten von 2004 errechnet sich ein geschätztes mittleres Risiko über alle Patienten von $\pi = 26.9\%$ und für die Patienten aus High-Volume-Krankenhäusern ein Risiko von $\pi_{HV} = 26.2\%$. Damit ergibt sich eine geschätzte Risikoreduktion von $\delta = 0.67\%$ (95%-KI: 0.59 bis 0.75%) oder anders ausgedrückt eine Population Impact Number von PIN=150 (95%-KI: 134 bis 170). Das bedeutet, dass von 150 Knie-TEP-Patienten ein Fall mit nicht ausreichender Beweglichkeit aufgrund der Behandlung in einem Low-Volume-Krankenhaus zu erwarten ist. Aus den Daten von 2003 erhält man $\pi = 27.2\%$, $\pi_{HV} = 26.4\%$, $\delta = 0.82\%$ (95%-KI: 0.71 bis 0.94%) und PIN=122 (95%-KI: 107 bis 142). Die mit der Mindestmenge MM=50 assoziierte Risikoreduktion von 0.67% für 2004 bzw. 0.82% für 2003 ist damit um ein Vielfaches kleiner als die geforderte Risikoreduktion von 10%, die vorab als klinisch relevant festgelegt wurde. Ein solch hoher Effekt konnte allerdings aufgrund der Verteilung der Fallzahlen auch gar nicht erreicht werden (siehe Abb. 1).

Für die einfache Risikodifferenz ergeben sich aus den Daten von 2004 die deutlich höheren Ergebnisse $\gamma = 11.01\%$ (95%-KI: 9.71 bis 12.32%) bzw. NNT=9 (95%-KI: 8 bis 10). Das bedeutet, dass unter 9 Knie-TEP-Patienten in Low-Volume-Krankenhäusern ein Fall mit nicht ausreichender Beweglichkeit mehr zu erwarten ist als in einem High-Volume-Krankenhaus. Dies ist zunächst als deutlicher Effekt einer Mindestmengenregelung mit der Grenze MM=50 zu werten. Allerdings darf hier nicht vergessen werden, dass sich der Effekt umkehrt, wenn man die Risiken der Low-Volume-Krankenhäuser und der Krankenhäuser mit sehr hohen Fallzahlen vergleicht. Unterteilt man die Krankenhäuser in drei

Gruppen durch Low: 0 bis unter 50, Medium: 50 bis 600 und High: über 600 Fälle, so ergibt sich beim Vergleich der so definierten Low- und High-Volume-Krankenhäuser ein Risikoanstieg (HV–LV) von 15.90% (95%-KI: 13.85 bis 17.93%) bzw. NNT=6 (95%-KI: 5 bis 7). Das bedeutet, dass unter 6 Knie-TEP-Patienten in High-Volume-Krankenhäusern ein Fall mit nicht ausreichender Beweglichkeit mehr zu erwarten ist, als in einem Low-Volume-Krankenhaus. Die Verlagerung eines Patienten von einem Low-Volume-Krankenhaus in ein High-Volume-Krankenhaus wäre danach mit einer sehr ausgeprägten Risikoerhöhung assoziiert. Diese Ergebnisse verdeutlichen noch einmal den stark U-förmigen Verlauf der Risikokurve, bei dem nur die Krankenhäuser mit mittleren Fallzahlen (etwa zwischen 100 und 500) ein vergleichsweise geringes Risiko aufweisen.

5.2.4 Schlussfolgerungen

Es ist nicht möglich, aus den vorhandenen Daten für den Qualitätsindikator "Unbeweglichkeit" einen klaren Schwellenwert abzuleiten, der im Rahmen einer entsprechenden Mindestmengenregelung für die Knie-TEP mit einer eindeutigen Verbesserung der Ergebnisqualität assoziiert ist. 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, bezüglich des populationsbezogenen Effektmaßes und den entsprechenden Vorgaben nicht klinisch relevant. Drittens lag die Risikokurve komplett oberhalb des akzeptablen absoluten Risikos, so dass keine Mindestmenge gefunden werden kann, die zu einer Unterschreitung des akzeptablen Risikos führt. Ausgehend von dem gefundenen Risikoverlauf mit dem starken Anstieg ab einer Fallzahl von etwa 500, erscheint hier eher die Kombination einer Mindestmenge mit einer "Maximalmenge" sinnvoll. Die Berechnung eines optimalen Bereichs in diesem Sinne war jedoch nicht Gegenstand des Auftrags und wurde daher nicht weiter verfolgt. Zudem wären weitere Festlegungen notwendig, welche Eigenschaften ein solch "optimaler" Fallzahlbereich erfüllen soll, um diesen mit Hilfe statistischer Methoden zu berechnen.

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 7 ist der Risikoplot in Abhängigkeit von der Fallzahl dargestellt. Die Punktwolke weist einen leicht monoton fallenden Trend auf.

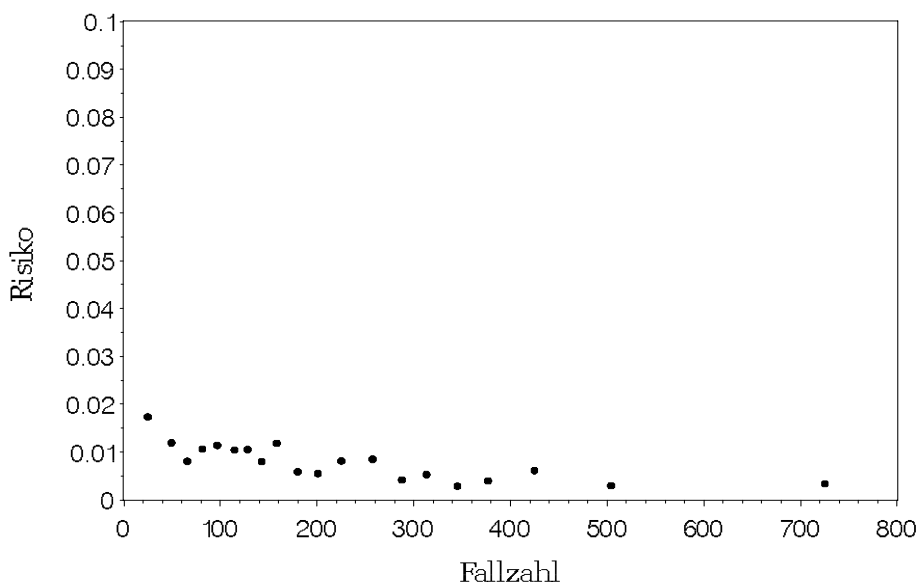


Abb. 7: Risikoplot für "Infektion" vs. Fallzahl (2004)

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 8 zeigt den aus dem Modell geschätzten Risikoverlauf mit 95%-Konfidenzband. 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.

