



Institut für Qualitätssicherung und
Transparenz im Gesundheitswesen

Datenauswertung zu Mindestmengen in der Versorgung von Frühgeborenen mit einem Aufnahmegewicht unter 1.250 g

Abschlussbericht

Erstellt im Auftrag des
Gemeinsamen Bundesausschusses

Stand: 22. Juni 2020

Impressum

Thema:

Datenauswertung zu Mindestmengen in der Versorgung von Frühgeborenen mit einem Aufnahmegegewicht unter 1.250 g. Abschlussbericht

Ansprechpartnerinnen und Ansprechpartner:

Teresa Thomas, Janina Sternal, PD Dr. Günther Heller

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

20. Juni 2019

Datum der Abgabe:

22. Juni 2020

Herausgeber:

IQTIG – Institut für Qualitätssicherung
und Transparenz im Gesundheitswesen

Katharina-Heinroth-Ufer 1
10787 Berlin

Telefon: (030) 58 58 26-0
Telefax: (030) 58 58 26-999

info@iqtig.org

<https://www.iqtig.org>

Inhaltsverzeichnis

Tabellenverzeichnis.....	5
Abbildungsverzeichnis.....	6
Abkürzungsverzeichnis.....	7
Glossar mathematischer Notationen.....	8
1 Einleitung.....	10
1.1 Historie.....	10
1.2 Erkenntnisse der bestehenden Literatur	10
1.3 Auftrag und Auftragsverständnis	12
2 Beschreibung der Datenbasis.....	14
3 Endpunkte der Analysen	18
3.1 Mortalität	18
3.2 Morbidität	18
4 Volume-Outcome-Analyse	22
4.1 Inferenzmodell.....	22
4.1.1 Relevante Fallzahl.....	24
4.1.2 Modellierung stetiger Variablen	25
4.1.3 Patientenseitige Risikofaktoren	26
4.1.4 Konkretes Modell	27
4.1.5 Schätzung der Modellkomponenten.....	28
4.1.6 Was liefert das Modell	30
4.2 Hauptergebnisse	31
4.2.1 Der geschätzte Fallzahleffekt	31
4.2.2 Sonstige Kovariablen	34
4.2.3 Die PNZ-spezifischen random intercepts	37
4.3 Analyse basierend auf den Fallzahlen analog zur Mindestmengenregelung	37
4.4 Sensitivitätsanalysen.....	39
4.4.1 Analyse basierend auf den QS-Daten.....	39
4.4.2 Weglassen des größten PNZ.....	40
4.4.3 Betrachtungen ohne Einschränkung auf Level-1-PNZ.....	41
5 Suche nach möglichen Schwellenwerten mit einem Bruchpunkt-Modell	45

6	Analyse möglicher Schwellenwerte: Simulation der Einführung einer Mindestmenge und Berechnung absoluter Effektmaße	47
6.1	Limitationen der Betrachtungen	48
6.2	Betroffene PNZ und umverteilte Fälle	48
6.3	Betrachtete Maßzahlen zur Auswirkungen auf die Sterblichkeit	50
6.4	Modellierung der Umverteilungen	50
6.5	Auswirkung der Umverteilungen auf die Sterblichkeit	52
7	Diskussion	58
8	Literatur	61

Tabellenverzeichnis

Tabelle 1: Statistische Kennzahlen des Modellfits.....	31
Tabelle 2: Odds-Ratios für verschiedene Fallzahlpaare.....	33
Tabelle 3: Regressionskoeffizient des linearisierten Volume-Outcome-Zusammenhangs sowie entsprechende Odds Ratio für Fallzahländerung.....	34
Tabelle 4: Parameter für nicht-stetige Kovariablen.....	34
Tabelle 5: Odds-Ratios für nicht stetige Kovariablen.....	35
Tabelle 6: Statistische Kennzahlen des Modellfits basierend auf den Fallzahlen analog zur bestehenden Mindestmengenregelung.....	38
Tabelle 7: Statistische Kennzahlen des Modellfits. Datenbasis: QS-Daten, nur Level-1-PNZ.....	39
Tabelle 8: Statistische Kennzahlen des Modellfits. Datenbasis: NICU-Daten, nur Level-1-PNZ, weglassen des größten PNZ.....	41
Tabelle 9: Statistische Kennzahlen des Modellfits. Datengrundlage: NICU-Daten, beide Versorgungsstufen.....	42
Tabelle 10: Statistische Kennzahlen des Modellfits. Datengrundlage: QS-Daten, alle Versorgungsstufen.....	43
Tabelle 11: Geschätzte Werte der Maßzahlen im Szenario 1.....	55
Tabelle 12: Geschätzte Werte der Maßzahlen im Szenario 2.....	56

Abbildungsverzeichnis

Abbildung 1: Histogramm der Verteilung der durchschnittlichen Gesamtfallzahl der PNZ in den Jahren 2014 bis 2018.	16
Abbildung 2: Lorenzkurve der Verteilung der Fälle auf die PNZ für die Jahre 2014-2018.....	17
Abbildung 3: Streudiagramm zur Darstellung des Zusammenhangs zwischen der Patienten- und Fallebene beim Indikator „30-Tages-Sterblichkeit“	18
Abbildung 4: Vergleich der Ergebnisse „ROP mit Intervention“ zwischen Patienten- und Fallebene aus dem Verlegungsgeschehen.....	20
Abbildung 5: Geschätzter Fallzahleffekt f_{FZ} auf logistischer Ebene	32
Abbildung 6: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl.	33
Abbildung 7: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit vom Aufnahmegewicht.....	36
Abbildung 8: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Gewichtsabweichung.....	36
Abbildung 9: Geschätzte PNZ-spezifische random intercepts als Funktion von der Fallzahl.	37
Abbildung 10: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl analog zur Mindestmengenregelung.....	38
Abbildung 11: Die geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl. Datenbasis: QS-Daten, nur Level-1-PNZ.	39
Abbildung 12: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl. Datenbasis: NICU-Daten, nur Level-1-PNZ, weglassen des größten PNZ.	40
Abbildung 13: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl. Datengrundlage: NICU-Daten, beide Versorgungsstufen.....	42
Abbildung 14: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl. Datengrundlage: QS-Daten, alle Versorgungsstufen.....	43
Abbildung 15: Marginale log-Likelihood als Funktion des Bruchpunkts M . Die horizontalen gepunkteten Linien markieren die Schwellenwerte an die log-Likelihood mit Schwellenwerten für die Likelihood-Quotienten-Konfidenzintervalle.	46
Abbildung 16: Die Anzahl von betroffenen PNZ als Funktion der möglichen Mindestmenge M	48
Abbildung 17: Anzahl von umverteilten Kindern als Funktion des Schwellenwerts M	49
Abbildung 18: Geschätzte Anzahl an Sterbefällen, die durch die neue Mindestmenge M verhindert werden könnte, als Funktion von M	53
Abbildung 19: Geschätzte Risikoreduktion $\delta u(M)$ für die umverteilten Fälle als Funktion von M	53
Abbildung 20: Geschätzte durchschnittliche Risikoreduktion $\delta d(M)$ als Funktion von M	54

Abkürzungsverzeichnis

Abkürzung	Bedeutung
BPD	bronchopulmonale Dysplasie
esQS	externe stationäre Qualitätssicherung
G-BA	Gemeinsamer Bundesausschuss
IK-Nummer	Institutionskennzeichen des Krankenhauses
IN	impact number
InEK	Institut für das Entgeltsystem im Krankenhaus
IQTIG	Institut für Qualitätssicherung und Transparenz im Gesundheitswesen
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IVH	Intraventrikuläre Hämorrhagie
NEK	Nekrotisierende Enterokolitis
NICU	Neonatal Intensive Care Unit
NNT	number needed to treat
ML	Maximum Likelihood
Mm-R	Mindestmengenregelungen
MOR	Median Odds-Ratio
PNZ	Perinatalzentrum
QFR-RL	Qualitätssicherungs-Richtlinie Früh- und Reifegeborene
QS	Qualitätssicherung
ROP	Retinopathie
SE	Standard Error (Standardfehler)
SMR	Standardisierte Mortalitäts-Ratio
SSW	Schwangerschaftswoche
VLBW	Very Low Birth Weight (Frühgeborene unter 1.500 g)
VLBW-I	Very Low Birth Weight (Frühgeborene unter 1.250 g)

Glossar mathematischer Notationen

Notation	Erläuterung
Zähl-Indizes und Gesamtfallzahl	
i	Index für die PNZ
$j \in \{1, 2, \dots, n_i^k\}$	Index für die Frühgeborenen bei PNZ i
$k \in \{2014, \dots, 2018\}$	Index für das Erfassungsjahr
n_i^k	Gesamtfallzahl von PNZ i in Erfassungsjahr k .
Variablen des Inferenzmodells (vgl. Gleichung (2))	
π_{ij}^k	Sterbewahrscheinlichkeit für Fall j in PNZ i im Erfassungsjahr k
y_{ij}^k	binäres Outcome für das Frühgeborene j bei PNZ i im Erfassungsjahr k
\tilde{n}_i^k	kumulierte durchschnittliche Fallzahl von PNZ i zu Erfassungsjahr k
$\beta_0^k \in \mathbb{R}$	jahresspezifischer Effekt
\mathbf{x}_{ij}^k	Vektor der patientenseitigen diskreten Risikofaktoren
$\boldsymbol{\beta}$	Koeffizientenvektor der patientenseitigen diskreten Risikofaktoren
A_{ij}^k	Aufnahmegegewicht
D_{ij}^k	Gewichtsabweichung
$f_{\text{Gew}}: \mathbb{R} \rightarrow \mathbb{R}$	Effekt des Aufnahmegegewichts
$f_{\text{Diff}}: \mathbb{R} \rightarrow \mathbb{R}$	Effekt der Gewichtsabweichung
$f_{\text{FZ}}: \mathbb{R} \rightarrow \mathbb{R}$	Effekt der relevanten Fallzahl
$u_i \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(0, \tau^2)$	standortspezifischer random intercept
Bestimmung und Analyse von Schwellenwerten	
M	Bruchpunkt
Allgemeine mathematische Notation	
$\text{logit}(x)$	Logit-Transformation von x : $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$ für $x \in (0, 1)$
$\text{logit}^{-1}(y)$	inverse Logit-Transformation:

Notation	Erläuterung
	$\text{logit}^{-1}(y) = \frac{\exp(y)}{1 + \exp(y)}$ für $y \in (-\infty, \infty)$
$\text{Odds}(y = 1 z)$	Verhältnis von Wahrscheinlichkeit und Gegenwahrscheinlichkeit für das Ereignis $y = 1$ gegeben einen Wert z .
$\text{OR}(y = 1 z, z')$	Verhältnis der Odds (Odds-Ratio) für das Ereignis $y = 1$ gegeben einen Wert z zu den Odds für das Ereignis $y = 1$ gegeben einen Wert z'
$\text{MOR}(\tau)$	Median-Odds-Ratio: (vgl. die Definition in Abschnitt 4.1.6)
$h \equiv g$	Gleichheit der Funktionen h und g an allen Punkten des Definitionsbereiches
$h \not\equiv g$	Funktionen h und g sind nicht an allen Punkten des Definitionsbereiches gleich.
\hat{x}	Datenbasierte Schätzung von Variable x
$\text{Ber}(\pi)$	Bernoulli-Verteilung mit Erfolgswahrscheinlichkeit $\pi \in [0, 1]$
$\mathcal{N}(\mu, \sigma^2)$	Normalverteilung mit Erwartungswert μ und Varianz σ^2
$\Phi(z)$	Verteilungsfunktion der Standard-Normalverteilung

1 Einleitung

1.1 Historie

Die Versorgung von sehr kleinen Frühgeborenen (VLBW) in Zentren mit adäquater Ausstattung und Fallzahl wird seit Jahrzehnten in der wissenschaftlichen Literatur thematisiert (bspw. Heller 2009, Hentschel et al. 2019, Kutschmann et al. 2012, Rochow et al. 2016, Phibbs et al. 2007, Jochum et al. 2008, Heller et al. 2007). Vor diesem Hintergrund werden bereits seit dem Jahr 2004 im Gemeinsamen Bundesausschuss (G-BA) Beratungen über die Einführung einer Mindestmenge gemäß § 137 SGB V für diese Kinder durchgeführt. Da die Einschätzung der Literatur zunächst wenig einheitlich war, wurde das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) in diesem Rahmen mit einer Auswertung der Literatur zum Zusammenhang zwischen Leistungsmenge und Ergebnis bei der Versorgung von Früh- und Neugeborenen mit sehr geringem Geburtsgewicht beauftragt und legte seinen Abschlussbericht 2008 vor (IQWiG 2008). Im August 2009 wurde über die „Vereinbarung des Gemeinsamen Bundesausschusses über Maßnahmen zur Qualitätssicherung der Versorgung von Früh- und Neugeborenen“ unter anderem eine jährliche Mindestmenge von 14 Frühgeborenen mit einem Geburtsgewicht < 1.250 g für Perinatalzentren (PNZ) festgelegt (G-BA 2009) und im Juni 2010 auf 30 Frühgeborene erhöht, anschließend jedoch vor Gericht angefochten und bis zur Entscheidung vor dem Bundessozialgericht vom G-BA ausgesetzt. Das Bundessozialgericht sah in Entscheidungen im Dezember 2012 und im November 2015 die Vereinbarung von Mindestmengen durch den G-BA als grundsätzlich rechters, die Erhöhung der Mindestmenge auf 30 aber als nicht ausreichend begründet an.¹ Um eine konkrete Mindestmenge anhand eines möglichen Schwellenwertes besser begründen zu können, beauftragte der G-BA am 20. Juni 2019 das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) mit der Auswertung von vorhandenen Daten zur Versorgung von Frühgeborenen mit einem Aufnahmegewicht von unter 1.250 g (VLBW-I) (G-BA 2019b).

1.2 Erkenntnisse der bestehenden Literatur

Erste Analysen zum Zusammenhang zwischen Ausstattung bzw. Fallzahl von perinatalen Einheiten und dem Behandlungsergebnis von VLBW finden sich bereits in den 1980er Jahren (Paneth et al. 1987, Gortmaker et al. 1985). Dabei wird mit der Einführung einer Mindestmenge die Erwartung verbunden, durch eine Konzentrierung der Zentren eine stärkere Spezialisierung und eine Steigerung der Erfahrungen seitens des Personals bei der Versorgung von VLBW zu erreichen (Obladen 2007) und in der Folge so eine erhöhte Ergebnisqualität für diese Patientengruppe zu erzielen. Dabei wird die Ergebnisqualität in der Regel in Form von risikoadjustierten Mortalitätsraten, erheblich seltener auch über risikoadjustierte Morbiditätsraten, operationalisiert.

¹ Bundessozialgericht Az.: B1 KR34/12R, Az.: B1 KR15/15R.

Dabei kann konstatiert werden, dass die Fallzahlen pro PNZ in Deutschland im internationalen Vergleich vergleichsweise gering ausfallen (Rossi 2015). Wird nun konkreter die deutsche Versorgungssituation der Perinatalzentren gemessen an ihren Fallzahlen betrachtet, so zeigt sich, dass bei Einführung einer Mindestmenge 2009 bis 2010 nur 3 von 4 Zentren mindestens 30 VLBW-I pro Jahr behandelten (Kutschmann et al. 2012). 2013 haben nur ein Drittel aller Krankenhäuser mehr als 48 VLBWs behandelt (Miedaner et al. 2018). Für die Jahre 2014 bis 2018 zeigt sich auf der Website perinatalzentren.org, dass 27 Level-1-PNZ weniger als 30 VLBWs im Jahr versorgen. Gemessen an den bisher geplanten Mindestmengen seitens des G-BA kann also der Rückschluss gezogen werden, dass in Deutschland die Versorgung von Frühgeborenen stark dezentralisiert stattfindet.

Mittlerweile liegen verschiedene veröffentlichte Arbeiten über Fallzahleffekte auf die Ergebnisqualität von VLBW in Deutschland vor. Mehrheitlich ergab sich ein inverser Zusammenhang zwischen der Anzahl behandelter VLBW (bzw. VLBW-I) pro Jahr und der Mortalitätsrate. Dabei wird die Effizienz einer Umverteilung über eine Mindestmenge gegenüber einer möglichen Umverteilung über vorherige Ergebnisqualität, den Analysen von Rogowski et al. (2004) folgend, aber mitunter kritisch diskutiert (Rogowski et al. 2004, Kutschmann et al. 2012, Rochow et al. 2016, Heller 2018). Zudem muss an dieser Stelle aufgeführt werden, dass in einigen Studien ein gegenläufiger Zusammenhang beobachtet wurde (Shah et al. 2015).

Zahlreiche Arbeiten haben sich mit der Thematik befasst, welche Ursachen für eine Volume-Outcome-Beziehung kausal wirken können. Neben dem Grad an klinischer Routine und der Erfahrung des medizinischen Personals werden unter anderem bessere bauliche oder apparative Ausstattung größerer Einheiten diskutiert (Hentschel et al. 2019, Kutschmann et al. 2012). Auch mütterliche Faktoren, wie Alter, Status, (Vor-)Erkrankungen oder Rauchen können einen Einfluss auf das untersuchte Outcome ausüben (Shah et al. 2015, Bartels et al. 2006). Eine Analyse möglicher Gründe eines Volume-Outcome-Zusammenhanges ist jedoch nicht Bestandteil dieses Auftrages.

Insgesamt ist zu bemerken, dass bislang nur wenige Arbeiten vorliegen, die Volume-Outcome-Zusammenhänge nicht ausschließlich auf der Basis von kategorialen Fallzahlgruppen untersucht haben, sondern Analysen durchführen, mit denen eine Identifikation optimaler Schwellenwerte möglich ist (Rogowski et al. 2004, Grouven et al. 2008, Schröder et al. 2007, Rochow et al. 2016).

Alle hier berichteten Erkenntnisse aus der Literatur stammen aus Studien, welche teilweise mit sehr alten Datenjahren arbeiten. In den vergangenen Jahren hat jedoch die Gesamtmortalität abgenommen und große und kleine Perinatalzentren haben sich in der Ergebnisqualität weiter angeglichen (Hentschel et al. 2019, Heller 2018). Zudem muss berücksichtigt werden, dass einerseits viele Studien auf einer (meist regional begrenzten) Stichprobe von Perinatalzentren basieren. Durch die geringe Anzahl der verwendeten Fälle kann eine ausreichende Power der Studie nicht gewährleistet werden. Andererseits verwenden viele Studien ausländische Daten (bspw. Shah et al. 2015, Phibbs et al. 2007), die zwar eine hohe Fallzahl aufweisen, aber nicht ohne weiteres auf deutsche Gegebenheiten übertragen werden können. In Deutschland sind außerdem regionale Unterschiede bekannt, beispielsweise ein deutliches Nord-Süd-Gefälle bei der Sterblichkeit sehr kleiner Frühgeborener (Obladen 2007). Daher ist es von hoher Bedeutung,

dass für die Bestimmung einer Mindestmenge alle zur Verfügung stehenden Daten aus Deutschland von allen Perinatalzentren und zudem mehrere Jahrgänge zusammen betrachtet werden.

1.3 Auftrag und Auftragsverständnis

Am 20. Juni 2019 wurde das IQTIG beauftragt, die Versorgung von Frühgeborenen mit einem Aufnahmegegewicht unter 1.250 g auszuwerten. Hauptgegenstand der Beauftragung ist es, den Zusammenhang von Fallzahl und Qualität und ggf. geeignete Schwellenwerte zu untersuchen (G-BA 2019b). Als führende Qualitätsmerkmale wurden hierzu die Morbidität und Mortalität der Frühgeborenen ausgewählt. Das IQTIG hat den Auftrag, zu prüfen, ob sich die Morbidität als weiterer Aspekt neben der Mortalität als Endpunkt eignet.

Für die Auswertungen stehen zwei Datengrundlagen zur Verfügung. Zum einen werden die Daten der externen stationären Qualitätssicherung (esQS) aus dem QS-Verfahren *Neonatalogie* verwendet. Zum anderen stehen die Daten der verpflichtenden zentralen Ergebnisveröffentlichung nach der Qualitätssicherungs-Richtlinie Früh- und Reifgeborene (QFR-RL)² zur Verfügung. Hierbei werden die Daten der letzten Ergebnisveröffentlichung, mit den Datenjahren 2014 bis 2018, deren Veröffentlichung am 01. Dezember 2019 stattfand, verwendet.

Als primärer Endpunkt für die Analysen wird die risikoadjustierte Krankenhausmortalität dieser Subgruppe untersucht. Sekundär soll laut Beschluss die risikoadjustierte Morbidität betrachtet werden, insbesondere sollen die nekrotisierende Enterokolitis (NEK), intraventrikuläre Hämorrhagie (IVH), Retinopathie (ROP), bronchopulmonale Dysplasie (BPD), Pneumonie, Infektionen einschließlich Sepsis in Betracht bezogen werden.

Es wurden keine regionalen Analysen vorgenommen und es wurde nicht untersucht, wie sich die Erreichbarkeit der Krankenhäuser ändert, wenn Perinatalzentren, die potenziell unter der Mindestmenge liegen, geschlossen werden. Zudem wurden keine Folgenabschätzungen zu Fahrtzeiten und Transportrisiken durchgeführt sowie keine Effekte von sekundärer Umverteilung untersucht. Mit sekundärer Umverteilung ist der Sachverhalt umschrieben, dass durch die Einführung oder Erhöhung einer Mindestmenge Kinder, die bislang in PNZ unterhalb der Mindestmenge versorgt worden wären, umverteilt werden müssen. Sekundär könn(t)en im Rahmen von Übergangsfristen PNZ, die knapp unterhalb der Mindestmenge liegen, diese nach diesen Umverteilungen wieder erreichen. Darüber hinaus werden keine Analysen für weitere Gewichtsgrenzen als für Frühgeborene mit einem Aufnahmegegewicht von unter 1.250 g durchgeführt. Anders als bei der aktuellen Mindestmengenregelung (Mm-R) ist nun nicht mehr das Geburtsgewicht der Bezug, sondern das Aufnahmegegewicht.

Der vorliegende Bericht ist nach dem Bericht zu „Volumen-Outcome-Beziehungen bei Revisionseingriffen in der Knieendoprothetik“ (IQTIG 2019e, Stand: 20. Dezember 2019) der zweite des IQTIG, der den Zusammenhang von Fallzahl und Outcome untersucht. Das IQTIG hat dabei eine Methodik für Volume-Outcome-Beziehungen entwickelt, die auch in dem vorliegenden Bericht

² Richtlinie des Gemeinsamen Bundesausschusses über Maßnahmen zur Qualitätssicherung der Versorgung von Früh- und Reifgeborenen gemäß § 136 Absatz 1 Nummer 2 SGB V in Verbindung mit § 92 Abs. 1 Satz 2 Nr. 13 SGB V. In der Fassung vom 20. September 2005, zuletzt geändert am 14. Mai 2020, in Kraft getreten am 14. Mai 2020. URL: <https://www.g-ba.de/informationen/richtlinien/41/> (abgerufen am 18.06.2020).

mit gewissen Anpassungen angewandt wird. Die Methode der Schwellenwertsuche wird vom IQTIG in der vorliegenden Analyse erstmals angewandt.

2 Beschreibung der Datenbasis

Für die Analysen wurden zwei verschiedenen Datenquellen genutzt. Zum einen stehen die Daten des QS-Verfahrens *Perinatalmedizin* bezogen auf den Leistungsbereich Neonatologie zur Verfügung (im Folgenden: QS-Daten). Zum anderen werden Daten genutzt, die im Rahmen der verpflichtenden Ergebnisveröffentlichung im Rahmen der QFR-RL (im Folgenden: NICU-Daten) dem IQTIG zur Verfügung stehen. Die NICU-Daten werden für die Veröffentlichung auf der Website perinatalzentren.org verwendet.

Bei beiden Datensätzen handelt es sich um auf Basis einer vorab festgelegten Spezifikation von den Krankenhausstandorten selbstdokumentierte Daten. Die NICU-Daten stellen dabei eine Teilmenge der QS-Daten dar. Derzeit werden ca. 30 Datenfelder aus dem Leistungsbereich Neonatologie übermittelt, diese sind in Anhang 1 zu Anlage 4 QFR-RL, festgeschrieben. Zudem ist in dieser Richtlinie geregelt, dass jeweils ein Fünfjahreszeitraum auszuwerten und auf der Website perinatalzentren.org darzustellen ist. Dadurch ist es notwendig, den Wechsel der Institutionskennzeichen (IK-Nummer) oder Fusionen von Krankenhäusern eindeutig nachvollziehen zu können. Somit liegen in den NICU-Daten die tatsächlichen IK-Nummer vor und die Daten sind über die fünf vorliegenden Jahre gemappt. Derzeit liegt ein aktueller Datenpool der Erfassungsjahre 2014 bis 2018 vor. Da die PNZ in den NICU-Daten bekannt sind, ist weiterhin die Versorgungsstufe der Krankenhäuser eindeutig bekannt. Die zentrale Ergebnisveröffentlichung auf perinatalzentren.org ist für PNZ mit Versorgungsstufe I (Perinatalzentrum Level 1) und II (Perinatalzentrum Level 2) relevant und somit auf Daten von Kindern mit einem Geburtsgewicht bis unter 1.500 g begrenzt. Hingegen sind in den QS-Daten auch die Daten von Standorten mit den Versorgungsstufen III (Perinataler Schwerpunkt), IV (Geburtsklinik) und sonstigen Krankenhäusern enthalten.³ Aus rechtlichen Gründen können die beiden Datensätze nicht verknüpft werden. In den NICU-Daten ist aktuell noch kein einheitlicher Standortbezug hergestellt, dies wird voraussichtlich für die Veröffentlichung im Dezember 2020 avisiert. In den QS-Daten wird hingegen seit dem Jahr 2014 eine standortbezogene Auswertung vorgenommen. Für die Analysen wird im Folgenden der Fokus auf die NICU-Daten gerichtet, da sich in den QS-Daten für die vorliegenden Analysen keine ausreichende Validität für die Angabe der Versorgungsstufen gezeigt hat. Darüber hinaus können in den QS-Daten die Standorte über die verschiedenen Jahre hinweg nicht verlässlich zugeordnet werden, da hier im Gegensatz zu den NICU-Daten keine Klarnamen der Standorte vorliegen, sondern nur Krankenhauspseudonyme, welche sich zwischen den Jahren unterscheiden können. Zudem liegen die NICU-Daten bereits in einem einheitlichen Datenpool über den gesamten Fünfjahreszeitraum vor, bei dem zeitliche Veränderungen der Datenerfassung nach Möglichkeit ausgeglichen wurden. Die QS-Daten sind hingegen inhomogener, da in der Regel jedes Erfassungsjahr separat ausgewertet wird. Die Analysen in diesem Bericht beruhen daher hauptsächlich auf den NICU-Daten, während die QS-Daten für zusätzliche Sensitivitätsanalysen herangezogen werden.

³ Die Aufnahme- und Zuweiskriterien der Versorgungsstufen sind in Anlage 1 der QFR-RL geregelt.

Da sich die aktuelle Mindestmengenregelung nur auf Frühgeborene in Level-1-PNZ bezieht, wird die Datengrundlage für die Hauptanalyse entsprechend eingeschränkt. Im Rahmen von Sensitivitätsanalysen wird diese Bedingung aber gelockert und jeweils alle gelieferten Fälle der NICU- bzw. QS-Daten analysiert. In diesem Kontext wäre es denkbar, die Versorgungsstufe als Kovariable im Volume-Outcome-Modell einfließen zu lassen. Da jedoch nahezu alle verstorbenen VLBW-I in Level-1-PNZ behandelt wurden (Ergebnisse nicht dargestellt), konnte erwartet werden, dass eine Adjustierung nach Level die im Folgenden berichteten Ergebnisse nur unwesentlich beeinflusst. Zusätzlich kann mit Blick auf die Beauftragung, wie auch mit Blick auf die Formulierung in § 136b Abs. 1 SGB V zumindest infrage gestellt werden, inwieweit eine Adjustierung nach Strukturvariablen für die Ermittlung einer Mindestmenge überhaupt im Sinne des Auftrages bzw. im Sinne der gesetzlichen Regelung ist.

Im Folgenden werden entsprechend diejenigen Fälle als *Grundgesamtheit* bezeichnet, die ein Aufnahmegegewicht unter 1.250 g hatten, die in Level-1-PNZ behandelt wurden und bei denen zusätzlich folgende Bedingungen erfüllt sind:

- keine letalen Fehlbildungen,
- ein Gestationsalter von mindestens 24 + 0 SSW,
- keine primär palliative Therapie.

Kinder, die mindestens eine der letzten drei Kriterien nicht erfüllen, werden aus der Analyse ausgeschlossen, da sie kaum Überlebenschancen aufweisen und da bei Verfehlen der ersten beiden Kriterien in Absprache mit den Eltern häufig eine primäre palliative Therapie vorgenommen wird. Dies kann daher nicht den Krankenhäusern angelastet werden. Für die Auswertungen in den Bereichen der QS- und NICU-Daten wird die gleiche, eingeschränkte Grundgesamtheit zugrunde gelegt.

Für die Schätzung des Volume-Outcome-Zusammenhangs bildet diese Grundgesamtheit die Datenbasis. Der resultierende Datenpool für die Jahre 2014 bis 2018 enthält 26.043 Datensätze aus 163 PNZ von VLBW-I, davon waren 1.891 verstorben (ca. 7,26 %). Die Mortalität berechnet sich dabei aus der Angabe des Entlassungsgrundes im Dokumentationsbogen. Weiterhin werden die Morbiditäten ebenfalls im Bogen durch Selbstangabe der PNZ dokumentiert und regelhaft ausgewiesen. Die NEK (Stadium II oder III) wurde in den NICU-Daten bei 1.068 Fällen (4,10 %) dokumentiert, eine höhergradige IVH (Grad III oder PVH) bei 1.912 Fällen (7,34 %). Weiterhin fanden sich bei 1.228 Fällen (4,72 %) eine Angabe zur höhergradigen ROP (Grad III oder höher) und bei 2.795 Fällen (10,73 %) eine moderate oder schwere BPD. Die Angaben zur Pneumonie und Sepsis (nosokomiale Infektionen) sind nicht in den NICU-Daten enthalten, sodass mithilfe dieser Daten keine Analysen zu den beiden Morbiditäten durchgeführt werden konnten. In den QS-Daten des Erfassungsjahres 2018 haben die nosokomialen Infektionen (später als 3 Tage nach Geburt) eine Häufigkeit von 782 Fällen (14,03 %).

Bei der Einschätzung, ob ein PNZ jeweils eine mögliche Mindestmenge einhält, ist allerdings nicht die Anzahl der Fälle aus dieser Grundgesamtheit ausschlaggebend, sondern die (größere) Anzahl aller Fälle mit einem Aufnahmegegewicht von unter 1.250 g. Diese *Gesamtfallzahl* (in Level-1-PNZ) besteht in den vorliegenden Daten von 2014 bis 2018 aus 29.048 Fällen (davon 3.789

verstorben, also ca. 13,0 %). Im Folgenden bezieht sich der Begriff Fallzahl daher stets auf diese Gesamtfallzahl und wird bei den Berechnungen zum Volume-Outcome-Zusammenhang verwendet.

Die folgende Abbildung 1 zeigt ein Histogramm der Verteilung der durchschnittlichen Gesamtallzahl der PNZ, also dem Durchschnitt der Jahresfallzahlen pro PNZ ab 2014, die mindestens einen Fall hatten. Die kleinste durchschnittliche Gesamtfallzahl beträgt 4,6; die größte 123,4.

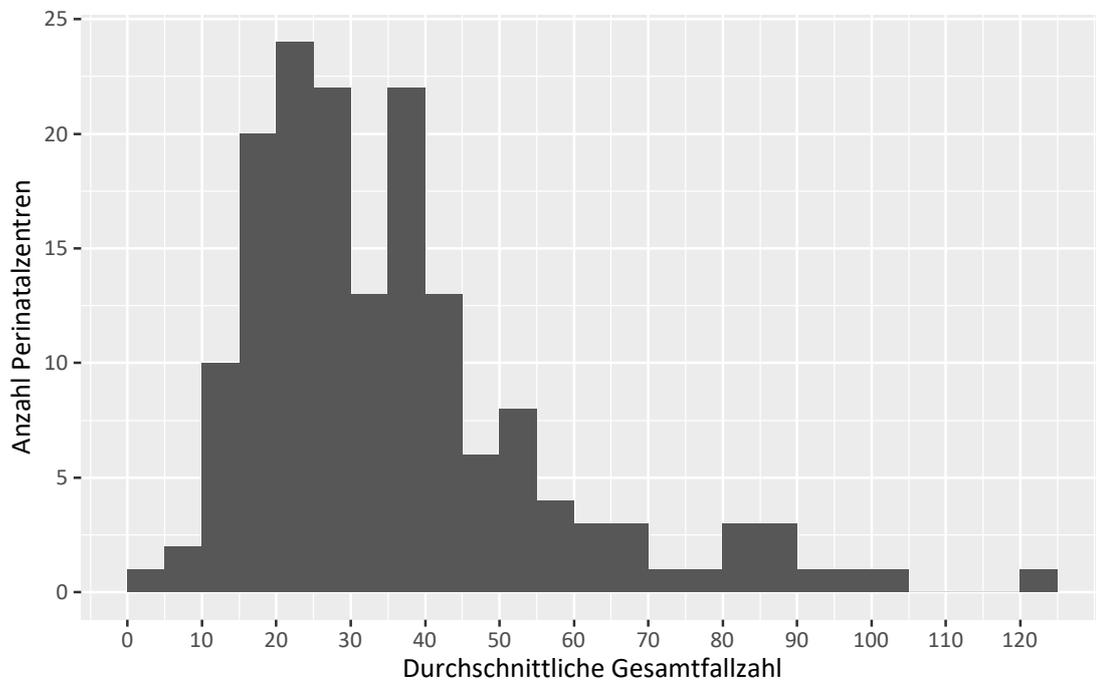


Abbildung 1: Histogramm der Verteilung der durchschnittlichen Gesamtfallzahl der PNZ in den Jahren 2014 bis 2018.

Die Lorenzkurve in der folgenden Abbildung verdeutlicht die Verteilung der Fälle auf PNZ mit entsprechenden Fallzahlen. Dabei wurde die durchschnittliche Gesamtfallzahl im Zeitraum 2014 bis 2018 betrachtet

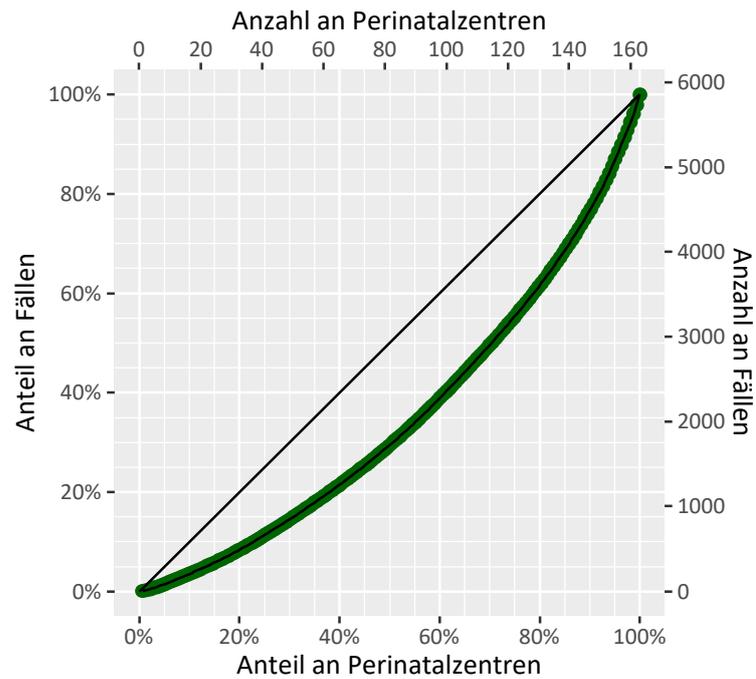


Abbildung 2: Lorenzkurve der Verteilung der Fälle auf die PNZ für die Jahre 2014-2018.

Aus der Abbildung 2 lässt sich ablesen, welcher Anteil an Fällen jeweils in den x % „kleinsten“ PNZ behandelt wurde, d. h. in den PNZ mit den niedrigsten Fallzahlen. Beispielsweise wurden in den 50 % „kleinsten“ PNZ ca. 30 % aller Fälle behandelt.

3 Endpunkte der Analysen

3.1 Mortalität

Die Krankenhausmortalität stellt den primären Endpunkt dieser Analysen dar. Der Vorteil dieses Endpunktes besteht in der eindeutigen Operationalisierung. Im Rahmen der Beauftragung zur Abbildung des Verlegungsgeschehens wurde festgestellt, dass die PNZ-spezifischen Mortalitätsraten nur vergleichsweise geringe Abweichungen zwischen Fall- und Patientenebene aufwiesen (IQTIG 2019d) (vgl. Abbildung 3).

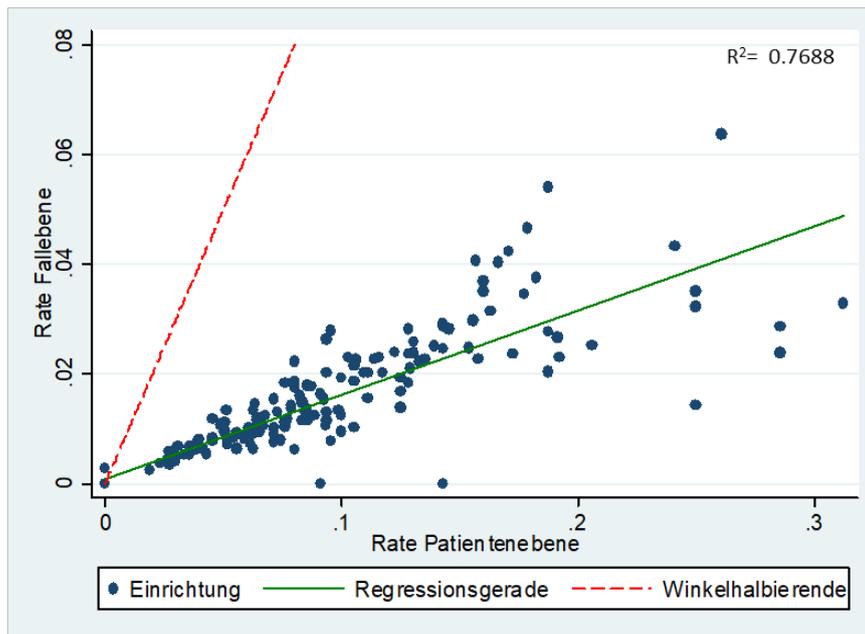


Abbildung 3: Streudiagramm zur Darstellung des Zusammenhangs zwischen der Patienten- und Fallebene beim Indikator „30-Tages-Sterblichkeit“

Auf dieser Grundlage scheint der Endpunkt Mortalität gut geeignet für Volume-Outcome-Analysen zu sein. Nahezu alle Studien zu diesem Thema nutzen ebenfalls den Endpunkt Mortalität.

3.2 Morbidität

Gemäß der Beauftragung soll wissenschaftlich geprüft werden, ob „Morbiditäten“ wie beispielsweise nekrotisierende Enterokolitis, intraventrikuläre Hämorrhagie (IVH), Retinopathie (ROP), bronchopulmonale Dysplasie (BPD), Pneumonie oder Infektionen einschließlich Sepsis sich als Endpunkte für eine Volume-Outcome-Analyse eignen. Dabei handelt es sich um traditionelle Ergebnisindikatoren der Frühgeborenenversorgung, wobei die Betrachtung und Analyse dieser Indikatoren sich ausschließlich auf definierte höhere Schweregrade bezieht (vgl. bspw. IQTIG 2019a).

Allerdings ist mit den vorhandenen Daten lediglich eine Analyse mit Fallbezug möglich. Werden Kinder verlegt, erscheinen sie als neuer Fall in einem anderen PNZ, ohne dass diese Fälle ver-

knüpft werden können. Die vulnerable Phase für das Auftreten dieser Erkrankungen scheint dabei überwiegend sehr früh (innerhalb der ersten Lebensstunden oder -tage) zu liegen, eine Diagnose ist aber erst ab einem bestimmten Gestationsalter möglich. Die Erkrankungen selbst treten aber erst zu einem späteren Zeitpunkt auf oder können erst ab einem bestimmten Gestationsalter diagnostiziert werden. Hinzu kommt, dass verschiedene schwerwiegende Morbiditäten (bzw. NEK) nur bzw. bevorzugt in PNZ mit entsprechender (z. B. kinderchirurgischer) Expertise behandelt werden, obgleich sie mutmaßlich zuvor entstanden sind. Dadurch könnten PNZ mit großer Fallzahl benachteiligt werden, da diese bspw. in Bezug auf kinderchirurgische Einrichtungen häufig besser ausgestattet sind als kleinere PNZ. Andere Erkrankungen wie z. B. ROP scheinen, was die Therapie und Diagnose einer höhergradigen Form der Erkrankung angeht, von ophthalmologischen Angebotsstrukturen vor Ort nicht unabhängig. In der Konsequenz muss davon ausgegangen werden, dass der Beginn der Morbidität und deren Zuordnung zum jeweiligen PNZ mit den hier vorliegenden Datenpools nicht darstellbar ist. Daher muss kein notwendiger kausaler Zusammenhang zwischen der Entstehung der Morbidität und dem aktuell behandelnden PNZ bestehen.