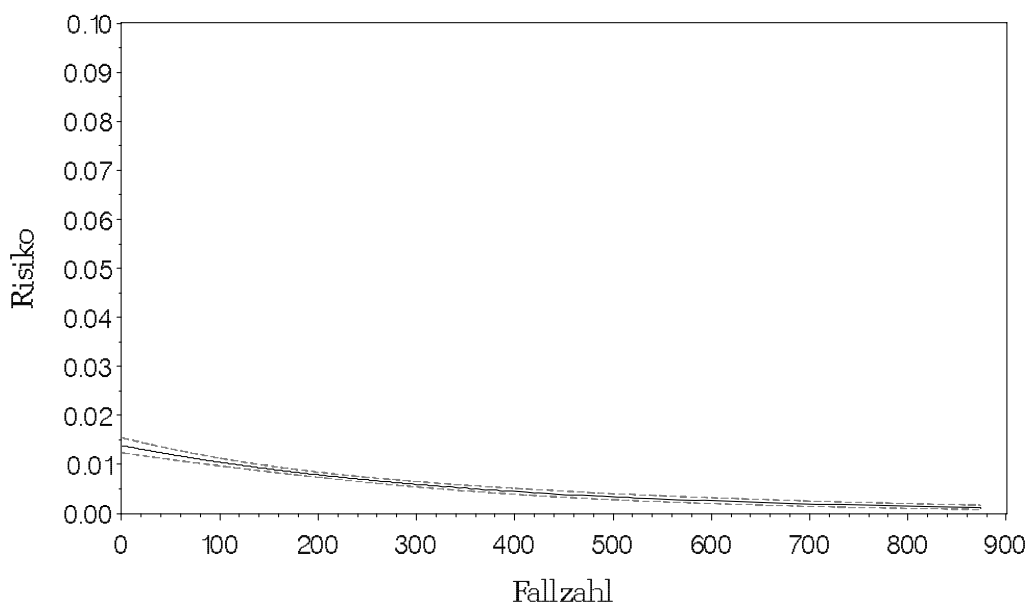


Abb. 8: Risikoverlauf mit 95%-Konfidenzbändern für "Infektion" aus Modell I (2004)

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 9 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						

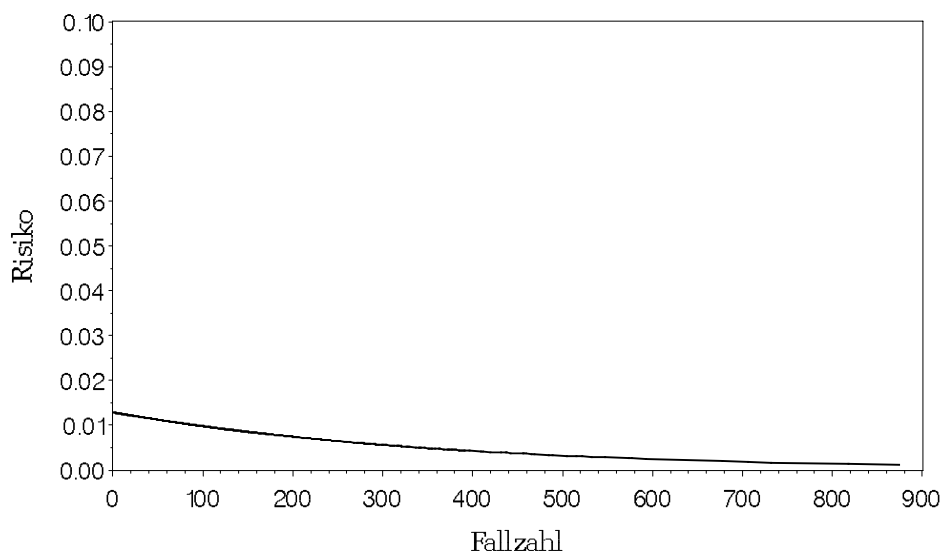


Abb. 9: Adjustierter Risikoverlauf für "Infektion" aus Modell II (2004)

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 10 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. 9). 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 8 und 10 stark gedehnt dargestellt, was bei der Interpretation beachtet werden muss. Der dargestellte Bereich umfasst nur 3% des gesamten Wertebereichs eines Risikos.

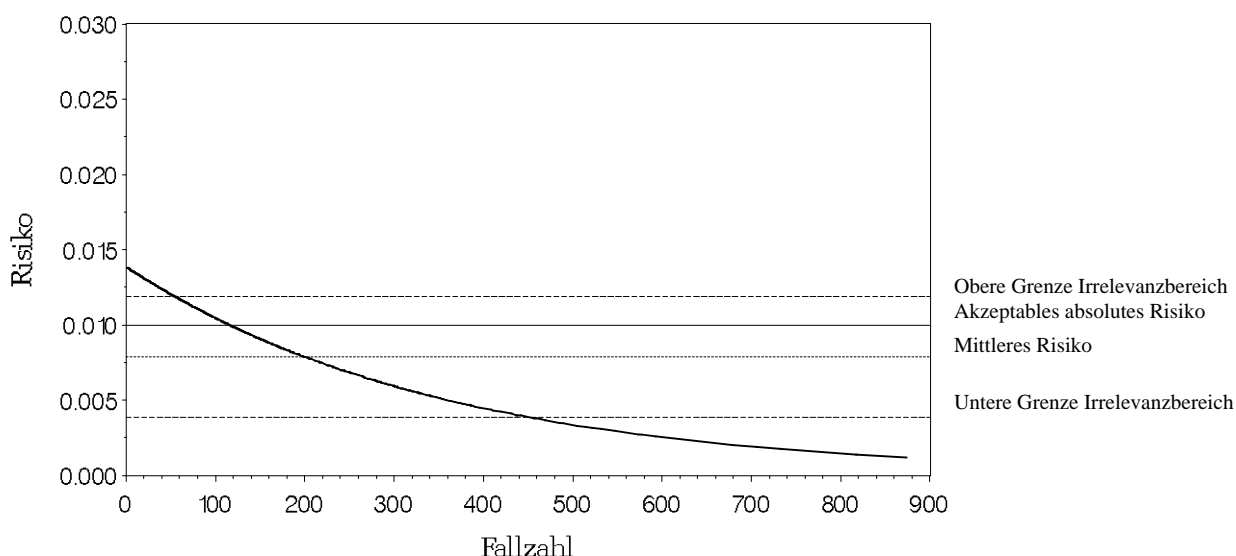


Abb. 10: Risikoverlauf für "Infektion" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos bei stark gedehnter Y-Achse (2004)

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 11 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.

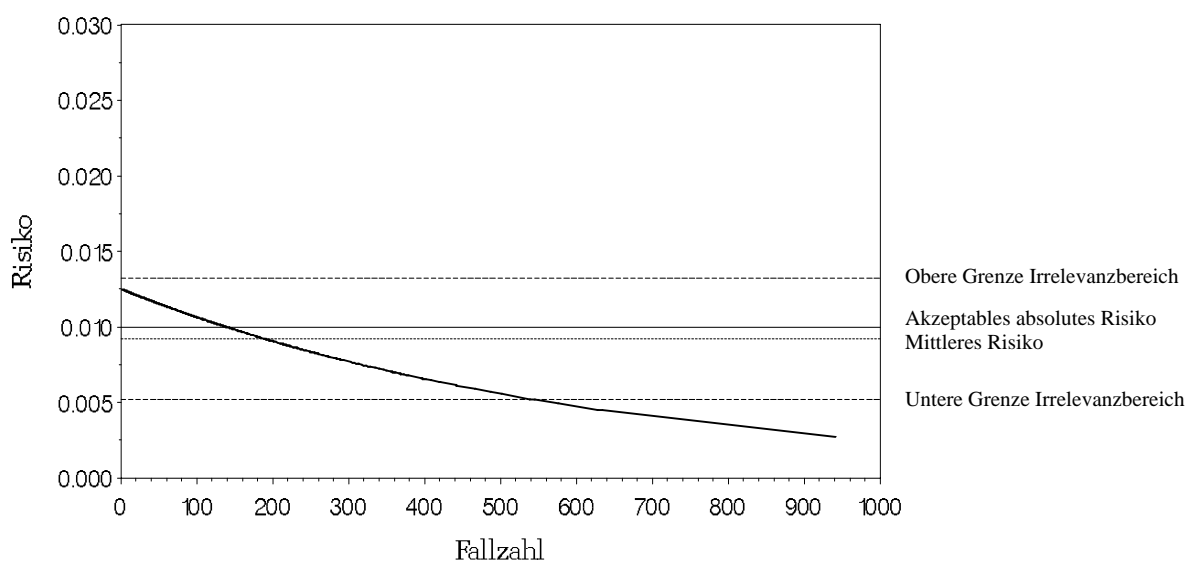


Abb. 11: Risikoverlauf für "Infektion" (Modell I) mit Kennzeichnung des mittleren Risikos, des Irrelevanzbereichs und des akzeptablen Risikos bei stark gedehnter Y-Achse (2003)

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 [16] 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 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 12 grafisch veranschaulicht, wobei die Y-Achse wieder stark gedehnt dargestellt ist.

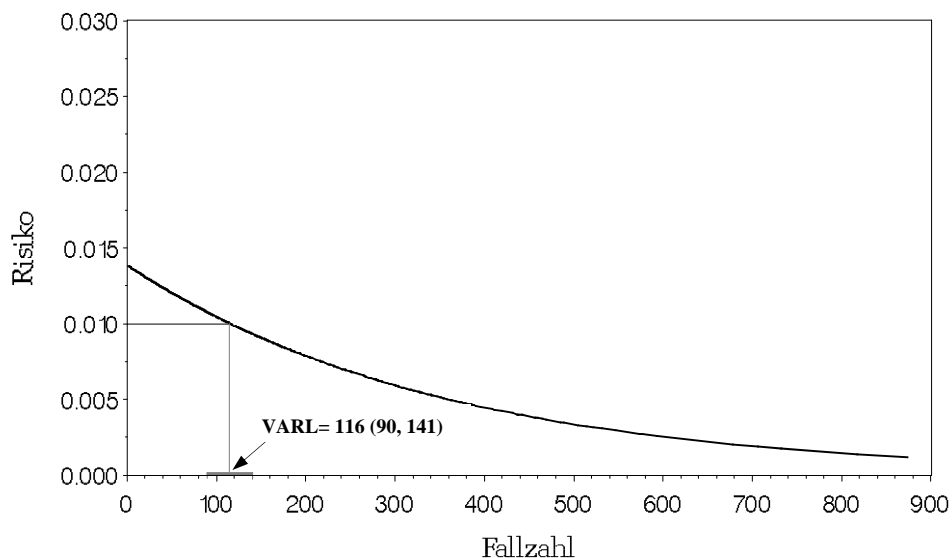


Abb. 12: Risikoverlauf für "Infektion" nach Modell I mit Kennzeichnung des berechneten VARL-Wertes mit 95%-KI bei stark gedehnter Y-Achse (2004)

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 errechnet sich eine geschätzte Risikoreduktion von $\delta = 0.15\%$ (95%-KI: 0.11 bis 0.19%) bzw. PIN=664 (95%-KI: 529 bis 872). Das bedeutet, dass unter 664 Knie-TEP-Patienten ein Fall von Wundinfektion aufgrund der Behandlung in einem Low-Volume-Krankenhaus zu erwarten ist. Aus den Daten des Jahres 2003 erhält man

für den Benchmark VARL=139 die Ergebnisse $\pi = 0.92\%$, $\pi_{HV} = 0.74\%$, $\delta = 0.18\%$ (95%-KI: 0.13 bis 0.23%) und PIN=556 (95%-KI: 426 bis 791).

Die mit diesen Benchmarks assoziierten Risikoreduktionen von 0.15% für 2004 bzw. 0.18% für 2003 sind damit deutlich kleiner als die geforderte Reduktion von 0.4%, die 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 mit populationsbezogener Sichtweise eine klinisch relevante Qualitätsverbesserung zu erwarten ist.