Im den Methodischen Grundlagen des IQTIG wird dieser Umstand ebenfalls adressiert. Dabei wird herausgearbeitet, dass es bei der Zuweisung eines Indikatorwerts von hoher Relevanz ist, dass dieser Leistungserbringer den Indikatorwert auch maßgeblich beeinflusst (IQTIG 2019b, S. 150). Dieses Kriterium kann für die Morbiditätsendpunkte in vielen Fällen angezweifelt werden.

Im Rahmen der Arbeit des IQTIG mit der Bundesfachgruppe *Perinatalmedizin* wie auch im Rahmen der Entwicklung und des Betriebs der Website perinatalzentren.org wurden dem IQTIG zahlreiche Meldungen zugeleitet, dass hierdurch erhebliche Verzerrungen in der Berechnung, Ausweisung und Zuschreibung der morbiditätsbedingten Ergebnisqualität entstehen. Dies war auch der Anlass für den G-BA, das AQUA-Institut (als damalige Institution nach § 137a SGB V) mit einer „Konzeptskizze zur Abbildung des Verlegungsgeschehens“ und darauffolgend das IQTIG mit der „Entwicklung eines Umsetzungskonzeptes zur Darstellung des Verlegungsgeschehens“ zu beauftragen. Ziel der Beauftragungen ist es, die Versorgungsqualität unter Berücksichtigung des Verlegungsgeschehens künftig abbilden zu können (IQTIG 2019d, AQUA 2015). Im Rahmen dieser Beauftragung wurden Analysen auf der Basis von Sozialdaten durchgeführt, so dass eine Zusammenführung der einzelnen stationären Aufenthalte und somit eine Berechnung auf Patientenebene möglich war. Die Ergebnisse der Indikatoren zur Mortalität zeigten akzeptable, diejenigen zur Morbidität dagegen nur sehr geringe Assoziationen zwischen der Patienten- und der Fallebene. Insgesamt bestätigten die Ergebnisse die Auffassung, dass die Morbidität der VLBW-I in den hier zur Verfügung stehenden Daten nur ungenügend dem verursachenden PNZ zugeordnet werden kann. Dabei soll nicht unerwähnt bleiben, dass vorherige Analysen aus Deutschland unter Nutzung von stationären Abrechnungsdaten nahegelegt hatten, dass unter Berücksichtigung des Verlegungsgeschehens ein Volume-Outcome-Effekt bei VLBW auch für Endpunkte unter Berücksichtigung der Morbidität, genauer für den Endpunkt „Sterblichkeit oder schwere Erkrankung“, existiert (Heller 2015).

Am Beispiel der ROP soll der Unterschied in der Berechnung auf Patienten- und Fallebene verdeutlicht werden, dies gilt jedoch auch für jede andere untersuchte Morbidität:

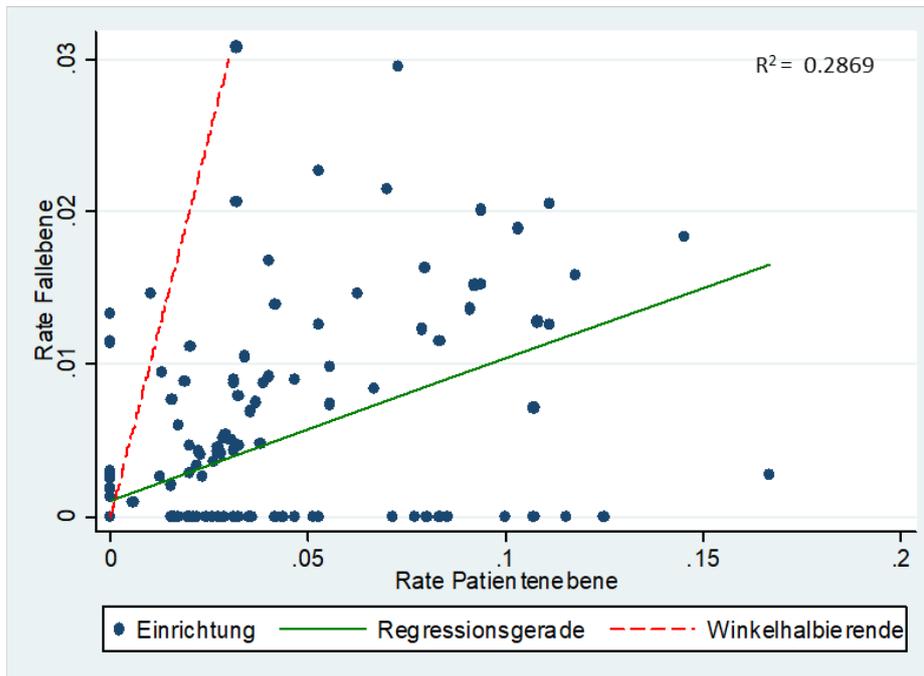


Abbildung 4: Vergleich der Ergebnisse „ROP mit Intervention“ zwischen Patienten- und Fallebene aus dem Verlegungsgeschehen.

In Abbildung 4 ist eine deutliche Streuung zwischen der Betrachtung auf Patienten- und Fallebene zu erkennen. So gibt die Regressionsgerade (r^2) einer durchgeführten linearen Regression an, dass mit Kenntnis des fallbezogenen Indikators (ROP mit Intervention) knapp 30 % der Variation des Indikators auf Patientenebene erklärt werden können. Im Umkehrschluss bedeutet dies, dass circa 70 % der Varianz des intendierten Indikators auf Patientenebene mit der bislang üblichen Berechnung auf Fallebene nicht abgebildet werden können.

Die Winkelhalbierende ist dabei die Gerade zwischen x- und y-Achse, auf der diejenigen Perinatalzentren liegen müssten, deren Ergebnisse auf Fallebene gleich mit den Ergebnissen auf der Patientenebene sind. In Abbildung 4 liegen acht Perinatalzentren oberhalb der Winkelhalbierenden. Dies kann dadurch erklärt werden, dass in diesen Perinatalzentren im Wesentlichen verlegte Kinder behandelt wurden und bei einer Analyse auf Patientenebene die Ereignisse einem anderen Perinatalzentrum zugeschrieben wurden (IQTIG 2019d).

Darüber hinaus kann postuliert werden, dass Kodierwahrscheinlichkeiten von bestehenden Erkrankungen von der Größe bzw. Fallzahl der PNZ abhängig sind. Beispielsweise kann postuliert werden, dass die Entdeckungsraten von Erkrankungen in größeren PNZ, ggf. aufgrund moderner Geräte, höher ist. Insofern könnte der Volume-Outcome-Zusammenhang für Morbiditätspunkte durch einen diagnostischen Bias verzerrt sein, ohne das eingeschätzt werden kann, wie groß dieser Bias ist. Der Entlassungsgrund Tod ist dagegen unzweifelhaft und operational sehr eindeutig.

Wir gehen davon aus, dass die genannten Gründe und Analysen auch dafür verantwortlich sind, warum nur sehr wenige Volume-Outcome-Analysen für VLBW mit Morbiditätsendpunkten in der wissenschaftlichen Literatur zu finden sind

Zusammenfassend und nach eingehenden Analysen der Daten kann festgehalten werden, dass Morbiditäten unter Nutzung der hier vorliegenden Datenpools keine geeigneten Endpunkte für Volume-Outcome-Analysen darstellen.

4 Volume-Outcome-Analyse

4.1 Inferenzmodell

In diesem Kapitel wird ein Inferenzmodell entwickelt und diskutiert, anhand dessen die Frage beantwortet werden soll, ob es einen Zusammenhang zwischen der Anzahl an behandelten Fällen in einem PNZ und der Sterbewahrscheinlichkeit von VLBW-I gibt, und welche Form dieser Zusammenhang gegebenenfalls annimmt.

Für die Zusammenhangsanalyse wird ein generalisiertes additives gemischtes Regressionsmodell (vgl. u.a. Wood 2017) verwendet, das die bedingte Wahrscheinlichkeit für das Versterben (Outcome) eines Kindes in Abhängigkeit der Fallzahl des behandelnden PNZ innerhalb eines bestimmten Zeitraums (siehe Abschnitt 4.1.1) und weiterer Einflussgrößen modelliert. Da es sich um ein binäres Outcome handelt, wird ein Modell mit Logit-Linkfunktion (logistische Regression) verwendet. Die Zielgröße für einen Fall j aus der Grundgesamtheit, der in PNZ i im Erfassungsjahr k behandelt wurde, wird im Folgenden als y_{ij}^k bezeichnet und nimmt den Wert „1“ an, wenn der Fall verstorben ist und den Wert „0“, wenn der Fall nicht verstorben ist.

Wie auch bei anderen Qualitätsmessungen kann die Sterblichkeit als komplexe Funktion verschiedener Einflussgrößen aufgefasst werden (vgl. IQTIG 2019b), Abschnitt 12.3 und Kapitel 15). Da das Ziel der Volume-Outcome-Analyse ist, den fallzahlabhängigen Anteil um andere Einflüsse bereinigt zu schätzen, werden beobachtete patientenseitige Einflüsse im Modell als Confounder berücksichtigt (siehe Abschnitt 4.1.3).

Neben weiteren Einflussgrößen, wie z. B. patientenseitigen Risikofaktoren, fließt die Fallzahl als einzige beobachtete PNZ-spezifische Einflussgröße in das Modell ein und kann als Surrogat-Parameter für unterschiedlichste systematische fallzahlabhängige Unterschiede zwischen PNZ aufgefasst werden, wie z. B. Erfahrung des Personals oder die apparative Ausstattung (vgl. z.B. Wetzel 2006, IQTIG 2019e), die nicht expliziter Teil der Modellierung sind. Nicht beobachtete Eigenschaften der PNZ, die einen systematischen Einfluss auf die Sterblichkeit haben, werden in Form von PNZ-spezifischen *random intercepts* im Modell berücksichtigt (vgl. Rogowski et al. 2004, George et al. 2017, IQTIG 2019e). In klassischen logistischen Regressionsmodellen ohne *random intercepts* werden die Outcomes aller Fälle, insbesondere derer, die beim selben PNZ behandelt wurden, als unabhängig angenommen. Dies kann dazu führen, dass die geschätzten Standardfehler (Standard Error; SE) der Effekte im Modell – speziell auch des Fallzahleffekts – unterschätzt werden. Insbesondere kann es somit zu einer Überschätzung der Sicherheit des Fallzahleffekts kommen (vgl. z.B. Urbach und Austin 2005, Panageas et al. 2003). Durch die Verwendung von *random intercepts* werden die geschätzten Standardfehler nach oben korrigiert (vgl. z.B. Fahrmeir et al. 2013). Darüber hinaus quantifiziert die Varianz der geschätzten PNZ-spezifischen *random intercepts* die Variabilität der Sterbewahrscheinlichkeiten in den unterschiedlichen PNZ aufgrund von PNZ-spezifischen Einflüssen, die nicht explizit im Modell berücksichtigt werden. Diese Variabilität kann mit der Variabilität der Sterbewahrscheinlichkeiten aufgrund der Fallzahl verglichen werden (siehe Abschnitt 4.2.1; vgl. auch Rogowski et al. 2004).

Durch die Betrachtung des Outcomes y_{ij}^k als Bernoulli-verteilte Zufallsvariable mit Sterbewahrscheinlichkeit π_{ij}^k werden unbeobachtete patientenseitige Einflüsse sowie sonstige Einflüsse, die innerhalb der PNZ variieren, als stochastische Einflüsse modelliert vgl. IQWiG (2008, Abschnitt 4.1).

Da das Modell auf Daten aus fünf verschiedenen Jahren (2014 bis 2018) beruht, muss entschieden werden, welche Einflüsse über den gesamten betrachteten Zeitraum als konstant angenommen werden können. Globale Veränderungen wie allgemeiner medizinischer Fortschritt werden im hier verwendeten Modell über jahresspezifische Effekte modelliert. Für alle anderen Einflüsse wird vereinfachend angenommen, dass eine zeitunabhängige Modellierung ausreichend ist. Dies betrifft insbesondere die Berücksichtigung der patientenseitigen Risikofaktoren. Diese Annahme gilt in gleicher Weise für das auf der Website perinatalzentren.org genutzte Risikoadjustierungsmodell (G-BA 2019a), dessen Risikofaktorvariablen auch in dem hier berechneten Modell genutzt wurden.

Auch die PNZ-spezifischen random intercepts werden als zeitunabhängig angenommen. Für einzelne PNZ mag diese Annahme verletzt sein (sei es aufgrund von strukturellen oder personellen Veränderungen am PNZ oder aufgrund einer erfolgreichen Maßnahme der internen oder externen Qualitätssicherung). Eine genauere Modellierung wäre sehr komplex und würde weitere Annahmen darüber benötigen, wie der random intercept eines PNZ von Jahr zu Jahr variiert. Die zusätzliche Clusterung, dass Outcomes im selben PNZ, die zeitlich näher beieinanderliegen, stärker korreliert sind als solche, die zeitlich auseinanderliegen, wird also vernachlässigt. Innerhalb von PNZ sind weitere Clustereffekte zu erwarten, z. B. durch die Gliederung der PNZ in Betriebsstätten oder Stationen. Da die nötigen Angaben fehlen ist eine Berücksichtigung dieser kleineren Cluster jedoch nicht möglich.

Letztlich wird auch der Fallzahleffekt als konstant angenommen (während die Fallzahl selbst von Jahr zu Jahr variiert). Grundsätzlich ist auch denkbar, dass der Effekt der Fallzahl auf die Mortalität über die Jahre variiert. Dies würde im Modell erfordern, dass der Fallzahleffekt vom Jahr abhängt, was einer Interaktion gleichkommt. Um auszuschließen, dass die Annahme zu restriktiv ist, wurde im Rahmen einer Sensitivitätsanalyse eine Interaktion berücksichtigt. Die Ergebnisse sind konsistent mit denen ohne die Interaktion zwischen Fallzahl und Jahr und legen nahe, dass die Annahme zeitkonstanter Fallzahleffekte getroffen werden kann.

Damit ergibt sich das folgende Inferenzmodell für die Sterbewahrscheinlichkeit π_{ij}^k für einen Fall j , der in PNZ i im Erfassungsjahr k behandelt wurde:

$$\text{logit}(\pi_{ij}^k) = \beta_0^k + \eta_{ij}^k + b_i^k, \quad k = 2014, \dots, 2018, \quad (1)$$

wobei β_0^k den jahresspezifischen Effekt bezeichnet, η_{ij}^k ist der Beitrag der patientenseitigen Risikofaktoren und $b_i^k = v_i^k + u_i$ ist der PNZ-spezifische Einfluss auf die Sterbewahrscheinlichkeit, der sich aus dem fallzahlabhängigen Teil, v_i^k , und dem fallzahlunabhängigen random intercept

u_i zusammensetzt. Im Folgenden wird detaillierter auf die konkrete Form der einzelnen Modellkomponenten eingegangen.⁴

4.1.1 Relevante Fallzahl

Zunächst stellt sich die Frage, welche Fallzahl genau als Prädiktor in der Volume-Outcome-Beziehung verwendet werden soll. Verschiedene Aspekte spielen hier eine Rolle.

1. Der Einfluss der Fallzahl auf das Outcome ist ein indirekter, da es sich bei der Fallzahl um einen Surrogatparameter für verschiedene Eigenschaften der PNZ handelt, die mit der Fallzahl zusammenhängen. Die jährlichen Fallzahlen schwanken von Jahr zu Jahr (die durchschnittliche jährliche relative Abweichung von der jeweiligen durchschnittlichen Fallzahl der PNZ beträgt ca. 17 %). Es empfiehlt sich daher, statt der jährlichen Fallzahl das Mittel der Fallzahlen über mehrere Jahre zu verwenden, da man damit ein verlässlicheres Maß für den Fallzahlbereich erhält, der für ein bestimmtes PNZ als typisch angesehen werden kann und somit auch einen besseren Proxy.
2. Die Fallzahl eines PNZ kann keinen kausalen Einfluss auf Behandlungen in zurückliegenden Zeiträumen haben. Beispielsweise kann die Fallzahl des Jahres 2018 keinen kausalen Einfluss auf die Frühgeborenen des Jahres 2016 haben. Daher ist es sinnvoll, als Prädiktor für die Volume-Outcome-Analyse eine Fallzahl basierend auf den jeweils vorangegangenen Jahren zu ermitteln. Bei der Verwendung des aktuellen Jahres ist diese Einschränkung nur teilweise gegeben
3. Ein Interesse an der Volume-Outcome-Beziehung besteht mit Blick auf eine mögliche Anhebung der Mindestmenge für Frühgeborene. Daher empfiehlt es sich, die Fallzahlen möglichst analog dazu zu berechnen, wie sie bei der Anwendung einer Mindestmenge zum Einsatz kommen. Dafür werden entsprechend der Mindestmengen-Regelung (Mm-R) zwei Fallzahlen basierend auf zwei überlappenden einjährigen Zeiträumen zur Bewertung der Fallzahlprognose eines PNZ herangezogen. Wenn bspw. der Effekt einer Mindestmenge für die Versorgungsstrukturen im Jahr 2020 untersucht werden soll, wären die dazu zu nutzenden Fallzahlen die des kompletten Jahres 2018, wie auch die Fallzahl vom 01.07.2018 bis zum 30.06.2019. Für die folgenden Berechnungen wird das Maximum der beiden soeben beschriebenen Fallzahlen genutzt.

Es wird der Argumentation in den Punkten 1 und 2 folgend angenommen, dass der kausale Mechanismus für einen Volume-Outcome-Zusammenhang besser durch eine über mehrere Jahre aggregierte Fallzahl abgebildet wird. Daher wird im Inferenzmodell eine im Folgenden als „kumulierte durchschnittliche Fallzahl“ bezeichnete Fallzahl verwendet. Diese ist für ein PNZ- i für Erfassungsjahr $k \geq 2014$ definiert als der Durchschnitt der von Null verschiedenen Fallzahlen n_i^l (Gesamtfallzahl) der vergangenen Jahre ab $l = 2012$ bis einschließlich des aktuellen Erfassungsjahres:

⁴ Die dargestellte Methodik in den Kapiteln 4 bis 6 ist unterexterner Beratung mit Herrn Prof. Dr. Helmut Küchenhoff entstanden.

$$\tilde{n}_i^k = \frac{\sum_{l=2012}^k n_i^l}{\sum_{l=2012}^k \mathbb{I}_{\{n_i^l > 0\}}}$$

Hierbei bezeichnet $\mathbb{I}_{\{n_i^l > 0\}}$ eine Indikatorfunktion, die den Wert „1“ annimmt, wenn das PNZ im Jahr l Frühgeborene unter 1.250 g behandelt hat und sonst den Wert „0“. Unter den 163 Level-1-PNZ gibt es lediglich drei, die in einem Jahr keine Frühgeborenen unter 1.250 g behandelten und vier, die in einem Jahr keinen Fall aus der Grundgesamtheit aufwiesen; dies ist neben zufälligen Schwankungen darauf zurückzuführen, dass einige Perinatalzentren nicht über den gesamten Zeitraum existierten.

Um den Überlegungen aus Punkt 3 Rechnung zu tragen wurde zusätzlich in einer separaten Analyse untersucht, wie sich der geschätzte Volume-Outcome-Zusammenhang verändert, wenn anstelle der kumulierten durchschnittlichen Fallzahl diejenige Fallzahlkombination verwendet wird, die aktuell in der Mindestmengenregelung vorgesehen ist (Kapitel 6) (IQTIG 2020). Die darauf basierende geschätzte Volume-Outcome-Beziehung ist qualitativ ähnlich, jedoch etwas schwächer ausgeprägt. Für die Schwellenwertbestimmung und -analysen in den Kapiteln 5 und 6 wird die kumulierte durchschnittliche Fallzahl und der darauf basierende Volume-Outcome-Zusammenhang verwendet.