Für die einfache Risikodifferenz ergeben sich aus den Daten von 2004 für den Benchmark VARL=116 als Mindestmenge wieder höhere Werte: $\gamma = 0.53\%$ (95%-KI: 0.41 bis 0.67%) bzw. NNT=187 (95%-KI: 149 bis 246). Das bedeutet, dass unter 187 Knie-TEP-Patienten in Low-Volume-Krankenhäusern ein Fall von Wundinfektion mehr zu erwarten ist als in einem High-Volume-Krankenhaus.

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 errechnet 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: 0.05 bis 0.09%)** und PIN=1460 (95%-KI: 1098 bis 2031). Aus den Daten des Jahres 2003 erhält man $\pi = 0.92\%$, $\pi_{HV} = 0.87\%$, $\delta = 0.05\%$ (95%-KI: 0.02 bis 0.07%) und PIN=2165 (95%-KI: 1374 bis 4258).

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 des Qualitätsindikators Wundinfektion mit populationsbezogener Sichtweise eine klinisch relevante Qualitätsverbesserung abzuleiten ist.

Für die einfache Risikodifferenz ergeben sich aus den Daten von 2004 die Ergebnisse: $\gamma=0.92\%$ (95%-KI: 0.66 bis 1.22%) bzw. $NNT=109$ (95%-KI: 82 bis 152). Diese Resultate verdeutlichen die Nachteile des Effektmaßes γ . Für die Mindestmenge $MM=116$ ergibt sich mit $\gamma=0.53\%$ ein geringerer Effekt im Vergleich zur Mindestmenge $MM=50$ mit $\gamma=0.92\%$, obwohl für die Patienten bei Einführung der Mindestmenge $MM=116$ bezüglich des Wundinfektionsrisikos eine höhere Ergebnisqualität zu erwarten ist (geringeres Wundinfektionsrisiko).

Eine Mindestmenge, die populationsbezogen mit einer relevanten Qualitätsverbesserung bezüglich des Infektionsrisikos assoziiert ist, müsste den Wert **$MM=400$** oder mehr erreichen (Daten aus 2004). Für $MM=400$ werden die folgenden Effekte erreicht: **$\delta = 0.42\%$** , **$PIN=236$** , **$\gamma=0.49\%$** und **$NNT=223$** . 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.

5.3.5 Schlussfolgerungen

Für den sekundären Qualitätsindikator "Infektion" sind aus den Daten des Bundesdatenpools 2004 und 2003 der BQS Schwellenwerte ableitbar, deren Eignung im Rahmen einer Mindestmengenregelung unter Berücksichtigung anderer Indikatoren und des Versorgungskontextes beurteilt werden muss. Es konnte ein statistisch signifikanter negativer Zusammenhang zwischen der Fallzahl und dem Infektionsrisiko gezeigt werden. Allerdings ist die entsprechende Risikokurve sehr flach und der Erklärungswert der Fallzahl letztlich sehr gering, was dazu führt, dass nur sehr hohe Werte mit einer klinisch relevanten Qualitätsverbesserung assoziiert sind. Weder der nach dem VARL-Ansatz abgeleitete Wert von 116 Fällen, noch die bereits eingeführte Mindestmenge von 50 Fällen ist aus populationsbezogener Sicht mit einer klinisch bedeutsamen Reduktion des Risikos assoziiert. Höhere Mindestmengen, die möglicherweise für den Indikator "Infektion" sinnvoll sind, widersprechen allerdings den Ergebnissen zum Indikator "Unbeweglichkeit".

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 im Rahmen dieses Auftrags nicht überprüfbar waren. Eine solche Überprüfung der Krankenhausangaben wäre jedoch notwendig, um Analysen auf der Basis einer gesicherten Datenqualität durchzuführen. 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 von der BQS auf fehlenden Wert gesetzt. Die Stabilität der Berechnungen bezüglich der fehlenden Werte wurde durch Sensitivitätsuntersuchungen analysiert. Ein weiteres Problem stellt der unklare Messzeitpunkt des Indikators "Unbeweglichkeit" dar, der nicht unbedingt auf das Ende der Verweildauer fällt. Dadurch war es nicht möglich, den Qualitätsindikator "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, so dass hier selbst hypothesengenerierend kein eindeutiger Schwellenwert ableitbar war, der sich möglicherweise als Mindestmenge eignet. Zudem zeigte sich das Problem, dass Mindestmengen, die sich – unter gewissen Vorgaben – für den einen Qualitätsindikator möglicherweise eignen, für einen anderen Qualitätsindikator gerade nicht zu empfehlen sind. Dieses Ergebnis aus den verfügbaren deutschen Daten steht in Übereinstimmung mit den Resultaten aus internationalen Daten, dass sich bei der Knie-TEP kein klarer und einheitlicher Grenzwert bezüglich der Fallzahl pro Krankenhaus und Jahr für die Unterscheidung zwischen höherer und geringerer Qualität finden lässt [21,22].

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. Die vorliegende Auswertung unterstützt damit die Hypothese, dass es bei der Knie-TEP einen Zusammenhang zwischen der Leistungsmenge und der Ergebnisqualität gibt. Somit kann aus diesen Ergebnissen nicht abgeleitet werden, dass eine Mindestmengenregelung zu keiner Veränderung der Ergebnisqualität führt.

Der Zusammenhang zwischen dem primären Qualitätsindikator "Unbeweglichkeit" und der Fallzahl zeigte allerdings 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. Eine geeignete Maßnahme zur Verbesserung der Ergebnisqualität scheint hier eher die Definition eines mittleren Leistungsbereichs zu sein. Hierfür sind weitere Untersuchungen notwendig.

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 klaren eindeutigen Schwellenwert abzuleiten. Eine Assoziation mit einer deutlichen Qualitätsverbesserung ergab sich nur für Mindestmengen in höheren Fallzahlbereichen. Hier muss jedoch der Zusammenhang zwischen der Leistungsmenge und anderen Qualitätsindikatoren sowie der Versorgungskontext beachtet werden.

Ein wissenschaftlicher Nachweis, dass eine Mindestmengenregelung für Patienten mit Knie-Totalendoprothesen-Erstimplantation zu einer Verbesserung der Ergebnisqualität führt, kann nur über eine kontrollierte Interventionsstudie geführt werden.