4.1.2 Modellierung stetiger Variablen

Im Rahmen generalisierter additiver gemischter Modelle kann der Einfluss von kontinuierlichen (oder nahezu kontinuierlichen) Einflussgrößen jeweils durch eine kontinuierliche Funktion modelliert werden, ohne dass eine parametrische Form vorgegeben werden muss (vgl. z.B. Wood 2017). Dies wird im hier verwendeten Inferenzmodell für alle kontinuierlichen Variablen genutzt. Insbesondere wird der Einfluss der Fallzahl auf die Sterbewahrscheinlichkeit durch eine stetige Funktion modelliert (vgl. z.B. IQWIG 2006, Grouven et al. 2008, IQWIG 2019, IQTIG 2019e Abschnitt 4.2 (S.30) und 4.4 (S.32)). Dies erlaubt, dass eine große Bandbreite von möglichen Zusammenhängen abgedeckt ist und sowohl monotone als auch nicht monotone Zusammenhänge detektiert werden können, ohne dabei die Möglichkeit einfacher, linearer Verläufe auszuschließen. Eine Analyse durch Bildung von Volumengruppen, wie sie häufig in der Literatur gefunden werden kann (vgl. z.B. Birkmeyer et al. 2002, Stefoski Mikeljevic et al. 2003, Varaganam et al. 2015), lässt sich grundsätzlich als Spezialfall dieses Ansatzes betrachten (vgl. IQTIG 2019e, Abschnitt 4.2 (S. 31)). Bei dem hier gewählten Ansatz wird jedoch zusätzlich die Glattheit des geschätzten Zusammenhangs sichergestellt, was abrupte Sprünge an den Gruppengrenzen vermeidet, die im Allgemeinen nicht realistisch sind. Die Verwendung unterschiedlicher Volumengruppen erschwert zudem den Vergleich der Ergebnisse verschiedener Arbeiten. Darüber hinaus können die Ergebnisse von Analysen mit Volumengruppen erheblich von der Anzahl, Größe und Begrenzung der gewählten Volumengruppen abhängen (vgl. Grouven et al. 2008, Wetzel 2006, IQWIG 2019). Heller et al. (2007) führen daher ihre Analyse mehrfach mit verschiedenen Einteilungen in Volumengruppen (Terzile, Quartile, Quintile und Sextile) durch. Um eine Kurve zu schätzen, die die Daten möglichst gut beschreibt, aber gleichzeitig keine Überanpassung (Overfitting) erfolgt, müssen auch für die kontinuierliche Modellierung Parameter festgelegt werden. Es zeigt sich jedoch, dass die Ergebnisse mit der hier gewählten Schätzmethode (penalisierter

Spline-Ansatz, vgl. z. B. Wood (2017)) und unter Beachtung einiger grundlegender Regeln im Allgemeinen (vgl. Eilers und Marx 1996, Ruppert 2002) für diese Analyse im Speziellen relativ robust gegenüber der Wahl dieser Parameter sind (für mehr Details siehe Abschnitt 4.1.5).

Im Folgenden bezeichnet $f_{FZ}()$ den kontinuierlichen Effekt der (kumulierten durchschnittlichen) Fallzahl auf die Sterbewahrscheinlichkeit. Somit lässt sich der fallzahlabhängige Teil des PNZ-spezifischen Einflusses auf die Sterbewahrscheinlichkeit aus Modell (1) schreiben als

$$v_i^k = f_{FZ}(\tilde{n}_i^k).$$

4.1.3 Patientenseitige Risikofaktoren

Für die beiden zur Verfügung stehenden Datenquellen (NICU-Daten und QS-Daten) existiert jeweils eine Risikoadjustierung mittels indirekter Standardisierung mit zum Teil unterschiedlichen Risikofaktoren (IQTIG 2019c, G-BA 2018). Im Risikoadjustierungsmodell für die NICU-Daten gehen die folgenden Einflussgrößen ein:

- Bestehen einer schweren Fehlbildung (binär)
- Geschlecht (weiblich/nicht weiblich)
- Aufnahmegewicht (stetig)
- Abweichung des erwarteten Aufnahmegewichts vom beobachteten Aufnahmegewicht in Abhängigkeit des Gestationsalters bei Aufnahme und des Geschlechts (stetig),

wobei die letzte Größe sich wie folgt bestimmt: Es ist bekannt, dass das Gestationsalter bei Aufnahme und das Geschlecht eng mit dem Aufnahmegewicht in Verbindung steht (es besteht etwa eine hohe positive Korrelation bzw. Kollinearität), was zu großer Schätzunsicherheit oder numerischer Instabilität führen kann. Um dies zu vermeiden, ohne auf die Information aus dem Gestationsalter zu verzichten, wurde ein Orthogonalisierungsansatz verwendet: Das beobachtete Aufnahmegewicht A_B wird als lineare Funktion des Kindesalters (Anzahl abgeschlossener Gestationsstage, wochengenau, plus Lebenstag bei Aufnahme) und des Geschlechts (wie oben) modelliert. Die resultierende Abweichung zwischen dem erwarteten (A_E) und dem beobachteten Aufnahmegewicht korreliert kaum mit dem beobachteten Aufnahmegewicht. Die daraus abgeleitete stetige Größe

$$D = \frac{A_E - A_B}{A_E + A_B}$$

ist betragsmäßig durch 1 beschränkt. Diese Größe wird im Folgenden kurz als Gewichtsabweichung bezeichnet.

Anstelle des Aufnahmegewichts und der Gewichtsabweichung wird im Risikoadjustierungsmodell für die QS-Daten das Gestationsalter berücksichtigt (vgl. z.B. Tyson et al. 2008). Im hier verwendeten Volume-Outcome-Modell werden die Risikofaktoren des NICU-Modells berücksichtigt, da der Fokus der Auswertungen auf die NICU-Daten gerichtet ist und basierend auf ihnen auch die Schwellenwertanalyse erfolgt (G-BA 2019a).

Das Ziel der Berücksichtigung der patientenseitigen Risikofaktoren im Volume-Outcome-Modell ist es, den Fallzahleffekt möglichst um andere Einflüsse bereinigt schätzen zu können. Die gemeinsame Schätzung aller Effekte in einem Modell verhindert Confounding der patientenseitigen Risikofaktoren bei der Schätzung des Fallzahleffektes und umgekehrt. Darüber hinaus werden durch die Berücksichtigung der Korrelation der Outcomes von Fällen, die beim selben PNZ behandelt wurden, die geschätzten Standardfehler der patientenseitigen Effekte korrigiert sowie die statistische Unsicherheit der Schätzung der Risikokoeffizienten in der Volume-Outcome-Modellierung berücksichtigt.

Der Beitrag der patientenseitigen Risikofaktoren auf die Sterbewahrscheinlichkeit in Modellformel (1) für einen Fall j , der in PNZ i im Erfassungsjahr k behandelt wurde, nimmt somit die folgende Form an:

$$\eta_{ij}^k = \mathbf{x}_{ij}^k \boldsymbol{\beta} + f_{\text{Gew}}(A_{ij}^k) + f_{\text{Diff}}(D_{ij}^k),$$

wobei \mathbf{x}_{ij}^k ein Vektor ist, der die faktoriellen Risikofaktoren, Bestehen einer schweren Fehlbildung sowie Geschlecht, enthält und $\boldsymbol{\beta}$ bezeichnet den dazugehörigen Koeffizientenvektor. Die Funktion $f_{\text{Gew}}(A_{ij}^k)$ modelliert den kontinuierlichen Effekt des Aufnahmegegewichts A_{ij}^k und $f_{\text{Diff}}(D_{ij}^k)$ den kontinuierlichen Effekt der Gewichtsabweichung D_{ij}^k .

4.1.4 Konkretes Modell

Setzt man die Komponenten aus den Abschnitten 4.1.1 und 4.1.3 in das Inferenzmodell in Formel (1) ein, ergibt sich das folgende Gesamtmodell für die Sterbewahrscheinlichkeit π_{ij}^k für einen Fall j aus der Grundgesamtheit, der in PNZ $i \in \{1, \dots, 163\}$ im Erfassungsjahr $k \in \{2014, \dots, 2018\}$ behandelt wurde:

$$\begin{aligned} \text{logit}\left(\mathbb{P}(y_{ij}^k = 1 \mid \mathbf{x}_{ij}^k, A_{ij}^k, D_{ij}^k, \tilde{n}_i^k, u_i)\right) &= \text{logit}(\pi_{ij}^k) = \beta_0^k + \eta_{ij}^k + b_i \\ &= \beta_0^k + \mathbf{x}_{ij}^k \boldsymbol{\beta} + f_{\text{Gew}}(A_{ij}^k) + f_{\text{Diff}}(D_{ij}^k) + f_{\text{FZ}}(\tilde{n}_i^k) + u_i. \end{aligned} \quad (2)$$

Dabei gehen folgende Größen ein:

Größe	Beschreibung
$\beta_0^k \in \mathbb{R}$	jahresspezifischer Effekt
\mathbf{x}_{ij}^k	patientenseitige diskrete Risikofaktoren
$\boldsymbol{\beta}$	Koeffizientenvektor der patientenseitigen diskreten Risikofaktoren
A_{ij}^k	Aufnahmegegewicht
$f_{\text{Gew}}: \mathbb{R} \rightarrow \mathbb{R}$	Effekt des Aufnahmegegewichts
D_{ij}^k	Gewichtsabweichung
$f_{\text{Diff}}: \mathbb{R} \rightarrow \mathbb{R}$	Effekt der Gewichtsabweichung
\tilde{n}_i^k	kumulierte durchschnittliche (relevante) Fallzahl

Größe	Beschreibung
$f_{FZ}: \mathbb{R} \rightarrow \mathbb{R}$	Effekt der relevanten Fallzahl
$u_i \in \mathbb{R}$	PNZ-spezifischer random intercept

Bedingt auf alle Kovariablen, insbesondere auf die PNZ-spezifischen random intercepts, werden die Sterbewahrscheinlichkeiten als unabhängig angenommen, d. h.

$$y_{ij}^k \mid \mathbf{x}_{ij}^k, A_{ij}^k, D_{ij}^k, \tilde{n}_i^k, u_i \stackrel{\text{unabhängig}}{\sim} \text{Ber}(\pi_{ij}^k).$$

Für die PNZ-spezifischen random intercepts wird, wie in der Literatur üblich (vgl. z. B. auch Agresti (2013), Stroup (2013), Gelman und Hill (2007) und McCulloch und Neuhaus (2011)), angenommen, dass sie unabhängig und identisch normalverteilt sind, d. h.

$$u_i \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(0, \tau^2).$$

Eine solche Verteilungsannahme stabilisiert die Schätzung der insgesamt 163 zu schätzenden random intercepts u_i (vgl. z. B. Fahrmeir et al. (2013)). Die Varianz der PNZ-spezifischen random intercepts lässt sich als Maß für die Heterogenität der PNZ abzüglich der im Modell berücksichtigten PNZ-spezifischen Einflüsse auffassen (Between-Provider-Varianz, vgl. Abschnitt 4.2). Es gibt Arbeiten, die zeigen, dass Modelle mit random intercepts relativ robust gegenüber der konkreten Verteilungsannahme der random intercepts sind (vgl. z. B. McCulloch und Neuhaus (2011)). Die Berücksichtigung der Clusterung in Form der PNZ-spezifischen random intercepts kann sich auf die geschätzten Standardfehler der restlichen Modellkomponenten auswirken. Grundsätzlich existieren neben Modellen mit random intercepts auch weitere Möglichkeiten, die Clusterung zu berücksichtigen. Für eine Diskussion zur Verwendung von generalisierten additiven gemischten Modellen mit random intercepts versus marginalen Modellen in diesem Kontext sei an dieser Stelle auf den Abschlussbericht zur Sonderauswertung zu Volumen-Outcome-Beziehungen bei Revisionseingriffen in der Knieendoprothetik verwiesen (IQTIG 2019e).

Die PNZ-spezifischen random intercepts und die restlichen Kovariablen, insbesondere die relevanten Fallzahlen, \tilde{n}_i^k , werden als voneinander unabhängig angenommen (sog. random effects assumption; vgl. z. B. Townsend et al. (2013)).

Aufgrund der ohnehin hohen Modellkomplexität und der daraus zu erwartenden statistischen Schätzprobleme mit der zur Verfügung stehenden Datenbasis werden im verwendeten Modell keine Interaktionen zwischen den patientenseitigen Risikofaktoren und den PNZ-spezifischen Einflüssen berücksichtigt. Wie oben beschrieben werden alle Modellkomponenten außer dem jahresspezifischen Effekt als über die betrachteten Erfassungsjahre konstant angenommen.

4.1.5 Schätzung der Modellkomponenten

Das Hauptinteresse liegt in der Funktion $f_{FZ}()$, welche die Information über die Volume-Outcome-Beziehung enthält. Die Verwendung des penalisierten Spline-Ansatzes ermöglicht, dass der Fallzahleffekt und die anderen kontinuierlichen Effekte gemeinsam mit allen anderen Mo-

dellkomponenten, insbesondere den PNZ-spezifischen random intercepts, im Rahmen eines penalisierten Maximum-Likelihood-Ansatz geschätzt werden kann (vgl. z.B. Wood (2017) für eine genaue Beschreibung des Schätzverfahrens). Für die Schätzung wurde das R-Paket `mgcv` verwendet, welches flexible Spezifikationen, insbesondere für Modelle mit random intercepts in der Kombination mit der kontinuierlichen Modellierung stetiger Variablen bietet (vgl. R Core Team 2020, Wood 2019). Das R-Paket liefert insbesondere Standardfehler und Wald-ähnliche Konfidenzintervalle für alle Modellkomponenten, die auf dem penalisierten Maximum-Likelihood-Ansatz beruhen.

Die Schätzung der kontinuierlichen Effekte, insbesondere des Fallzahleffektes $f_{FZ}()$, basiert auf einem Glättungsansatz anhand von sogenannten penalisierten Splines (vgl. Eilers und Marx 1996). Für die Spline-Konstruktion wird der Bereich der Fallzahlen in eine gewisse Anzahl von möglicherweise verschiedenen großen Teilstücken zerlegt (gesteuert durch die Anzahl und Position der sogenannten Knoten); auf diesen Teilstücken werden die unbekannt Funktionen dann derart durch Polynome approximiert, dass die resultierende zusammengesetzte Funktion glatt ist und keine sehr erratische Form hat (siehe z.B. Fahrmeir et al. 2013), Kapitel 8 für eine Einführung). Im Rahmen dieser Analyse wurden kubische Splines gewählt, die abrupte Sprünge vermeiden und lineare, aber auch höhergradige Zusammenhänge abbilden können. Als Basis wurden sogenannte „cardinal“-Splines verwendet, die numerisch sehr effizient sind (vgl. Wood 2017). Grundsätzlich hat die konkrete Basiswahl sowie der Polynomgrad wenig Einfluss auf die Schätzung (vgl. Perperoglou et al. 2019), was insbesondere auch vorangegangene Analysen gezeigt haben (vgl. IQTIG 2019e). Eine starke Krümmung der jeweiligen geschätzten Funktion wird durch einen Bestrafungsterm kontrolliert, der das Integral der quadrierten zweiten Ableitung bestraft. Ein Glättungsparameter kontrolliert dabei den Einfluss des Bestrafungsterms und somit die Glattheit der geschätzten Kurve. Dieser wird derart aus den Daten geschätzt, dass die geschätzte Kurve die Daten möglichst gut beschreibt, gleichzeitig aber keine Überanpassung (Overfitting) erfolgt (für technische Details siehe (Wood 2017), dort Abschnitt 5.3.1). Die datengesteuerte Schätzung der Glättung ist eine besonders vorteilhafte Eigenschaft von penalisierten Splines, da hierdurch die konkrete Wahl der Anzahl und Lage der Knoten – sofern sie ausreichende Flexibilität gewährleisten – nicht sehr entscheidend ist (vgl. Eilers und Marx (1996), Fahrmeir et al. (2013) und Perperoglou et al. (2019)). Für die Approximation der Funktionen $f_{Gew}()$, $f_{Diff}()$ und $f_{FZ}()$ wurden jeweils 20 quantilbasierte Knoten verwendet. Als Sensitivitätsanalyse wurde die Anzahl an Knoten variiert und der Einfluss auf die Schätzung überprüft. Dabei bestätigte sich, dass die Form der geschätzten Kurven robust gegenüber der Wahl der Anzahl der Knoten ist. Es wurden zusätzlich die drei folgenden statistischen Kriterien für die verschiedenen Anzahlen an Knoten evaluiert:

- AUC, in-sample und out-of-sample (Maß zur Vorhersagekraft).
- Konditionales AIC (Abwägung zwischen Modellanpassung und -komplexität, siehe Wood (2017)).
- Geschätzte Standardfehler im Modell (vgl. Wood (2017)).

Sie alle blieben für die betrachteten Knotenanzahlen relativ stabil. Grafische Darstellungen der geschätzten Effekte finden sich jeweils in Abschnitt 4.2.1.

Die Identifizierbarkeit des Volume-Outcome-Modells ist gegeben durch die Bedingung, dass jeweils die mittleren geschätzten Werte von $f_{\text{Gew}}()$, $f_{\text{Diff}}()$ und $f_{\text{FZ}}()$ (gemittelt über alle Ausprägungen) sowie der mittlere geschätzte Wert von u_i (gemittelt über alle PNZ) jeweils gleich 0 sind (sum-to-zero-constraint und Verteilungsannahme der random intercepts) (vgl. Wood, 2017).

4.1.6 Was liefert das Modell

Wie oben beschrieben, soll das Volume-Outcome-Modell die Frage beantworten, ob es einen Zusammenhang zwischen der Anzahl an behandelten Fällen in einem PNZ und der Sterbewahrscheinlichkeit bei Frühgeborenen mit einem Aufnahmegegewicht unter 1.250 g gibt, und welche Form dieser Zusammenhang gegebenenfalls annimmt.

Der erste Teil der Frage entspricht dem statistischen Testproblem $H_0: f_{\text{FZ}} \equiv 0$ gegen $H_1: f_{\text{FZ}} \neq 0$, also der Nullhypothese, dass f_{FZ} konstant gleich 0 ist, gegen die Alternativhypothese, dass f_{FZ} irgendwo im Fallzahlbereich von 0 abweicht. Ein geeigneter statistischer Test ist in Marra und Wood (2012) beschrieben und in der hier verwendeten Software implementiert (vgl. Wood 2019). Der Test lässt sich aus simultanen Konfidenzbändern um die geschätzte Funktion \hat{f}_{FZ} ableiten. Zur Entscheidung über statistische Signifikanz wird hier das Niveau $\alpha = 5\%$ verwendet. Die zugehörigen simultanen 95 % Konfidenzbänder werden in den grafischen Darstellungen in Abschnitt 4.2.1 entsprechend abgebildet. Generell gilt, dass ein in diesem Sinne statistisch signifikantes Ergebnis keinen besonders starken Volume-Outcome-Zusammenhang signalisiert, sondern dass die Daten unter den oben genannten Modellannahmen der Hypothese widersprechen, dass es keinen Zusammenhang gibt. Bei der abschließenden Bewertung der Ergebnisse sollten neben der statistischen Signifikanz auch die klinische Relevanz einbezogen werden.

Der zweite Teil der Frage lässt sich, basierend auf den geschätzten Modellkomponenten, visuell darstellen (vgl. Abschnitt 4.2.1). Da dort auch die Konfidenzbänder eingezeichnet werden können, lässt sich auch direkt beurteilen, auf wie viel Informationen der geschätzte Zusammenhang in den jeweiligen Fallzahlbereichen basiert. Durch die Berechnung von geschätzten Odds-Ratios für bestimmte (exemplarische) Fallzahlkonstellationen lässt sich die Form des Zusammenhangs zusätzlich auf andere Art veranschaulichen (vgl. Abschnitt 4.2.1). Die Odds-Ratios lassen sich dabei wie folgt bestimmen: Für die Odds für das Versterben bei einem PNZ mit Fallzahl \tilde{n}_1 gegenüber einem PNZ mit Fallzahl \tilde{n}_2 mit dem gleichen random intercept u besteht für zwei sonst identische Fälle j und j' (behandelt im selben Jahr und mit $\eta_{ij}^k = \eta_{i'j'}^k$) das Verhältnis

$$\frac{\text{Odds}(y_{ij}^k = 1 | \eta_{ij}^k, u, \tilde{n}_1)}{\text{Odds}(y_{i'j'}^k = 1 | \eta_{i'j'}^k, u, \tilde{n}_2)} = \exp(f_{\text{FZ}}(\tilde{n}_1) - f_{\text{FZ}}(\tilde{n}_2)).$$

Neben der Beurteilung der statistischen Signifikanz des Fallzahleffekts sowie seiner Form liefert das Modell auch eine Schätzung für die Variabilität der PNZ-spezifischen random intercepts, welche die Variabilität der Sterbewahrscheinlichkeiten zwischen PNZ quantifiziert, die auf Unterschiede jenseits der Fallzahl, von patientenseitigen Risikofaktoren oder Einflüssen des Erhebungsjahres zurückgeführt werden kann. Diese geschätzte Variabilität kann mithilfe sog. Median Odds-Ratios (MOR) der random intercepts (vgl. Larsen et al. 2000) direkt mit den Odds-Ratios

für bestimmte Fallzahlkonstellationen verglichen werden, was eine Beurteilung der Größenordnung des Fallzahleffektes gegenüber der Größenordnung nicht fallzahlabhängiger Unterschiede erlaubt (vgl. Abschnitt 4.2.1). Der MOR der random intercepts ist dabei wie folgt definiert: Würden zwei Fälle mit identischen Risikofaktoren von zwei zufällig ausgewählten PNZ mit gleicher relevanter Fallzahl im gleichen Jahr behandelt, so wäre der Odds-Ratio zwischen der Behandlung bei dem PNZ mit dem niedrigeren Risiko und der Behandlung bei PNZ mit dem höheren Risiko im Median

$$\text{MOR}(\tau) = \exp\left(-\sqrt{2\tau^2}\Phi^{-1}(3/4)\right),$$

wobei Φ^{-1} die inverse Verteilungsfunktion der Standardnormalverteilung bezeichnet. Der (geschätzte) MOR liegt grundsätzlich zwischen null und eins. Ein Ergebnis nahe eins deutet darauf hin, dass die im Modell nicht explizit berücksichtigten Einflussfaktoren auf Ebene der PNZ nur einen geringen Einfluss auf die Sterbewahrscheinlichkeit haben.