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] Robert-Koch-Institut (Hrsg.). *Definition nosokomialer Infektionen (CDC-Definitionen), 4. Aufl.* Berlin: Nationales Referenzzentrum für Surveillance von nosokomialen Infektionen, 2004 (http://www.nrz-hygiene.de/dwnld/cdc_definitionen.pdf, Zugriff am 01.12.2005).
- [6] Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. *Ann Rheum Dis* 1957; 16: 494-502.
- [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] Kleinbaum DG, Klein M. *Logistic Regression: A Self-Learning Text, 2nd Ed.* New York: Springer, 2002.
- [11] 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.

- [12] 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.
- [13] Nagelkerke NJD. A note on the general definition of the coefficient of determination. *Biometrika* 1991; 78: 691-692.
- [14] Mittlböck M, Schemper M. Explained variation for logistic regression. *Stat Med* 1996; 15: 1987-1997.
- [15] Pregibon D. Logistic regression diagnostics. *Ann Stat* 1981; 9: 705-724.
- [16] Bender R. Quantitative risk assessment in epidemiological studies investigating threshold effects. *Biom J* 1999; 41: 305-319.
- [17] Heller RF, Dobson A. Disease impact number and population impact number: Population perspectives to measures of risk and benefit. *BMJ* 2000; 321: 950-952.
- [18] Newcombe RG. Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Stat Med* 1998; 17: 873-890.
- [19] Bender R. Calculating confidence intervals for the number needed to treat. *Control Clin Trials* 2001; 22: 102-110.
- [20] Cook RJ, Sackett DL. The number needed to treat: A clinically useful measure of treatment effect. *BMJ* 1995; 310: 452-454.
- [21] Klakow-Franck R, Wetzel H. Mindestmengenregelung ohne Mindestmengen. *Dtsch Ärzteztg* 2004; 101: A-2926-2928.
- [22] Schröder P, Rath T. Mindestmengen in der Kniegelenkendoprothetik. *Orthopäde* 2005; 34: 198-209.

Anhang: Stellungnahmen

A1. Stellungnahme der BQS

Formblatt Stellungnahme zum Vorbericht

Auftrag des IQWiG/Projektbezeichnung mit Auftragsnummer:
<p>IQWiG-Vorbericht b05-01a "Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für die Knie-Totalendoprothese" vom 17.10.2005</p>

Name; Vorname des/der Stellungnehmenden:
<p>Dr. med. Oliver Boy Dipl.-Inform. Med. Burkhard Fischer Dipl.-Stat. Joachim Kötting Dr. med. Volker D. Mohr</p>

Kontakt für weitere Nachfragen:
<p>Anschrift: BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH, Tersteegenstr. 12, 40474 Düsseldorf</p> <p>Tel./Fax-Nr.: 0211-280729-0</p> <p>E-Mail-Adresse: info@bqs-online.de</p>

**Stellungnahme der BQS Bundesgeschäftsstelle
Qualitätssicherung gGmbH zum IQWiG-Vorbericht b05-01a
"Entwicklung und Anwendung von Modellen zur Berechnung
von Schwellenwerten bei Mindestmengen für die Knie-
Totalendoprothese" vom 17.10.2005**

Generelle Stellungnahme

Die gefundenen Ergebnisse entsprechen denen, die mit den vorab festgelegten Methoden des Berichtsplans und den von den Fachexperten getroffenen Ergebnissen zu erzielen sind.

Die im Berichtsplan (IQWiG 2005) vorgesehene Berechnung nach den Ansätzen von Budtz-Jørgensen (2001) und ggf. Ulm (1991) fehlt. Hier sollte eine Begründung erfolgen oder die Ergebnisse sollten ergänzt werden.

Kritisch anzumerken ist, dass nur die Eignung der Datenbasis und der gewählten Qualitätsindikatoren diskutiert wird, nicht jedoch die Eignung der gewählten Methoden und der Festlegungen durch die Experten zu einer "klinisch relevanten Veränderung des absoluten Risikos" und zu einem "akzeptablen absoluten Risiko".

Vorschläge für zukünftige "Schwellenwertberechnungen für Mindestmengen"

Die Festlegungen durch die Fachexperten "klinisch relevante Veränderung des absoluten Risikos" und zu "akzeptables absolutes Risiko" haben sich als unrealistisch herausgestellt. Hier sollten für zukünftige Berechnungen andere Festlegungen durch die Fachexperten erreicht werden. Was kann hier geändert werden? Wie können die Fachexperten unterstützt werden?

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 Berichtes nicht oder nicht ausreichend beantwortet. Eine Klärung der Fragen ist wünschenswert.

Wo?	Was?	Kommentar
Seite 9	Haben die Schritte 4 bis 8 stattgefunden? In dieser Reihenfolge? 4.3.1 zeigt eine andere Reihenfolge Die Ergebnisse von 5 und 8 sind in 4.2 dargestellt	Vermutung: Die Festlegung der Experten bei "Nicht ausreichende Beweglichkeit" erfolgte ohne Kenntnis oder adäquate Berücksichtigung der Punktwolke von 2003. Bei Einhaltung der Reihenfolge wären die Vorgaben der Fachexperten möglicherweise realistischer ausfallen. Die Frage, ob dies allerdings zu anderen Ergebnissen geführt hätte, kann für die "nicht ausreichende Beweglichkeit" verneint werden.
Seite 9	Wo sind die Ergebnisse zu Schritt 2? In 4.3.1?	Zu 4.2 nehmen und dort mit Datum aufführen.

Wo?	Was?	Kommentar
Seite 12 4.3.2	"Zur Darstellung der Effektstärke wurden Odds-Ratio aus der LR ermittelt, und zwar jeweils für eine Änderung der Fallzahl um 50 Fälle."	Wo werden diese Ergebnisse dargestellt?
Seite 14 4.4 Ergänzungen im Vergleich zum Berichtsplan	Im Berichtsplan wurde eine Berechnung nach Budtz-Jørgensen und Ulm vorgesehen. Diese Berechnung wird nicht dargestellt. Wurden diese Modelle gerechnet und nicht dargestellt?	Ergebnisse darstellen oder Begründung ergänzen, warum nicht berechnet bzw. nicht in den Bericht aufgenommen.
Seite 20	Warum hier mit Einschlussgrenze $p=0,01$ und sonst mit Signifikanzniveau $0,05$ gerechnet?	Begründung ergänzen.
Seite 21	Warum bleiben die nicht signifikanten Variablen in den Modellen?	Begründung ergänzen.
Seite 38	Warum "U" unerwartet? Vor der Untersuchung steht doch kein Ergebnis fest.	"unerwartet" streichen?