4.2 Hauptergebnisse

Tabelle 1 zeigt einige statistische Kennzahlen des Modellfits (vgl. Abschnitte 4.1.3, 4.2 und Kapitel 5 in IQTIG (2019e)). Die relativ hohen AUC-Werte zeigen eine gute Beschreibung der Daten durch das Modell an. Der Volume-Outcome-Zusammenhang f_{FZ} hat einen p-Wert von 0,014. Damit ist der Volume-Outcome-Zusammenhang signifikant zum Signifikanzniveau 0,05; d. h. die Hypothese, dass kein Fallzahleffekt besteht, wird durch den statistischen Test (vgl. Abschnitt 4.1.6) abgelehnt. Genauer wird der geschätzte Zusammenhang im nächsten Abschnitt 4.2.1 analysiert.

Tabelle 1: Statistische Kennzahlen des Modellfits

Größe	Wert
p-Wert zum Volume-Outcome-Zusammenhang	0,014
AUC zum Volume-Outcome-Modell	in-sample: 0,831; out-of-sample: 0,821
Geschätzte Varianz der random intercepts mit Konfidenzintervall	0,183, KI: [0,126, 0,2645]
MOR der random intercepts mit Konfidenzintervall	0,665, KI: [0,612, 0,713]

4.2.1 Der geschätzte Fallzahleffekt

Der oben genannte Glättungsansatz führt zu einer geschätzten Funktion \hat{f}_{FZ} für den Fallzahleffekt f_{FZ} , die monoton fallend und (auf Logitebene) nahezu linear ist. Höhere Fallzahlen sind demnach mit einer geringeren Sterblichkeit assoziiert. Die Funktion ist in Abbildung 5 dargestellt. Die gestrichelten Linien in dieser Abbildung liegen bei $\hat{f}(\tilde{n}) \pm \hat{\tau}$ und veranschaulichen die geschätzte Schwankung der random intercepts. Über der Abbildung ist ein Histogramm der Fall-

zahlverteilung abgebildet. Man sieht, dass es nur wenige PNZ mit einer kumulierten durchschnittlichen Fallzahl von über 90 gibt. In höheren Fallzahlbereichen ist der Kurvenverlauf daher mit großer Unsicherheit behaftet und mit Vorsicht zu interpretieren.

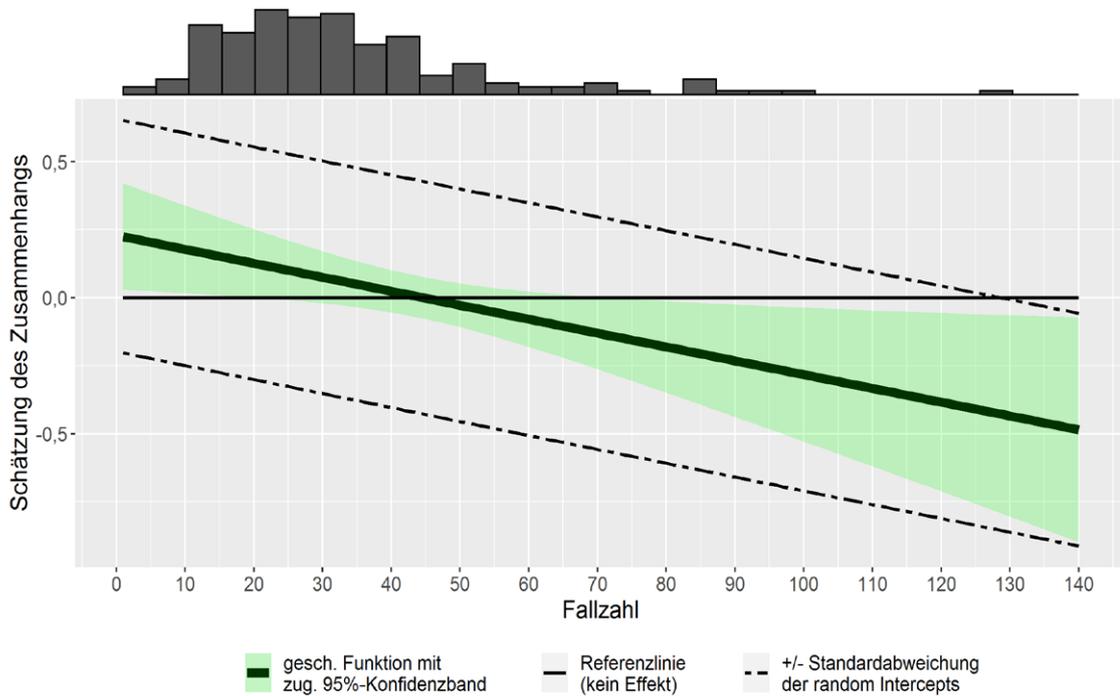


Abbildung 5: Geschätzter Fallzahleffekt f_{FZ} auf logistischer Ebene

Die dickere schwarze Kurve in Abbildung 6 zeigt die geschätzte Wahrscheinlichkeit des Versterbens für einen durchschnittlichen Fall, der in einem PNZ mit random intercept $u^* = 0$ und fester Fallzahl \tilde{n} behandelt wird. Der Fall ist in dem Sinne durchschnittlich, dass er die mittlere geschätzte Sterbewahrscheinlichkeit p^* aller in das Modell eingehenden Fälle unter Weglassen des Fallzahleffekts hat.

Die eingezeichneten Werte sind demnach

$$\mathbb{P}(y = 1 \mid \eta^*, \tilde{n}) = \text{logit}^{-1}(\eta^* + \hat{f}_{FZ}(\tilde{n})),$$

wobei $\eta^* = \text{logit}(p^*)$. Die waagerechte Referenzlinie folgt dem analogen Schema, wobei der Fallzahleffekt f_{FZ} auf null gesetzt ist. Zusätzlich eingezeichnet ist eine waagerechte Referenzlinie beim Wert p^* . Sie liegt mit 7,28 % sehr nah an der empirischen Sterblichkeitsrate in der betrachteten Grundgesamtheit (vgl. Kapitel 2).

Die gestrichelten Linien veranschaulichen die geschätzte Schwankung der random intercepts anhand der zugehörigen Standardabweichung τ . Sie entsprechen den Werten

$$\text{logit}^{-1}(\eta^* + \hat{f}_{FZ}(\tilde{n}) \pm \hat{\tau}).$$

Die Wahl eines Wertes η^* , der hier zur genannten durchschnittlichen Sterblichkeit gehört, ist dabei für die grafische Darstellung notwendig. Die statistischen Größen wie die Signifikanz des Fallzahleffekts, die Odds-Ratios und Aussagen zum qualitativen Verlauf der Kurve (Monotonieverhalten) hängen jedoch nicht von η^* ab.

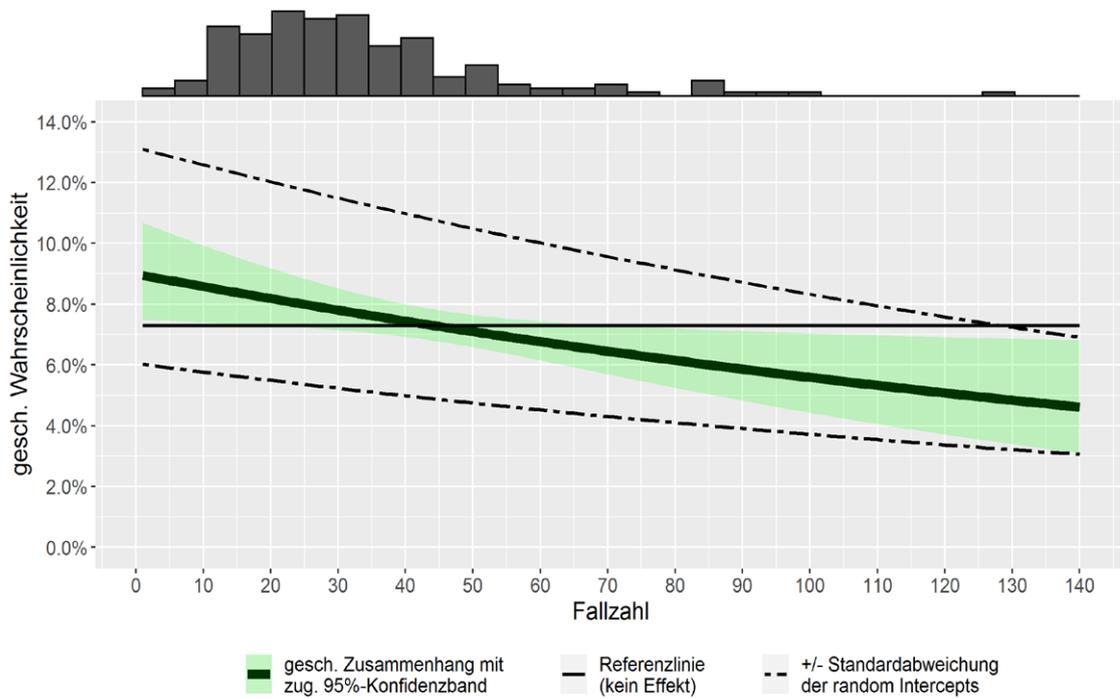


Abbildung 6: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl.

Bei der Interpretation der geschätzten Funktion f_{FZ} in Abbildung 6 muss beachtet werden, dass der Fallzahleffekt zur Visualisierung für einen durchschnittlichen Fall dargestellt wird. Zum Beispiel ist es so, dass Frühgeborene mit hohem Risiko (im Sinne der betrachteten Risikofaktoren) nur sehr selten in PNZ mit sehr geringer Fallzahl auftreten. Die dargestellte Wahrscheinlichkeit bei einer Fallzahl darf also nicht mit der durchschnittlichen Sterbewahrscheinlichkeit für PNZ bei dieser Fallzahl verwechselt werden. Eine Analyse des geschätzten Volume-Outcome-Zusammenhangs unter Berücksichtigung der Risikoprofile findet in Kapitel 6 statt.

Neben der grafischen Darstellung des Fallzahleffekts werden zusätzlich die geschätzten Odds-Ratios für bestimmte Fallzahlkonstellationen ermittelt (vgl. Abschnitt 4.1.6). Tabelle 2 zeigt, mit welchem Faktor sich die geschätzten Odds für verschiedene exemplarisch gewählte Fallzahlpaare $(\tilde{n}_1, \tilde{n}_2)$ verändern:

Tabelle 2: Odds-Ratios für verschiedene Fallzahlpaare.

Fallzahl \tilde{n}_1 \tilde{n}_2	1	5	10	20	40	60	80	100
1	1	0,98	0,955	0,907	0,819	0,739	0,667	0,602
5		1	0,975	0,926	0,836	0,755	0,681	0,615
10			1	0,95	0,858	0,774	0,699	0,631
20				1	0,903	0,815	0,735	0,664

Fallzahl \tilde{n}_1 \tilde{n}_2	1	5	10	20	40	60	80	100
40					1	0,903	0,815	0,735
60						1	0,903	0,815
80							1	0,903
100								1

Da die geschätzte Funktion \hat{f}_{FZ} nahezu linear ist, lassen sich die Odds-Ratios basierend auf der linearen Approximation bestimmen. Dies vereinfacht auch die Interpretation der Ergebnisse und die Angabe von Konfidenzintervallen. Durch die lineare Approximation lässt sich allgemeiner angeben, welchen Einfluss z. B. jeweils eine Erhöhung der Fallzahl um 10 auf die Odds hat, das heißt, um welchen Faktor sich die Odds für das interessierende Ereignis verändern, wenn die Fallzahl um 10 erhöht wird. Tabelle 3 gibt den geschätzten Fallzahleffekt im Falle linearer Modellierung sowie den darauf basierenden Odds-Ratio mit zugehörigen transformierten Wald-ähnlichen Konfidenzintervallen an:

Tabelle 3: Regressionskoeffizient des linearisierten Volume-Outcome-Zusammenhangs sowie entsprechende Odds Ratio für Fallzahländerung.

	Wert	Konfidenzintervall
Koeffizientenschätzer	-0,00512	[-0,00105 , -0,00919]
OR bei Erhöhung um 10 Fälle	0,95000	[0,912 , 0,99]

Die angegebenen Werte für die Odds-Ratios lassen sich mit dem MOR der random intercepts in Tabelle 1 vergleichen, der ein Maß dafür ist, wie sehr die Sterbewahrscheinlichkeit bei unterschiedlichen PNZ mit identischer Fallzahl für Fälle mit identischem Risikoprofil variiert (vgl. Abschnitt 4.1.6). Es zeigt sich, dass diese Variabilität groß ist im Vergleich mit der Fallzahlabhängigkeit der risikoadjustierten Mortalität. Dies deckt sich mit den Ergebnissen von (Rogowski et al. 2004, Heller 2009, Kutschmann et al. 2012, Heller 2018), die ebenfalls eine große Variabilität der Sterblichkeit zwischen PNZ vergleichbarer Fallzahl gefunden haben.

4.2.2 Sonstige Kovariablen

Für die diskreten Kovariablen erhält man die im Folgenden in Tabelle 4 angegebenen geschätzten Parameterwerte für den Endpunkt Mortalität:

Tabelle 4: Parameter für nicht-stetige Kovariablen

Größe	Koeffizient	Standardfehler	KI
Jahr 2014	-3,058	0,079	[-3,214 , -2,902]
Jahr 2015	-3,026	0,078	[-3,178 , -2,873]

Größe	Koeffizient	Standardfehler	KI
Jahr 2016	-3,020	0,078	[-3,172 , -2,868]
Jahr 2017	-2,999	0,077	[-3,15 , -2,849]
Jahr 2018	-2,932	0,077	[-3,082 , -2,781]
schwere Fehlbildung (Referenzkategorie: keine oder leichte Fehlbildung)	2,209	0,090	[2,032 , 2,386]
Geschlecht weiblich (Referenzkategorie: nicht weiblich)	-0,551	0,054	[-0,656 , -0,446]

Die jahresabhängigen Effekte unterscheiden sich nur wenig zwischen den Jahren. Dies deckt sich mit dem NICU-Modell (G-BA 2018).

Für die beiden binären patientenseitigen Risikofaktoren schwere Fehlbildungen und weibliches Geschlecht ergeben sich aus den Koeffizienten die folgenden Odds-Ratios in Tabelle 5:

Tabelle 5: Odds-Ratios für nicht stetige Kovariablen

Größe	Odds Ratio	KI
schwere Fehlbildung (Referenzkategorie: keine oder leichte Fehlbildung)	9,108	[7,629 , 10,874]
Geschlecht weiblich (Referenzkategorie: nicht weiblich)	0,576	[0,519 , 0,64]

Wie in Abschnitt 4.1.2 beschrieben erfolgt auch eine glatte Schätzung des Effekts des Aufnahmegewichts sowie der Gewichtsabweichung. Abbildung 7 und Abbildung 8 zeigen die entsprechenden Ergebnisse mit demselben Grundprinzip wie für die Fallzahl (siehe Abschnitt 4.2.1), wobei entsprechend das Aufnahmegewicht bzw. die Gewichtsabweichung die Rolle der Fallzahl übernehmen.

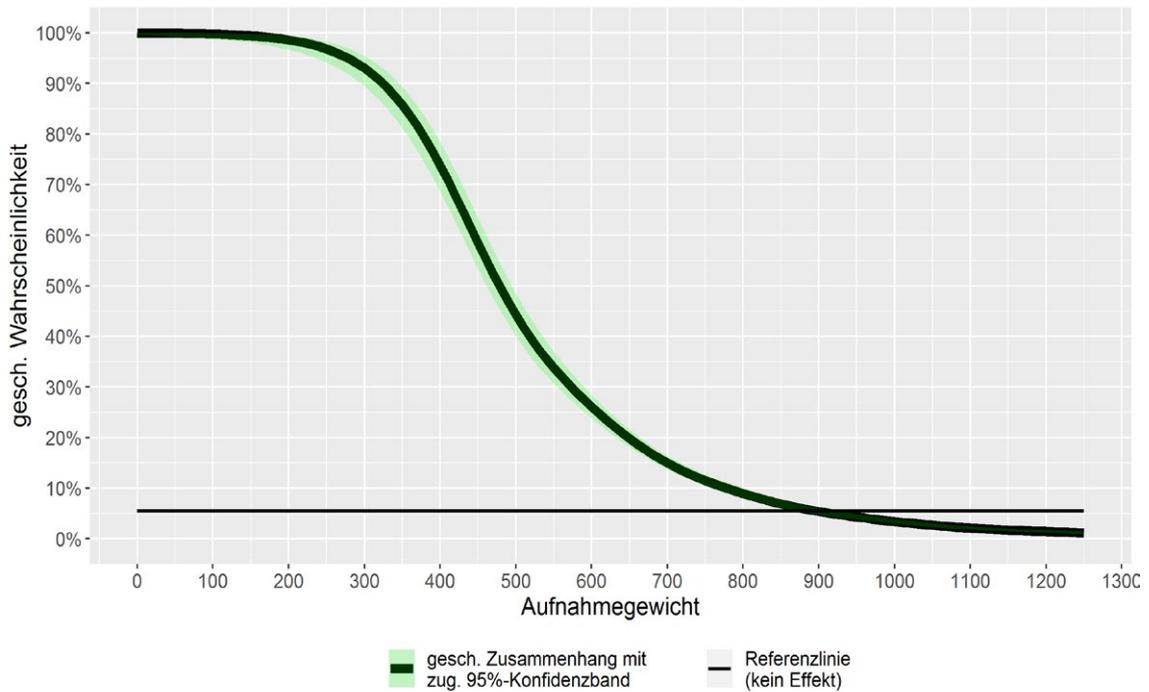


Abbildung 7: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit vom Aufnahmegegewicht.

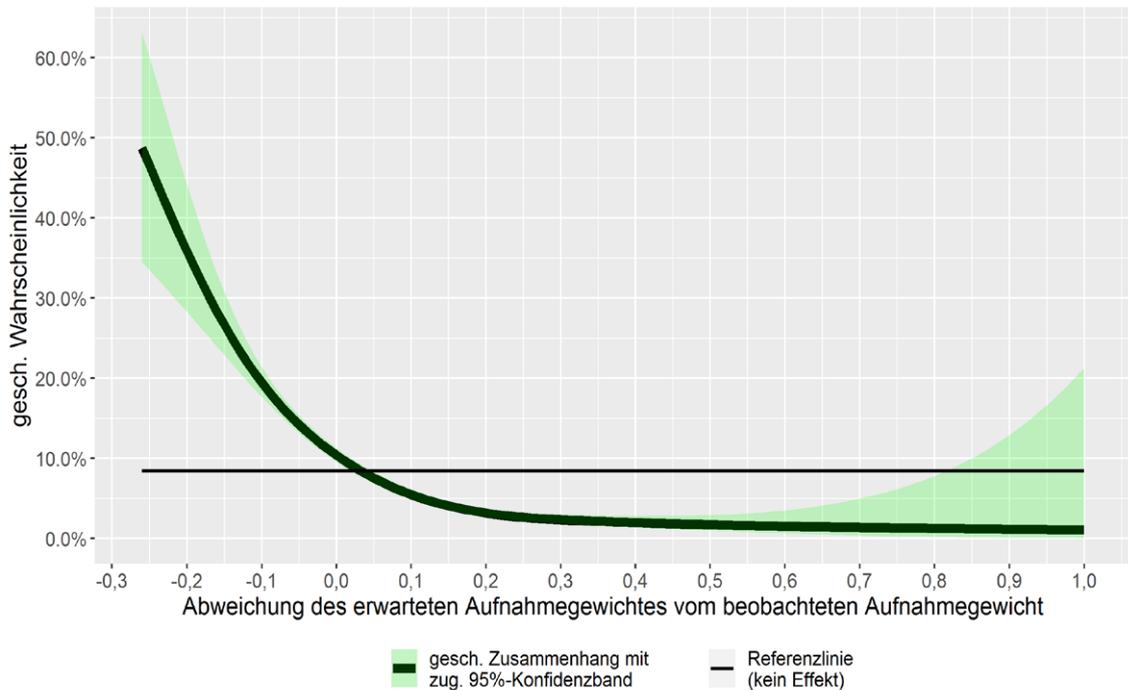


Abbildung 8: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Gewichtsabweichung.

Die Unsicherheit im hinteren Bereich spiegelt gut die Anzahl der vorhandenen Daten wider: Nur 34 der vorhandenen 26.043 Werte liegen über 0,5, und nur noch ein Wert liegt über 0,8 und beträgt ca. 0,985.

4.2.3 Die PNZ-spezifischen random intercepts

In Abbildung 9 sind die geschätzten PNZ-spezifischen random intercepts nach der Fallzahl aufgetragen. Dies dient der Überprüfung der Annahme, dass die PNZ-spezifischen random intercepts von der Fallzahl unabhängig sind (vgl. Abschnitt 4.1.4). Zur besseren Interpretierbarkeit wurde zusätzlich analog zu Abschnitt 4.1.5 eine Glättungsfunktion („smoother“; dicke schwarze Linie) mit zugehörigem Konfidenzband durch die Daten gelegt, wobei die Linkfunktion die Identität anstatt des Logits ist. Die Abbildung weist auf keine Verletzung dieser Annahme hin.

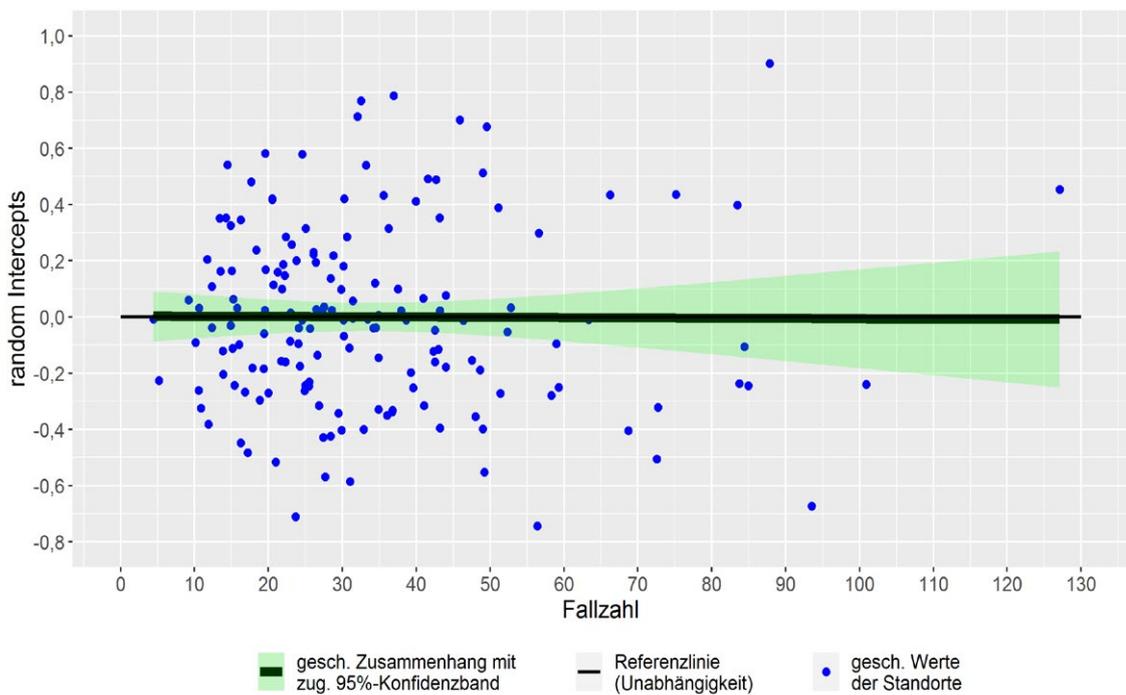


Abbildung 9: Geschätzte PNZ-spezifische random intercepts als Funktion von der Fallzahl.

4.3 Analyse basierend auf den Fallzahlen analog zur Mindestmengenregelung

Wie in Abschnitt 4.1.1 beschrieben, wird im Folgenden zusätzlich die Volume-Outcome-Analyse betrachtet, bei der die relevanten Fallzahlen analog zur bestehenden Mindestmengenregelung definiert werden. Abbildung 10 zeigt die geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit dieser Fallzahlen.

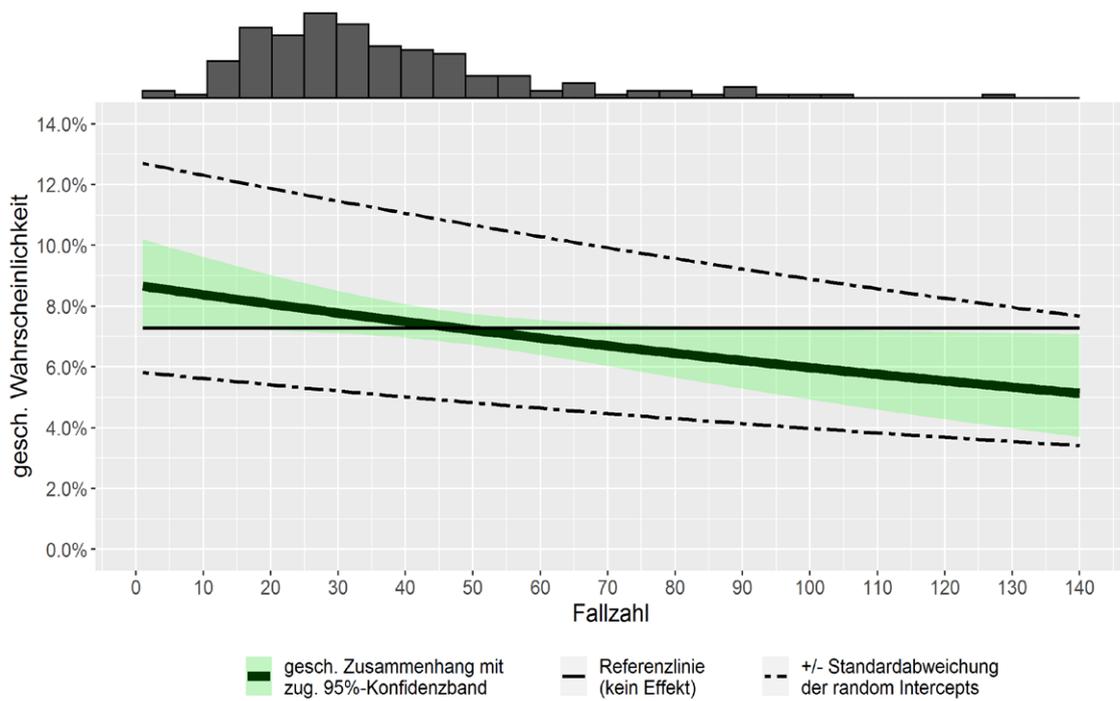


Abbildung 10: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl analog zur Mindestmengenregelung.

Die statistischen Kenngrößen basierend auf dieser Analyse finden sich in der folgenden Tabelle 6.

Tabelle 6: Statistische Kennzahlen des Modellfits basierend auf den Fallzahlen analog zur bestehenden Mindestmengenregelung.

Größe	Wert
p-Wert zum Volume-Outcome-Zusammenhang	0,024
AUC zum Volume-Outcome-Modell	in-sample: 0,831; out-of-sample: 0,821
Geschätzte Varianz der random intercepts	0,183, KI: [0,126 , 0,266]
MOR der random intercepts	0,665, KI: [0,611 , 0,712]

Die auf Basis der an der Mindestmengenregelung orientierten Fallzahl geschätzte Volume-Outcome-Beziehung ist qualitativ ähnlich, jedoch etwas schwächer ausgeprägt, sowohl was die Stärke des geschätzten Effekts angeht als auch die statistische Signifikanz – nichtsdestotrotz ist der p-Wert eher klein. Zur Einordnung dieses Ergebnisses soll an die Argumentation dafür erinnert werden, dass die kumulierten durchschnittlichen Fallzahlen eine stabilere, genauere Abbildung des gesuchten Volume-Outcome-Mechanismus ermöglichen (vgl. Abschnitt 4.1.1). Dies macht umgekehrt den hier beobachteten leicht abgeschwächten Effekt plausibel.

4.4 Sensitivitätsanalysen

4.4.1 Analyse basierend auf den QS-Daten

Die NICU-Daten enthalten auch die Fälle aus Level-2-PNZ, die QS-Daten (ebenfalls 2014 bis 2018) enthalten die Fälle für alle Versorgungsstufen. Da eine Verknüpfung der beiden Datensätze im Rahmen dieser Beauftragung nicht möglich ist, soll zunächst überprüft werden, ob die Hauptanalyse erwartungsgemäß zu einem vergleichbaren Ergebnis führt, wenn anstatt der NICU-Daten die QS-Daten verwendet werden. Dabei wird zur Sicherstellung der Vergleichbarkeit ebenso das im Abschnitt 4.1.4 vorgestellte Modell (2) verwendet. Es werden die gleichen Ein- bzw. Ausschlussbedingungen angewandt. Für die Einschränkung auf die Level-1-PNZ ist jedoch zu beachten, dass in den QS-Daten keine eindeutige Information über die Versorgungsstufe vorliegt. Es wird im Folgenden angenommen, dass die tatsächliche Versorgungsstufe jeweils der in den fallbezogenen QS-Daten am häufigsten genannten Versorgungsstufe entspricht. Abbildung 11 zeigt die geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit der kumulierten durchschnittlichen Fallzahl basierend auf den QS-Daten.

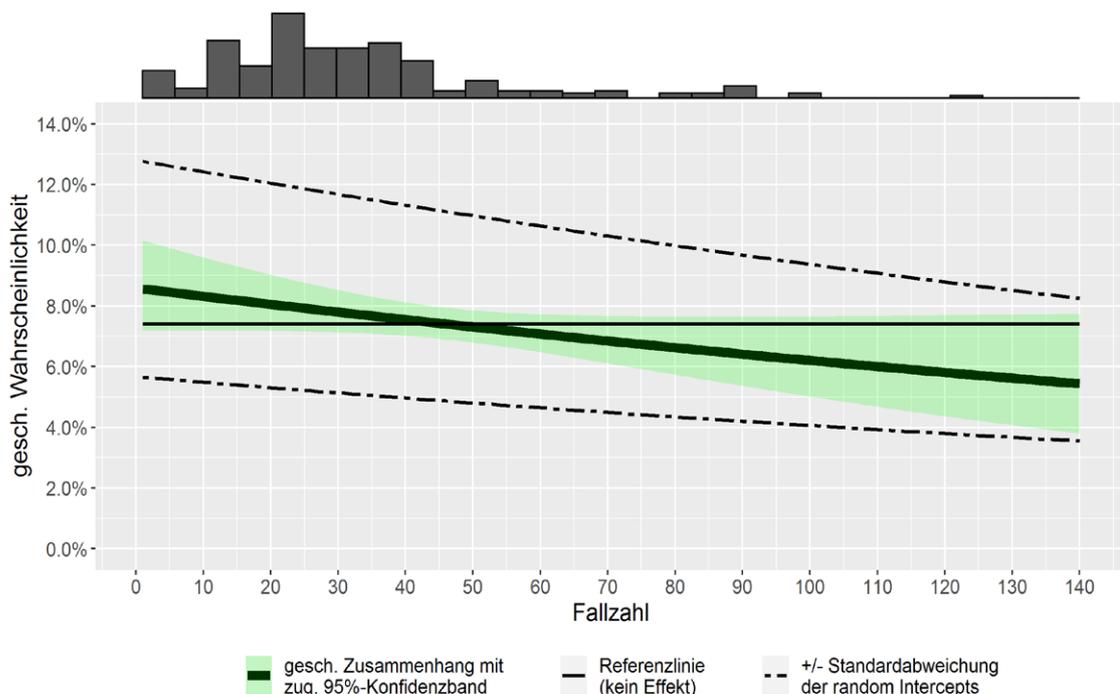


Abbildung 11: Die geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl. Datenbasis: QS-Daten, nur Level-1-PNZ.

Die statistischen Kenngrößen basierend auf dieser Analyse finden sich in Tabelle 7.

Tabelle 7: Statistische Kennzahlen des Modellfits. Datenbasis: QS-Daten, nur Level-1-PNZ.

Größe	Wert
p-Wert zum Volume-Outcome-Zusammenhang	0,071
AUC zum Volume-Outcome-Modell	in-sample: 0,832; out-of-sample: 0,82

Größe	Wert
Geschätzte Varianz der random intercepts	0,2, KI: [0,138 , 0,289]
MOR der random intercepts	0,653, KI: [0,599 , 0,702]

Auch diese Ergebnisse sind konsistent mit der obigen Hauptanalyse. Aufgrund der beschriebenen Limitationen beim QS-Datensatz ist die geringere beobachtete Stärke des Volume-Outcome-Zusammenhangs plausibel.

4.4.2 Weglassen des größten PNZ

Das PNZ mit der größten Fallzahl lässt sich hinsichtlich dieser Fallzahl sowie (in geringerem Maße) hinsichtlich der geschätzten random intercepts als Ausreißer interpretieren. Daher liegt es nahe, das größte PNZ im Rahmen einer Sensitivitätsanalyse herauszulassen, um seine Bedeutung für die zentrale Volume-Outcome-Analyse in Abschnitt 4.2 zu überprüfen. Die folgende Abbildung 12 zeigt die entsprechende geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit der Fallzahl.

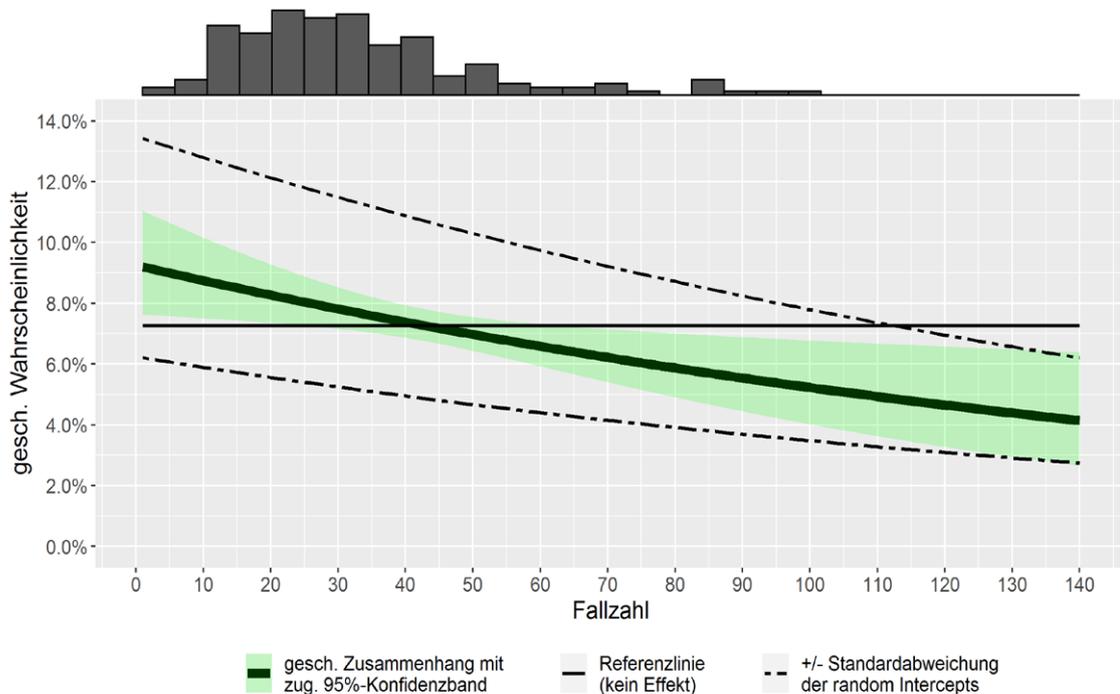


Abbildung 12: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl. Datenbasis: NICU-Daten, nur Level-1-PNZ, weglassen des größten PNZ.

Die statistischen Kenngrößen basierend auf dieser Ausreißeranalyse finden sich in Tabelle 8.

Tabelle 8: Statistische Kennzahlen des Modellfits. Datenbasis: NICU-Daten, nur Level-1-PNZ, weglassen des größten PNZ.

Größe	Wert
p-Wert zum Volume-Outcome-Zusammenhang	0,007
AUC zum Volume-Outcome-Modell	in-sample: 0,832; out-of-sample: 0,821
Geschätzte Varianz der random intercepts	0,181, KI: [0,125 , 0,264]
MOR der random intercepts	0,666, KI: [0,613 , 0,714]

Das Ergebnis ist grundsätzlich konsistent mit den Hauptergebnissen aus Abschnitt 4.2, welche weiterhin auf der vollständigen Grundgesamtheit basieren. Der kleinere p-Wert ist angesichts der random intercepts plausibel, aber insbesondere auch aufgrund der Tatsache, dass das größte, hier lediglich zur Überprüfung ignorierte, PNZ eine überdurchschnittliche Sterblichkeit aufweist.

4.4.3 Betrachtungen ohne Einschränkung auf Level-1-PNZ

Schließlich sollen gemäß Beauftragung die Analysen auch ohne Einschränkungen auf die Level-1-PNZ durchgeführt werden. Dies wird sowohl für die NICU- als auch für die QS-Daten betrachtet. Auch wenn dem Auftragstext nicht wörtlich zu entnehmen ist, welche Stufen der perinatalogischen Versorgung in den Analysen zu berücksichtigen sind, erklärt sich dies aus dem Kontext. Eine Beschränkung der Analysen nur auf Level-1-PNZ in allen Analysen wäre im Auftrag zu erwähnen gewesen. Die explizite Beauftragung die QS-Daten für die Analysen zu nutzen weist demgegenüber darauf hin, auch alle anderen Versorgungsstufen in gesonderten Analysen zu berücksichtigen.

Die folgende Abbildung 13 zeigt die entsprechende geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit der Fallzahl basierend auf den NICU-Daten.

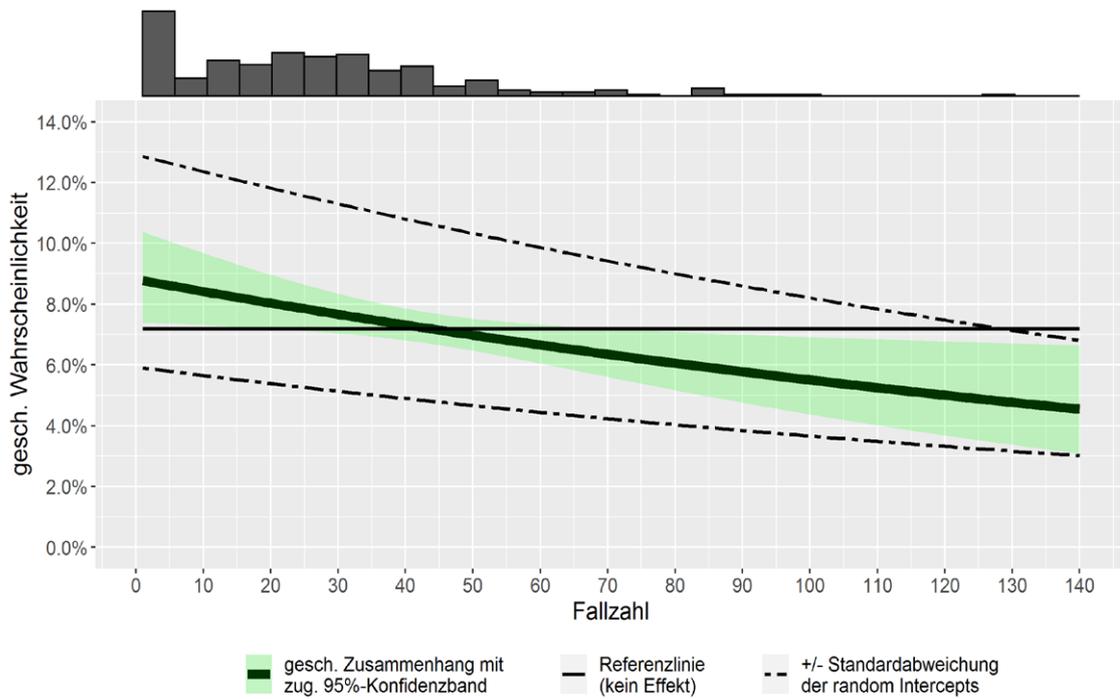


Abbildung 13: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl. Datengrundlage: NICU-Daten, beide Versorgungsstufen.

Die statistischen Kenngrößen basierend auf dieser Analyse für die NICU-Daten finden sich in der folgenden Tabelle 9.

Tabelle 9: Statistische Kennzahlen des Modellfits. Datengrundlage: NICU-Daten, beide Versorgungsstufen.

Größe	Wert
p-Wert zum Volume-Outcome-Zusammenhang	0,012
AUC zum Volume-Outcome-Modell	in-sample: 0,832; out-of-sample: 0,821
Geschätzte Varianz der random intercepts	0,183, KI: [0,127 , 0,265]
MOR der random intercepts	0,665, KI: [0,612 , 0,712]

Die folgende Abbildung 14 zeigt die entsprechende geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit der Fallzahl basierend auf den QS-Daten.