Verbesserungshinweise

Neben den gestellten Fragen, die ggf. einer Erläuterung oder Klarstellung im Bericht erfordern, sind folgende Änderungen notwendig, damit die Sachverhalte korrekt wiedergegeben werden:

Darstellung in der Einleitung

Wo?	Was?	Vorschlag
4.2 (Seite 10) und weitere Fundstellen	"Unbeweglichkeit" Untersucht wurde die eingeschränkte Beweglichkeit von weniger als 90 Grad.	Alternativ sollte es "nicht ausreichende Beweglichkeit" heißen, da das Komplement der ausreichenden Beweglichkeit nicht die Unbeweglichkeit ist.
Seite 6 und Seite 7	Bundesgeschäftsstelle für Qualitätssicherung	BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH
Seite 7 2.1 Auftrag	", da diese die Datengrundlage zur Bearbeitung des Auftrags zur Verfügung stellt"	Diese so nicht zutreffende Begründung, die auch nicht im Auftrag zu finden ist, bitte streichen.

Darstellung der Methoden (Datenquellen, Vorgaben der Fachexperten)

Wo?	Was?	Vorschlag
Seite 10 4.1	"Die Fallzahl der behandelten Patienten bezieht sich hierbei jeweils auf ein Krankenhaus."	Ergänzen: Hierbei handelt es sich um die Zahl der für die externe stationäre Qualitätssicherung im Jahr 2003 bzw. 2004 dokumentierten Fälle."
Seite 10 4.2	Durch die Fachexperten festgelegten QI und Einflussfaktoren fehlen hier.	Von Datenbeschreibung in 4.3.1 trennen und unter 4.2 darstellen.

Darstellung zu den Auswertungen (Datenbeschreibung, Modellbildung, ...)

Wo?	Was?	Vorschlag
Seite 11 4.3.1	"Unbeweglichkeit"	"Nicht ausreichende Beweglichkeit": Erläuterung der Neutral-Null-Angabe 0/0/90 ergänzen: kein Streckdefizit, Beugefähigkeit von mindestens 90 Grad

Wo?	Was?	Vorschlag
	Grund für die fehlenden Werte bei Beweglichkeit angeben: Von 110.349 Datensätzen lagen nur bei 91.714 Datensätzen plausible Angaben zur postoperativen Beweglichkeit nach der Neutral-Null-Methode vor.	Fehlende Werte sind in der Regel darauf zurückzuführen, dass diese nicht erhoben (gemessen) wurden. Die Systematik der QS-Dokumentation schließt aber die Möglichkeit der systematischen Verfälschung durch Nichtdokumentieren schlechter Werte nicht aus.
Seite 11 4.3.1	Definition "implausible Werte" fehlt (siehe Erläuterung der Rechenregel)	Rechenregel für den QI (BQS 2005b): Grundgesamtheit: Alle Patienten (eingeschränkt auf plausible Werte bei Angaben zur Extension und Flexion) Zähler: Patienten mit postoperativer aktiver Beweglichkeit von mindestens 0/0/90 Grad (Extension/Flexion) Erläuterung der Rechenregel: Für die Beweglichkeit gelten als plausible Werte bei Angaben zur Extension und Flexion: 0 - 10/0 - 150/0 - 150 Grad.
Seite 11	BQS-Arthrose-Score	Bitte ergänzen: "ein modifizierter Röntgenscore der Arthrosezeichen im Kniegelenk nach Kellgren-und-Lawrence (Kellgren und Lawrence 1957)" (vgl. BQS 2005a)

Darstellung der Ergänzungen zum Berichtsplan

Wo?	Was?	Vorschlag
Seite 14 4.4	"aufgrund des unerwarteten Risikoverlaufs"	"unerwarteten" streichen und durch "U-förmigen"

Darstellung der Ergebnisse

Wo?	Was?	Vorschlag
Seite 17 Abb. 1	Bildung der Klasseneinteilung für die Fallzahlen der Krankenhäuser und wo diese auf der x-Achse dargestellt werden (Median, Mittelwert?)	Bitte ergänzen, wie die Vingtillgrenzen berechnet wurden.
Seite 17 Abb. 1, weitere Fundstellen	Achsen werden zwar auf Bildschirm angezeigt, aber teilweise nicht mit ausgedruckt!	Bitte so ändern, dass diese auch im Ausdruck erscheinen.
Seite 36	"Die Qualität der Daten ist unklar, da sie von den Krankenhäusern selbst berichtet werden und nicht überprüfbar sind." Hinweis: Daten sind anhand der Patientenakten überprüfbar, wurden jedoch im Rahmen der Untersuchung nicht überprüft, weil es nicht Bestandteil des Auftrags war.	" und nicht überprüfbar sind." ersetzen durch ". Eine Überprüfung war nicht Bestandteil des Auftrags."

Referenzen

[BQS 2005a] BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH: Qualitätsindikator "Indikation" (Knie-Totalendoprothesen-Erstimplantation, Verfahrensjahr 2004). Online verfügbar unter: http://www.bqs-qualitaetsindikatoren.de/2004/ergebnisse/leistungsbereiche/knie_tep_erst/indikatoren/indikation/rationale/document_view [Recherchedatum: 14.11.2005]

[BQS2005b] BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH: Qualitätsindikator "Postoperative Beweglichkeit" (Knie-Totalendoprothesen-Erstimplantation, Verfahrensjahr 2004). Online verfügbar unter: http://www.bqs-qualitaetsindikatoren.de/2004/ergebnisse/leistungsbereiche/knie_tep_erst/indikatoren/postoperative_beweglichkeit%20rationale/document_view [Recherchedatum: 14.11.2005]

[Budtz-Jørgensen et al. 2001] Budtz-Jørgensen E, Kleiding N, Grandjean P: Benchmark dose calculation from epidemiological data. *Biometrics* 2001; 57: 698-706

[IQWiG 2005] Berichtsplan zum Auftrag "Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für die Knie-Totalendoprothese" [Auftrag B05/01a] Version 3.0 vom 18.07.2005. Köln: Institut für Qualität und Wirtschaftlichkeit in der Medizin.

[Kellgren & Lawrence 1957] Kellgren JH, Lawrence JS: Radiological Assessment of Osteo-Arthrosis. *Annals of the Rheumatic Diseases*. 1957; 16: 494-502.

[Ulm 1991] Ulm K: A statistical method for assessing a threshold in epidemiological studies. *Stat Med* 1991; 10: 341-349

Notwendige Voraussetzungen zur Berücksichtigung von Stellungnahmen ist ihre formale Richtigkeit. Anhand folgender Checkliste können Sie prüfen, ob Ihre Stellungnahme die formalen Kriterien erfüllt. Bitte beachten Sie darüber hinaus die Hinweise im „Leitfaden für Stellungnehmende“.

- Umfang der Stellungnahme ist nicht größer als **6 DIN A4-Seiten**
- Für alle Zitate sind **Originalpublikationen (Volltext)** beigelegt
- Formblatt zur Stellungnahme vollständig ausgefüllt und **von allen Stellungnehmenden** unterschrieben
- Formblatt zur Darlegung potenzieller Interessenskonflikte** liegt **für alle Stellungnehmenden** einzeln **ausgefüllt** vor **siehe Anschreiben**

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.

Düsseldorf, 15.11.2005
Ort/Datum



Unterschrift



Vielen Dank für Ihre Stellungnahme!




A2. Stellungnahme von Dr. Schröder

Formblatt Stellungnahme zum Vorbericht

Auftrag des IQWiG/Projektbezeichnung mit Auftragsnummer:
Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für die Knie-Totalendoprothese Auftrag B05/01a

Name; Vorname des/der Stellungnehmenden:
Dr. Schröder, Peter

Kontakt für weitere Nachfragen:	
Anschrift:	MDS e. V., Lützowstraße 53, 45 141 Essen
Tel./Fax-Nr.:	0201 / 8327-243 (-244)
E-Mail-Adresse:	p.schraeder@mds-ev.de

1. Folgende Originalstudien fehlen im Vorbericht (Bitte beachten Sie die Ein- und Ausschlusskriterien für Studien dieses Berichtes):**Beweglichkeit**

Die graphische Darstellung des Zusammenhangs zwischen Menge und Qualität hinsichtlich des Indikators „Unbeweglichkeit“ zeigt einen u-förmigen Verlauf. Menge und Qualität sind somit bis zu einem bestimmten Punkt (Fallzahl) positiv miteinander korreliert, danach kehrt sich diese Beziehung um. Wenn hieraus eine Empfehlung hinsichtlich einer Mindestmengenausgestaltung erfolgen sollte würde dieser Verlauf allenfalls dafür sprechen, einen Minimal- und Maximalwert der Leistungserbringung für Knie-TEP festzulegen, z.B. im Sinne eines Korridors für die optimale Versorgung. Auf eine Berechnung gänzlich zu verzichten würde dem eher starken Zusammenhang zwischen Menge und Qualität (unabhängig von der Richtung) widersprechen. Eine Berechnung dieses Korridors ist auch dann noch sinnvoll, wenn – wie hier – die gesamte Kurve oberhalb des Irrelevanzbereichs liegt. Die Ergebnisse werden dadurch relativiert, der Beziehung Menge zu Qualität im Hinblick auf diesen Indikator wird jedoch Rechnung getragen. Bis zu einer Menge von ca. 350 K-TEP/Jahr ist der Zusammenhang streng abfallend, erst bei einer Leistungserbringung von oberhalb ca. 350 Knie-TEP/a ist die Menge mit einem schlechteren Ergebnis korreliert. Weniger als 5% der Krankenhäuser in Deutschland erbringen mehr als 350 K-TEP/a, d.h. für 95% der Krankenhäuser ist die Menge mit einem positiven Behandlungsergebnis hinsichtlich der postoperativen Beweglichkeit korreliert. Gleichzeitig ist die Dokumentationsquote der „High-Volume“-Krankenhäuser höher als die der „Low-Volume“-Krankenhäuser. Eine Darstellung der Verteilung der Leistungserbringung der Krankenhäuser sowie der Dokumentationsrate würde die gefundenen Ergebnisse sinnvoll ergänzen und sollte in die Diskussion aufgenommen werden.

Es ist zusätzlich wichtig darauf hinzuweisen, dass die "schwarz-weiß-Entscheidung" Beweglichkeit/Unbeweglichkeit aus Praktikabilitätsgründen gewählt wurde. Es gibt durchaus Graubereiche wie z. B. eine Beugefähigkeit von etwa 80 Grad und eine fehlende Streckfähigkeit von etwa 5 Grad, die klinisch zum selben Ergebnis führen wie eine volle Beugefähigkeit von 90 Grad und eine volle Streckfähigkeit von 0 Grad.

Wundinfektion

Für den Bereich der Wundinfektion kann ein Zusammenhang zwischen Menge und Qualität nachgewiesen werden. Nach der Methode VARL kann unter dem vorgegebenen Risiko für das Jahr 2004 eine Mindestmenge von 116 und für 2003 von 139 angegeben werden. Die Ergebnisse werden auf Seite 35 dahingehend relativiert, dass bei einer solchen Mindestmenge eine klinisch relevante Qualitätsverbesserung nicht zu erwarten sei. Unter Berücksichtigung des Irrelevanzbereichs ist ein Schwellenwert von 400 prinzipiell berechenbar. In der Schlussfolgerung wird jedoch lediglich beschrieben, dass ein sinnvoller Schwellenwert nicht berechnet werden kann und die Schwellenwerte 116 und 50 keine klinisch relevante Reduktion des Risikos bedeutet. Dies sollte ergänzt werden durch die Aussage, dass eine Mindestmenge von 400 berechenbar ist und diese zu einer relevanten Reduktion des Risikos führen würde. Eine Diskussion aller Berechnungen (Mindestmenge 116, 50 und 400) insbesondere unter Berücksichtigung des Versorgungskontextes würde die getroffenen Schlussfolgerungen sinnvoll ergänzen.