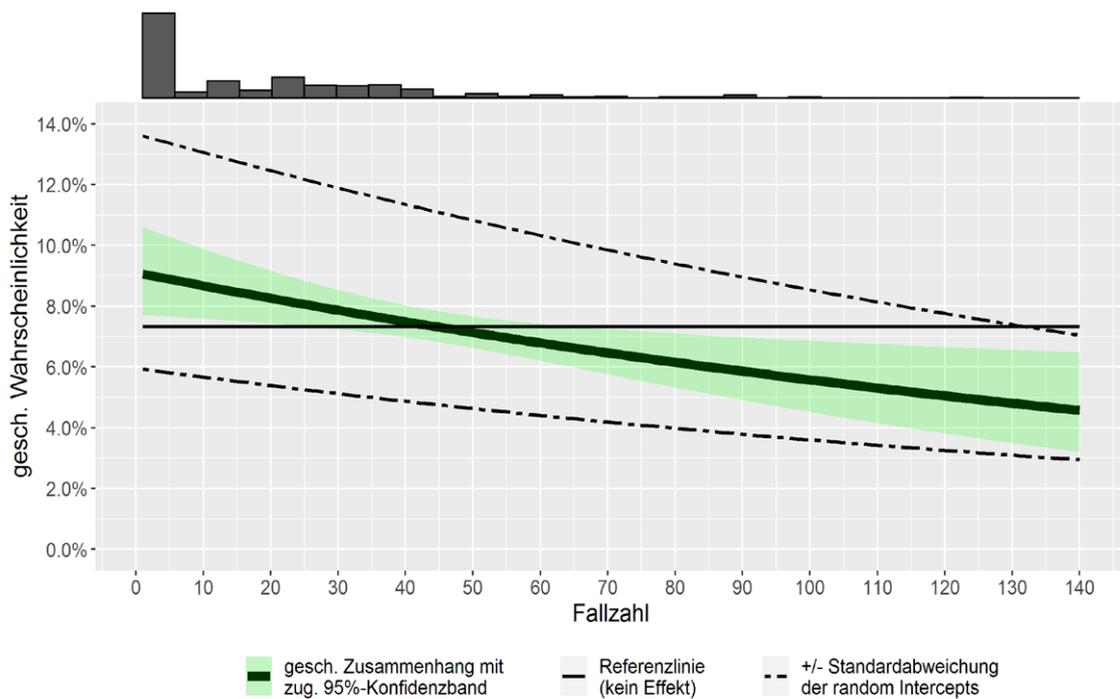


Abbildung 14: Geschätzte Sterbewahrscheinlichkeit für einen durchschnittlichen Fall in Abhängigkeit von der Fallzahl. Datengrundlage: QS-Daten, alle Versorgungsstufen.

Die statistischen Kenngrößen basierend auf dieser Analyse für die QS-Daten finden sich in der folgenden Tabelle 10.

Tabelle 10: Statistische Kennzahlen des Modellfits. Datengrundlage: QS-Daten, alle Versorgungsstufen.

Größe	Wert
p-Wert zum Volume-Outcome-Zusammenhang	0,004
AUC zum Volume-Outcome-Modell	in-sample: 0,832; out-of-sample: 0,819
Geschätzte Varianz der random intercepts	0,21, KI: [0,145 , 0,303]
MOR der random intercepts	0,646, KI: [0,591 , 0,695]

Auch diese Ergebnisse sind konsistent mit den Ergebnissen der Hauptanalyse. Insbesondere wird auch hier der Fallzahleffekt monoton fallend und (auf Logitebene) nahezu linear geschätzt.

Erwähnenswert ist hier der wesentlich stärkere Zusammenhang bei den QS-Daten im Vergleich zur Analyse basierend auf den Level-1-PNZ. Eine naheliegende Erklärung hierfür besteht darin, dass der Datensatz nun zum Beispiel im Bereich der kleinen Fallzahlen viele Notfälle enthält, die außerhalb von PNZ behandelt werden und entsprechend mit schlechteren Behandlungsergebnissen einhergehen – nicht zuletzt, da diese Standorte im Allgemeinen nicht auf die Behandlung von VLBW-I ausgelegt sind. Aus dieser Perspektive erscheint der starke Volume-Outcome-Zusammenhang plausibel. Insbesondere ist es auch plausibel, dass der Unterschied beim NICU-Datensatz (d. h. die Ergebnisse basierend auf den Level-1-PNZ, Abbildung 6, im Vergleich zu den Ergebnissen basierend auf beiden Levels, Abbildung 13) geringer ausfällt, da dieser Datensatz

nur Fälle aus PNZ enthält und nicht, so wie die QS-Daten, auch Fälle aus Standorten mit Versorgungsstufe III oder IV.

5 Suche nach möglichen Schwellenwerten mit einem Bruchpunkt-Modell

Im Kontext der Volume-Outcome-Beziehung sind zwei Interpretationen für das Wort Schwellenwert denkbar:

1. Unter dem Schwellenwert kann man eine Fallzahl verstehen, ab der das Sterberisiko für ein Outcome sich nicht mehr relevant und/oder signifikant verringert (Ulm 1991). Diese Fallzahl entspricht formal einem Strukturbruch in der Volume-Outcome-Beziehung.
2. Alternativ kann man unter einem Schwellenwert eine Fallzahl verstehen, ab der das Sterberisiko eine gewisse Schwelle unterschreitet (Bender 1999). Diese Schwelle entspricht einem maximalen tolerierbaren Risiko. Diese Schwelle kann dabei entweder auf Ebene der Wahrscheinlichkeiten festgelegt werden (Bender 1999) oder auf der logistischen Ebene der Prädiktoren (Budtz-Jørgensen et al. 2001).

Es gibt derzeit keine fachliche Festlegung eines maximal tolerierbaren Risikos, das zur Bestimmung eines möglichen Schwellenwerts hätte herangezogen werden können. Oft lassen sich solche Werte auch nicht im vornherein exakt angeben, sondern lediglich als Orientierungswert (IQWiG 2005), sodass auch der Schwellenwert durch solche Vorgaben nicht genau bestimmt werden kann, sondern gesetzt werden muss.

Daher wird in diesem Abschnitt der erste Ansatz verfolgt, in der Volume-Outcome-Beziehung einen Strukturbruch zu suchen. Ein derartiger Strukturbruch ist allerdings in der geschätzten Volume-Outcome-Beziehung nicht direkt erkennbar. Das geschätzte (bedingte) Risiko sinkt über den gesamten Fallzahlbereich kontinuierlich ab. Dabei muss natürlich die Schätzunsicherheit berücksichtigt werden: Innerhalb der Schätzungenauigkeit ist es durchaus möglich, dass es in der Volume-Outcome-Beziehung einen solchen Strukturbruch gibt.

Um zu untersuchen, ob die NICU-Daten statistische Evidenz auf einen möglichen Schwellenwert in der Volume-Outcome-Beziehung enthalten, wurde ein Bruchpunkt-Modell gefittet, welches sich an der Methodik von Ulm (1991) orientiert (vgl. auch Grouven et al. (2008)). Das Modell hat die folgende Gestalt:

$$\text{logit}(\pi_{ij}^k) = \beta_0^k + \eta_{ij}^k + \beta_M(M - \tilde{n}_i^k)_+ + u_i.$$

Hierbei ist M der Bruchpunkt-Parameter, β_M ein Regressionskoeffizient und

$$(M - \tilde{n}_i^k)_+ = \begin{cases} M - \tilde{n}_i^k & , \text{ falls } M > \tilde{n}_i^k, \\ 0 & , \text{ sonst.} \end{cases}$$

alle anderen Modellkomponenten tragen dieselbe Bedeutung wie in Abschnitt 4.1.4 Während also im Volume-Outcome-Modell der Hauptanalyse die Fallzahlabhängigkeit durch eine glatte Funktion $f_{FZ}(\tilde{n}_i^k)$ modelliert wurde, wird hier der Zusammenhang so modelliert, dass er linear ist bis zum Bruchpunkt M und ab dem Bruchpunkt konstant fortgesetzt wird.

Der Parameter M kann durch Maximierung der (marginalen) Likelihood geschätzt werden. Hält man den Wert von M konstant, so hat das Bruchpunkt-Modell die Gestalt eines gewöhnlichen generalisierten additiven gemischten Modells, sodass sich für die Parameter durch Standardmethoden Maximum-Likelihood-Schätzer (ML-Schätzer) gewinnen lassen. Durch Vergleich der Werte der Likelihood für verschiedene Werte von M erhält man dann den ML-Schätzer \hat{M} für M . Für die vorliegende Analyse ist es dabei ausreichend, sich auf die ganzzahligen möglichen Werte von M in dem Bereich zu beschränken, in dem auch die beobachteten Fallzahlen liegen. Ein asymptotisches Konfidenzintervall erhält man durch die Likelihood-Quotienten-Methode (Ulm 1991).

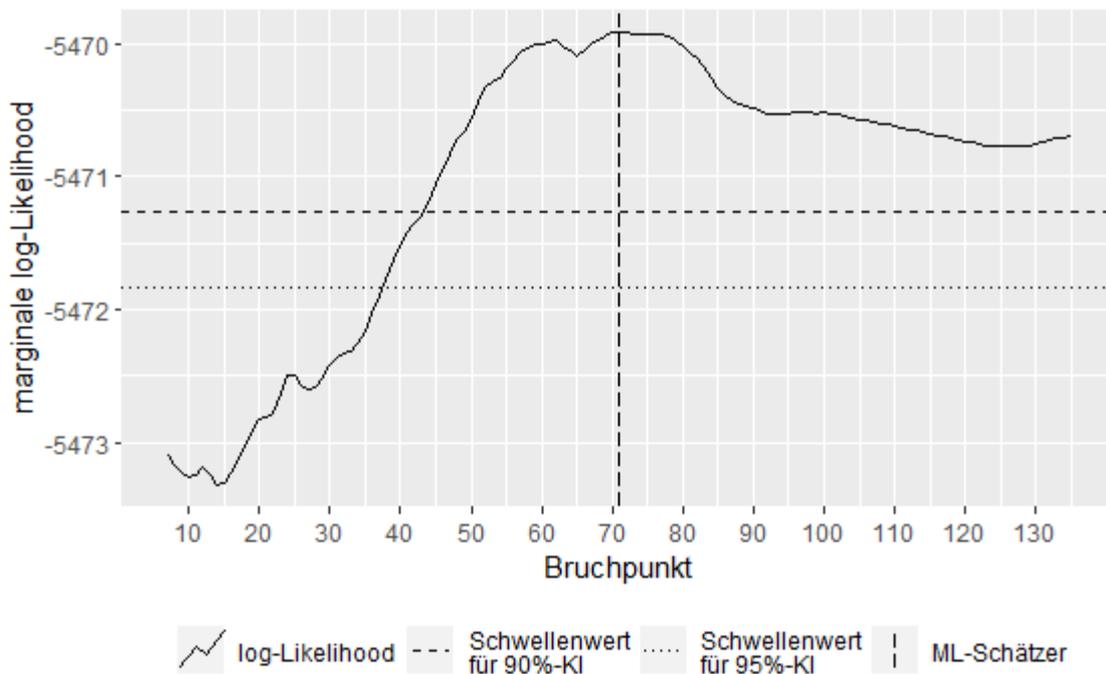


Abbildung 15: Marginale log-Likelihood als Funktion des Bruchpunkts M . Die horizontalen gepunkteten Linien markieren die Schwellenwerte an die log-Likelihood mit Schwellenwerten für die Likelihood-Quotienten-Konfidenzintervalle.

Auf den NICU-Daten wurde für den ML-Schätzer ein Wert von $\hat{M} = 71$ ermittelt. Die untere Schranke des 95 %-Konfidenzintervalls liegt bei 38 (die des 90 %-Konfidenzintervalls liegt bei 44). Die obere Schranke liegt jeweils beim Maximalwert. Dies bedeutet, dass ein Likelihood-Quotienten-Test die Nullhypothese, dass es keinen Bruchpunkt gibt, nicht ablehnt (weder bei einem Signifikanzniveau von $\alpha = 5\%$ noch bei einem Signifikanzniveau von $\alpha = 10\%$). In diesem Sinne ist der Bruchpunkt bei $\hat{M} = 71$ nicht signifikant.

6 Analyse möglicher Schwellenwerte: Simulation der Einführung einer Mindestmenge und Berechnung absoluter Effektmaße

Wie in Abschnitt 4.2.1 erläutert, enthält die Funktion f_{FZ} den reinen Zusammenhang zwischen Sterberisiko und Fallzahl des PNZ. Bei der Interpretation muss beachtet werden, dass die patientenseitigen Risiken mit der Fallzahl korrelieren. So ist das modellbasiert anhand der bekannten patientenseitigen Risiken vorhergesagte durchschnittliche Sterberisiko für Frühgeborene in PNZ mit hoher Fallzahl merklich höher als in PNZ mit geringer Fallzahl. Bewertet man die Volume-Outcome-Beziehung daher lediglich anhand der Sterbewahrscheinlichkeit für durchschnittliche Frühgeborene (Abbildung 6) oder auf der Basis von relativen Effektmaßen wie Odds-Ratios, so wird der Effekt gegebenenfalls falsch eingeschätzt.

Darüber hinaus sind Angaben von Odds-Ratios, SMR oder Funktionen von Änderungen von Sterbewahrscheinlichkeiten, welche sich auf VLBW-I mit durchschnittlichem Risiko beziehen, wenig aussagekräftig oder anschaulich, wenn eingeschätzt werden soll, wie sich die Einführung einer Mindestmenge gemäß dem statistischen Modell potentiell auswirken würde.

Vor diesem Hintergrund der Betrachtung von Risikofaktoren oder Interventionen in Populationen werden in den Bereichen Public Health und Epidemiologie seit langem die Berechnung von absoluten Effektmaßen wie Risikodifferenzen, attributable Risiken oder auch number needed to treat (NNT) gefordert (Bonita et al. 2006, Rothman et al. 2008). Im Bereich der Perinatalogie wurde von Heller (2009), Heller (2018), Heller et al. (2017) und Heller et al. (2003) der Begriff „potentiell vermeidbare Todesfälle“ oder „potentiell vermeidbare Ereignisse“ genutzt.

Daher enthält dieses Kapitel eine alternative Darstellung der geschätzten Volume-Outcome-Beziehung, welche die Risikoprofile der PNZ berücksichtigt. Dazu werden ähnlich wie in Heller (2009) und Heller (2018) für verschiedene Schwellenwerte die hypothetischen Folgen simuliert, wenn Fälle aus PNZ mit kleiner Fallzahl auf PNZ mit größerer Fallzahl umverteilt werden. Dazu werden die Fälle des Jahres 2018 betrachtet. Auf diesen Fällen werden die hypothetischen Folgen anhand der Volume-Outcome-Beziehung simuliert. Da die Volume-Outcome-Beziehung aus den Fällen aus allen fünf Jahren gewonnen wurde, liegen der Abschätzung der Effektmaße in diesem Kapitel letztlich die gesamten betrachteten NICU-Daten zugrunde. Die Folgen werden jedoch für ein einzelnes Jahr betrachtet, wobei die Fälle des aktuellsten verfügbaren Jahres in den betrachteten NICU-Daten als repräsentativ erachtet werden. Als Fallzahl wird die kumulierte durchschnittliche Gesamtfallzahl zugrunde gelegt (vgl. Abschnitt 4.1.1), also (da Fälle aus dem Datenjahr 2018 betrachtet werden) die durchschnittliche Fallzahl der Jahre 2012 bis 2018.

Für verschiedene mögliche Werte M für den Schwellenwert wird betrachtet:

- Wie viele PNZ haben eine Fallzahl kleiner als M ?
- Wie viele Fälle wurden in solchen PNZ behandelt?
- Wie verringert sich die Anzahl an Sterbefällen mit der Umverteilung?

6.1 Limitationen der Betrachtungen

Wie oben erwähnt, handelt es sich bei den folgenden Betrachtungen nicht um eine konkrete Vorhersage der Folgen, die eine Einführung oder Erhöhung der Mindestmenge hätte. Die Betrachtungen dienen vielmehr zur Illustration der geschätzten Volume-Outcome-Beziehung unter Berücksichtigung der Verteilung der Risikoprofile über PNZ mit kleiner oder großer Fallzahl. Zusätzlich zu den Limitationen der in diesem Bericht auf der Basis von Beobachtungsdaten durchgeführten Analysen (vgl. Kapitel 7) sollte bei der Interpretation dieser im Folgenden dargestellten Zahlen beachtet werden:

1. Der Betrachtung liegen die Fälle des Jahres 2018 zugrunde. Es wird also abgeschätzt, wie sich die hypothetische Umverteilung auf die Anzahl an Sterbefällen auswirkt, wenn ähnliche Fälle wie im Jahr 2018 auftreten.
2. Die folgenden Ergebnisse hängen davon ab, wie die Umverteilung modelliert wird. Durch die Umverteilung verändert sich die Fallzahlverteilung der PNZ und die Zuordnung der Fälle zu den PNZ. In der vorliegenden Analyse werden dafür zwei verschiedene Szenarien betrachtet.

6.2 Betroffene PNZ und umverteilte Fälle

Zunächst sollen die von einer Umverteilung betroffenen PNZ und Fälle in Abhängigkeit des Schwellenwerts M analysiert werden. Zur Überprüfung, ob die Fälle eines PNZ umverteilt werden sollen, wird dabei die kumulierte durchschnittliche Fallzahl verwendet (vgl. Abschnitt 4.1.1), also (da Fälle aus dem Datenjahr 2018 betrachtet werden) die durchschnittliche Gesamtfallzahl der Jahre 2012 bis 2018.

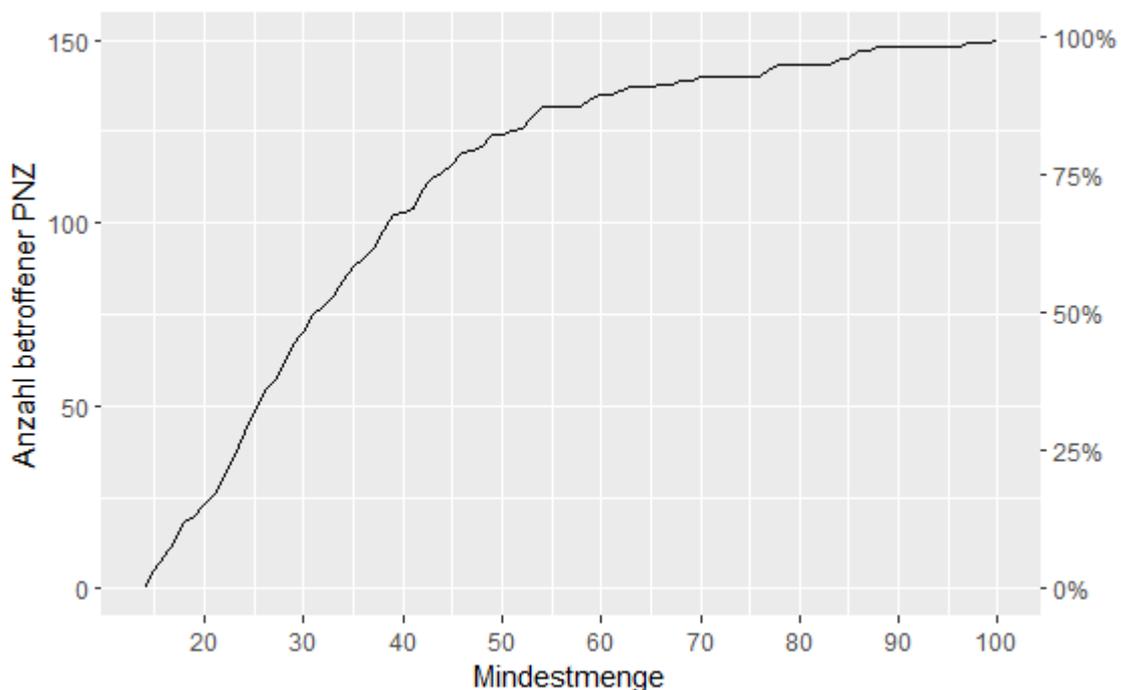


Abbildung 16: Die Anzahl von betroffenen PNZ als Funktion der möglichen Mindestmenge M .

Bei der Interpretation der Zahl der betroffenen PNZ ist zu beachten, dass diese Zahl nicht gleich der Anzahl an PNZ ist, die bei Erhöhung einer Mindestmenge aus der Versorgung von VLBW-I ausgeschlossen würden, weil gemäß der aktuellen Mindestmengenregelung Übergangsfristen vorgesehen sind.⁵ Demnach könnten Einrichtungen (knapp) unterhalb der Mindestmenge diese im Folgejahr (in den Folgejahren) dennoch erreichen, weil sie zusätzlich VLBW-I behandeln, die sonst in nunmehr geschlossenen Abteilungen behandelt worden wären.