Irrelevanzbereich

Zur Festlegung der klinisch relevanten Veränderung des absoluten Risikos sowie des akzeptablen absoluten Risikos wurden den Fachexperten am 27.07.2005 per Mail die Ergebnisse zum mittleren Risiko einer postoperativen Wundinfektion bzw. Unbeweglichkeit bei Knie-TEP u. a. auch graphisch aufbereitet dargelegt. Da dies die

Basis der Einschätzung der Fachexperten bildete wäre es sinnvoll, dies auch genau so im Bericht darzustellen. In diesem Zusammenhang ist sicher bemerkenswert festzustellen, dass z. B. hinsichtlich der postoperativen Wundinfektion diese in den BQS-Daten sowohl bei der 10 und 25 % Quartile als auch im Median bei 0% lag. Die Festlegung des Irrelevanzbereichs erfolgte aufgrund klinischer Erfahrung. Nationale oder internationale Daten für die Knie-TEP im Hinblick auf relevante Risikoreduktionen existieren nicht. Deshalb erscheint perspektivisch ein formales Konsensusverfahren zur Ermittlung des Irrelevanzbereichs sinnvoll.

Risikoadjustierung

Auf Seite 11 des Vorberichts wird die Risikoadjustierung beschrieben. Generell sei darauf hingewiesen, dass diese Adjustierung nur anhand der vorhandenen BQS-Daten erfolgt, da nur auf diese zurückgegriffen werden konnte. Aus klinischer Sicht wird die postoperative Beweglichkeit u. a. in erheblichem Ausmaß durch den präoperativen Status beeinflusst. Ursprünglich war deshalb in der Diskussion, diesen über die präoperative Beweglichkeit in die Berechnung einfließen zu lassen. Da dieser Parameter in nicht ausreichendem Maße in den BQS-Daten dokumentiert war, wurde der präoperative Status versucht über die Kombination aus Arthrosegrad und präoperativer Deformität festzulegen. Generell gilt, dass gerade Beinachsen die besten postoperativen Ergebnisse zeigen, gefolgt von präoperativ bestehendem o-Bein und danach von präoperativ bestehendem x-Bein. Ähnlich zeigt sich das Verhältnis im Hinblick auf den Arthrosegrad. Generell lässt sich sagen, dass ein geringerer präoperativ bestehender Arthrosegrad bessere postoperative Ergebnisse im Hinblick auf die Beweglichkeit bedeutet. Dieser Tatsache sollte hinsichtlich der Risikoadjustierung insbesondere des Parameters Beweglichkeit auch in besonderem Maße Rechnung getragen werden. Vorgeschlagen wurde deshalb eine Kombination aus präoperativer Beweglichkeit und Arthrosegrad zu

wählen, so dass quasi drei Risikogruppen entstehen: Ein hohes Risiko für ein postoperativ schlechtes Ergebnis hinsichtlich der Beweglichkeit besteht bei präoperativem x-Bein verbunden mit stark ausgeprägter Arthrose, ein mittleres Risiko besteht bei präoperativ vorhandenem o-Bein mit mäßiger Arthrose und ein geringes Risiko besteht bei präoperativ gerader Beinachse und geringem Arthrosetrad. Es wäre deshalb hilfreich, sowohl auf die Entstehung der Risikoadjustierung hinzuweisen als auch darauf, warum ein solches Kombinationsmodell, wie oben skizziert, nicht verwendet wurde. In diesem Zusammenhang wäre es ebenfalls sinnvoll, die Rohdaten der BQS zu demonstrieren. Hierbei zeigt sich u. a., dass im Jahr 2004 über 1.000 Patienten (knapp 1% aller Knie-TEP-Patienten) mit einem modifizierten Arthrosetrad von 0 Punkten operativ versorgt wurden. Das bedeutet, dass bei diesen Patienten keine osteophytären Randaufbauten, keine Gelenkspaltverschmälerung, keine Sklerose und keine Deformierung der Femurkondylen präoperativ vorhanden waren. Eine fehlerhafte Dokumentation oder eine falsche Indikationsstellung sind hierbei interpretatorisch möglich.

Ausblick

Der relevante Paragraph zu den Mindestmengen spricht von einem Zusammenhang zwischen Menge und Qualität "in besonderem Maße". Wenn man dieses "in besonderem Maße" dahingehend interpretiert, dass die Menge mehr als andere Parameter (Ausbildung des Arztes, Ausbildung der Schwestern, Finanzierung des Krankenhauses etc.) einen Einfluss auf die Behandlungsqualität hat, dann müssten konsequenterweise jetzt nach diesem ersten Schritt auch noch alle weiteren Faktoren anhand von BQS-Daten untersucht werden, um dann die Ergebnisse in Beziehung zu setzen.

Berechnungen anhand anderer Datenquellen (z.B. Routinedaten) oder Kombinationen von Datenquellen sind ebenfalls denkbar.

Formblatt_Stellungnahme_151105

5

Abschliessend sollte auch in der Zusammenfassung – ähnlich wie in der Diskussion – darauf hingewiesen werden, dass letztendlich valide Daten nur über eine prospektive Erhebung einschliesslich aller Co-Faktoren zu generieren sind (wie für die Begleitevaluation vorgesehen).

Notwendige Voraussetzungen zur Berücksichtigung von Stellungnahmen ist ihre formale Richtigkeit. Anhand folgender Checkliste können Sie prüfen, ob Ihre Stellungnahme die formalen Kriterien erfüllt. Bitte beachten Sie darüber hinaus die Hinweise im „Leitfaden für Stellungnehmende“.

- | | |
|--|-------------------------------------|
| Umfang der Stellungnahme ist nicht größer als 6 DIN A4-Seiten | <input checked="" type="checkbox"/> |
| Für alle Zitate sind Originalpublikationen (Volltext) beigelegt | <input type="checkbox"/> |
| Formblatt zur Stellungnahme vollständig ausgefüllt und von allen Stellungnehmenden unterschrieben | <input checked="" type="checkbox"/> |
| Formblatt zur Darlegung potenzieller Interessenskonflikte liegt für alle Stellungnehmenden einzeln ausgefüllt vor | <input checked="" type="checkbox"/> |

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.

15.11.2005
Ort/Datum _____



Unterschrift

Vielen Dank für Ihre Stellungnahme!