Wie bei Heller (2009) und Heller (2018) sollen im Folgenden allerdings lediglich die Anzahl der primär von einer potentiellen Mindestmenge betroffenen PNZ wie auch die Anzahl der primär potentiell umzuverteilenden Fälle bestimmt werden, also die Gesamtanzahl an Fällen, die 2018 in einem betroffenen PNZ unter einer definierten Mindestmenge behandelt wurden.

Der Unterschied zu den Analysen von Heller (Heller 2009, Heller 2018) besteht darin, dass dort Gesamtfallzahlen einer einjährigen Analyseperiode vor der simulierten Einführung einer Mindestmenge mit den potentiellen Effekten in der Periode nach der Einführung einer Mindestmenge berechnet wurden. Hier wird dagegen die durchschnittliche Fallzahl der Jahre 2012 bis 2018 genutzt, um simulierte Mindestmengen im Jahr 2018 zu definieren und deren Effekt im Jahr 2018 zu schätzen.

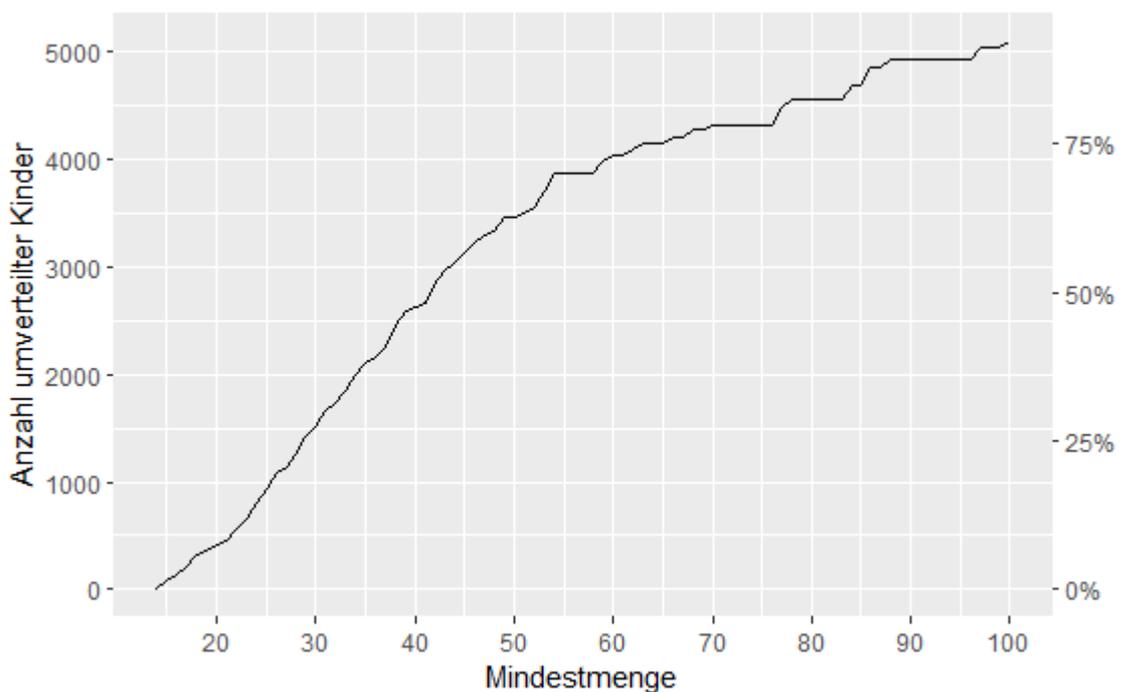


Abbildung 17: Anzahl von umverteilten Kindern als Funktion des Schwellenwerts M .

Was für die potentielle Umverteilung der Perinatalzentren gilt, gilt in gleicher Weise für die Zahl der umverteilten Fälle: Nicht jeder dieser Fälle müsste bei Erhöhung der Mindestmenge tatsächlich umverteilt werden. Wie bereits oben erwähnt wurde, wird nicht jeder betroffene PNZ, der

⁵ Regelungen des Gemeinsamen Bundesausschusses gemäß § 136b Absatz 1 Satz 1 Nummer 2 SGB V für nach § 108 SGB V zugelassene Krankenhäuser (Mindestmengenregelung, Mm-R). In der 1. Neufassung vom 21. März 2006, zuletzt geändert am 27. März 2020, in Kraft getreten am 27. März 2020. URL: <https://www.g-ba.de/richtlinien/5/> (zuletzt abgerufen am: 12.05.2020).

den Schwellenwert nicht erfüllt, zwangsläufig geschlossen. Daneben kann sich ein Level-1-PNZ, das die Mindestmenge nicht erfüllt, dazu entscheiden, als Level-2-PNZ weiterzuarbeiten. Frühgeborene mit unter 1.250 g könnten dann weiterhin als Notfälle behandelt werden. Beispielsweise multipliziert Heller (2009), von einem 10 %igen Anteil der Notfallversorgung bei VLBW ausgehend, die Anzahl der Umverteilungen pauschal mit 90 %.

6.3 Betrachtete Maßzahlen zur Auswirkungen auf die Sterblichkeit

Schließlich soll abgeschätzt werden, wie sich die Umverteilung auf die Sterblichkeit auswirkt. Dazu muss man modellieren, was mit den Fällen passiert, die umverteilt werden sollen. Diese Modellierung wird im Abschnitt 6.4 genauer beschrieben.

Die Verringerung der Sterblichkeit drückt sich dadurch aus, dass nach der Umverteilung die erwartete Gesamtzahl an verstorbenen Kindern kleiner ist als vorher. Die Differenz lässt sich als *Übersterblichkeit* (excess mortality) oder als *Anzahl an potentiell vermeidbaren Sterbefällen (oder Ereignissen)* vor der Umverteilung (Heller 2009, Heller 2018, Heller et al. 2017, Heller et al. 2002) auffassen. Das Verhältnis dieser Differenz der erwarteten absoluten Anzahlen zur Anzahl der umverteilten Frühgeborenen ergibt die *Risikoreduktion für die umverteilten Fälle*. Das Verhältnis dieser Differenz der absoluten Anzahlen zur Gesamtanzahl von VLBW-I ergibt die *durchschnittliche Risikoreduktion*. Zur Interpretation der Risikoreduktion kann diese in Beziehung zur absoluten Sterblichkeit gesetzt werden. Diese lag für die Grundgesamtheit innerhalb des betrachteten Zeitraumes bei 7,26 %. Formeln für diese Maßzahlen befinden sich in Abschnitt 6.4.

Die durchschnittliche Risikoreduktion lässt sich auch als geschätzte Differenz des durchschnittlichen Sterberisikos vor und nach Umverteilung interpretieren. Sie entspricht damit dem in Grouven et al. (2008) vorgeschlagenen populationsbasierten Effektmaß δ .

Der Kehrwert der Risikoreduktion für die Umverteilungen entspricht der *Anzahl der notwendigen Behandlungen* (NNT) aus der Epidemiologie: Wie viele Frühgeborene müssen umverteilt werden, damit im Mittel ein Sterbefall verhindert wird? Der Kehrwert der mittleren Risikoreduktion ist die *impact number* (IN): Unter wie vielen Frühgeborenen insgesamt (mit und ohne Umverteilung) wird durch die Umverteilung im Mittel ein Sterbefall verhindert?

6.4 Modellierung der Umverteilungen

Um den Effekt der Umverteilungen zu simulieren wird das Risiko basierend auf dem geschätzten Volume-Outcome-Zusammenhang vor der Umverteilung mit dem entsprechenden Risiko nach der Umverteilung verglichen. Für die Modellierung der Umverteilung basierend auf dem geschätzten Volume-Outcome-Zusammenhang sind von den Parametern in Gleichung (2) lediglich zwei relevant, nämlich die Fallzahl und die random intercepts des PNZ, in das der Fall umverteilt wird. Die patientenseitigen Risikofaktoren und das Jahr ändern sich durch eine Umverteilung nicht.

Um den reinen Effekt der Fallzahländerung durch die Umverteilungen zu schätzen ist es naheliegend, den random intercept des PNZ bei der simulierten Umverteilung konstant zu halten.

Jeder umverteilte Fall wird also in einen PNZ umverteilt, der den gleichen Wert des random Intercepts besitzt wie der ursprüngliche PNZ des Falles (vgl. die Formeln weiter unten).

Des Weiteren ist die Frage, wie sich durch die Änderung einer Mindestmenge die Größenverteilung der Level-1-PNZ auswirkt. Wie bereits oben erwähnt, wird die Änderung einer Mindestmenge dazu führen, dass manche Level-1-PNZ geschlossen oder auf Level-2-PNZ zurückgestuft werden, während andere Level-1-PNZ wachsen. Aus den vorhandenen Daten lässt sich nicht vorhersagen, wie sich dies auf die Volume-Outcome-Beziehung auswirken wird. Beispielsweise macht die Volume-Outcome-Beziehung, wenn man sie naiv kausal interpretiert, die Vorhersage, dass sich die Behandlungsqualität in PNZ, die aufgrund von Umverteilungen wachsen, bessert. Es ist jedoch nicht klar, in welchem Zeitraum dieser positive Effekt zu erwarten ist bzw. ob dieser Effekt überhaupt eintritt. Daher wird angenommen, dass sich bei den Fällen in den PNZ, deren Fallzahl über dem Schwellenwert liegt, durch die Umverteilungen am Outcome nichts ändert. Für die umverteilten Fälle muss jedoch für die Formel der Volume-Outcome-Beziehung eine Fallzahl nach der Umverteilung angegeben werden.

In diesem Bericht werden zwei Szenarien für die Fallzahl der PNZ, in den die Umverteilungen jeweils stattfinden, betrachtet:

1. In Szenario 1 erfolgt die Umverteilung der betroffenen Fälle in PNZ, deren Fallzahl genau dem Schwellenwert entspricht. Dies ist die kleinste sinnvolle Fallzahl. Aufgrund des monotonen Zusammenhangs der Volume-Outcome-Beziehung entspricht dieses Szenario bei gegebenem Schwellenwert einer unteren Schranke an den vorhergesagten Effekt.
2. In Szenario 2 wird der Median aller Fallzahlen derjenigen PNZ verwendet, die über dem Schwellenwert liegen. Dies entspricht einer Umverteilung in einen PNZ durchschnittlicher Größe (wobei die durch die Änderungen in der Fallzahlstatistik aufgrund der Umverteilungen nicht berücksichtigt werden). Der Median der Fallzahl wird dabei auf Fallebene gebildet; d. h. es wird der Median der Fallzahlen der PNZ über alle Frühgeborenen gebildet, die nicht umverteilt werden. Größere PNZ (mit dementsprechend mehr Frühgeborenen) haben so ein etwas größeres Gewicht als beim Median der Fallzahlen der PNZ auf Ebene der PNZ.

Zusammengefasst wird der Effekt der Umverteilungen also auf Ebene des Einzelfalles wie folgt modelliert: Vor der Umverteilung ist die geschätzte Wahrscheinlichkeit $\hat{\pi}_{ij}$ für das Versterben des j ten Kindes in PNZ i nach der Gleichung aus Abschnitt 4.1.4 gegeben durch

$$\text{logit}(\hat{\pi}_{ij}^{2018}) = \hat{\beta}_0^{2018} + \hat{\eta}_{ij}^{2018} + \hat{f}_{FZ}(\tilde{n}_i^{2018}) + \hat{u}_i.$$

Durch die Umverteilung verändert sich die geschätzte Wahrscheinlichkeit für die Fälle in PNZ mit $\tilde{n}_i^{2018} \leq M$ zu $\hat{\pi}'_{ij}$, wobei

$$\text{logit}(\hat{\pi}'_{ij}^{2018}) = \hat{\beta}_0^{2018} + \hat{\eta}_{ij}^{2018} + \hat{f}_{FZ}(\tilde{n}'_i^{2018}) + \hat{u}_i,$$

wobei \tilde{n}'_i^{2018} vom Szenario abhängt. Aufgrund der Annahme, dass sich der Wert des random Intercepts des PNZ bei der Umverteilung nicht ändert, geht in $\hat{\pi}_{ij}$ und $\hat{\pi}'_{ij}$ der selbe Wert \hat{u}_i ein. Für Fälle, die nicht umverteilt werden, gilt in jedem Szenario $\tilde{n}'_i^{2018} = \tilde{n}_i^{2018}$.

Die geschätzte Anzahl an Sterbefällen, die durch die Umverteilung mit Schwellenwert M verhindert werden könnte, ist dann gegeben durch

$$r(M) = \sum_{i=1}^I \sum_{j=1}^{n_i} (\hat{\pi}_{ij}^{2018} - \hat{\pi}'_{ij}{}^{2018}),$$

wobei zur Vereinfachung der Notation $r(M)$ und die folgenden Maßzahlen nicht explizit als Schätzungen gekennzeichnet sind, da diese nur als solche vorkommen. In dieser Summe spielen nur die umverteilten Fälle eine Rolle. Dies ist konsistent mit der Annahme, dass sich für die Fälle in PNZ, deren Fallzahl über dem Schwellenwert liegt, das Outcome nicht ändert.

Die Risikoreduktion für die umverteilten Fälle ist dann

$$\delta_u(M) = \frac{r(M)}{n_u(M)}, \quad \text{wobei } n_u(M) = \sum_{i:n_i < M} n_i,$$

und die durchschnittliche Risikoreduktion ist

$$\delta_d(M) = \frac{r(M)}{n}, \quad \text{wobei } n = \sum_{i=1}^I n_i.$$

6.5 Auswirkung der Umverteilungen auf die Sterblichkeit

Abbildung 18, Abbildung 19 und Abbildung 20 visualisieren die oben eingeführten Maßzahlen. Die numerischen Werte folgen in den Tabelle 11 und Tabelle 12. Zur Interpretation der Risikoreduktionen sei wiederholt, dass die durchschnittliche Sterblichkeit im betrachteten Zeitraum (2014 bis 2018) innerhalb der Grundgesamtheit bei 7,26 % lag.

In Tabelle 11 und Tabelle 12 ist in Klammern der geschätzte Standardfehler der Maßzahlen angegeben. Der Standardfehler ist dabei mittels der Deltamethode aus der in Abschnitt 4.2 beschriebenen Schätzunsicherheit des Modellfits des Volume-Outcome-Modells berechnet (siehe z. B. Agresti (2013)). Der Standardfehler quantifiziert also die Unsicherheit der Maßzahlen, die sich aus der Unsicherheit der Modellschätzung ergibt, während die Prädiktionsunsicherheit nicht berücksichtigt wird (vgl. Abschnitt 6.1). Die gestrichelten Linien in den Abbildungen entsprechen Wald-Konfidenzintervallen zum Konfidenzniveau 95 %.

Dabei zeigt sich analog zu den obigen Analysen in Kapitel 4 eine monoton steigende Anzahl von potentiell vermeidbaren Sterbefällen mit dem gewählten Schwellenwert. Im Szenario 1 könnten demnach bei einem Schwellenwert (Fallzahl) von 100 VLBW-I potentiell 74 Sterbefälle vermieden werden, gemäß dem Szenario 2 dagegen 103 (dabei gab es im Jahr 2018 insgesamt 382 Sterbefälle in der Grundgesamtheit). Dies entspräche im Szenario 1 einer NNT von 70, im Szenario 2 von 51. Allerdings müssen diese Angaben schon aufgrund der wenigen PNZ mit mindesten 100 VLBW-I pro Jahr als recht unsicher angesehen werden, worauf auch die breiten Konfidenzintervalle hinweisen. Bei einem Schwellenwert von 60 ergeben sich im Szenario 1 29 potentiell vermiedene Sterbefälle und im Szenario 2 52 potentiell vermiedene Sterbefälle. Die zugehörigen NNT für diesen Schwellenwert wären 146 (Szenario 1) bzw. 79 (Szenario 2).

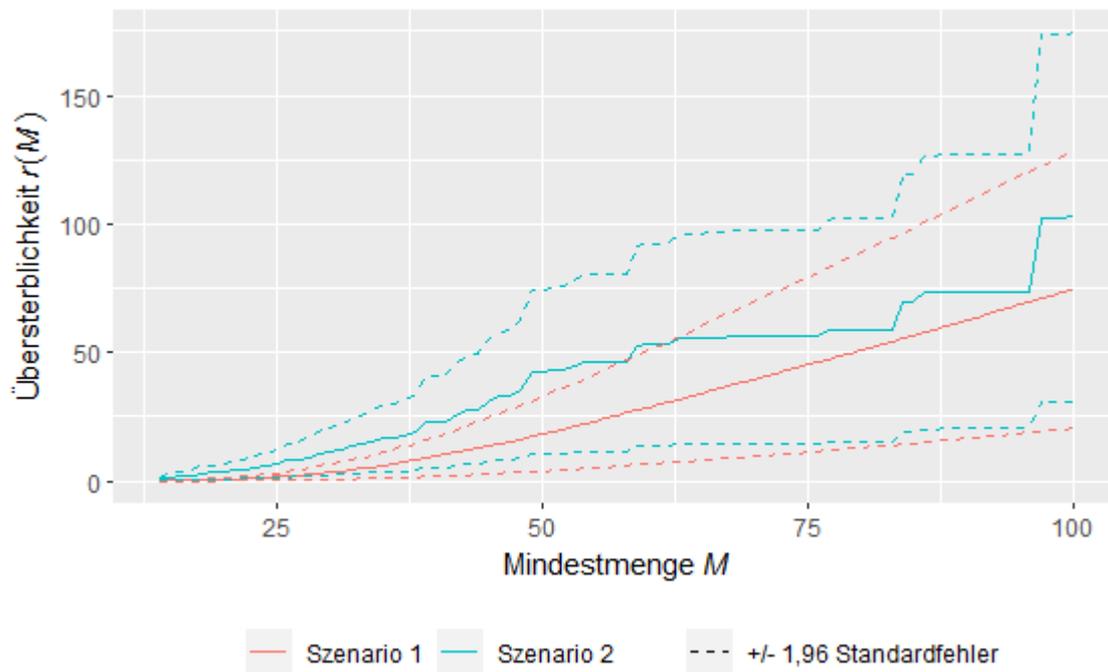


Abbildung 18: Geschätzte Anzahl an Sterbefällen, die durch die neue Mindestmenge M verhindert werden könnte, als Funktion von M .

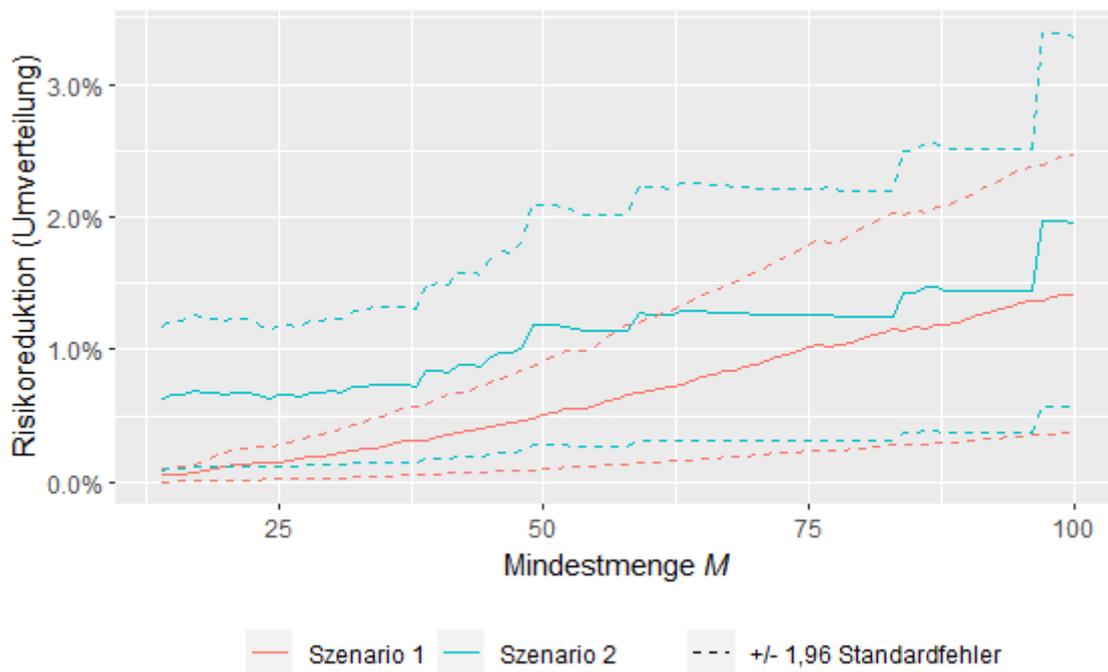


Abbildung 19: Geschätzte Risikoreduktion $\delta_u(M)$ für die umverteilten Fälle als Funktion von M .

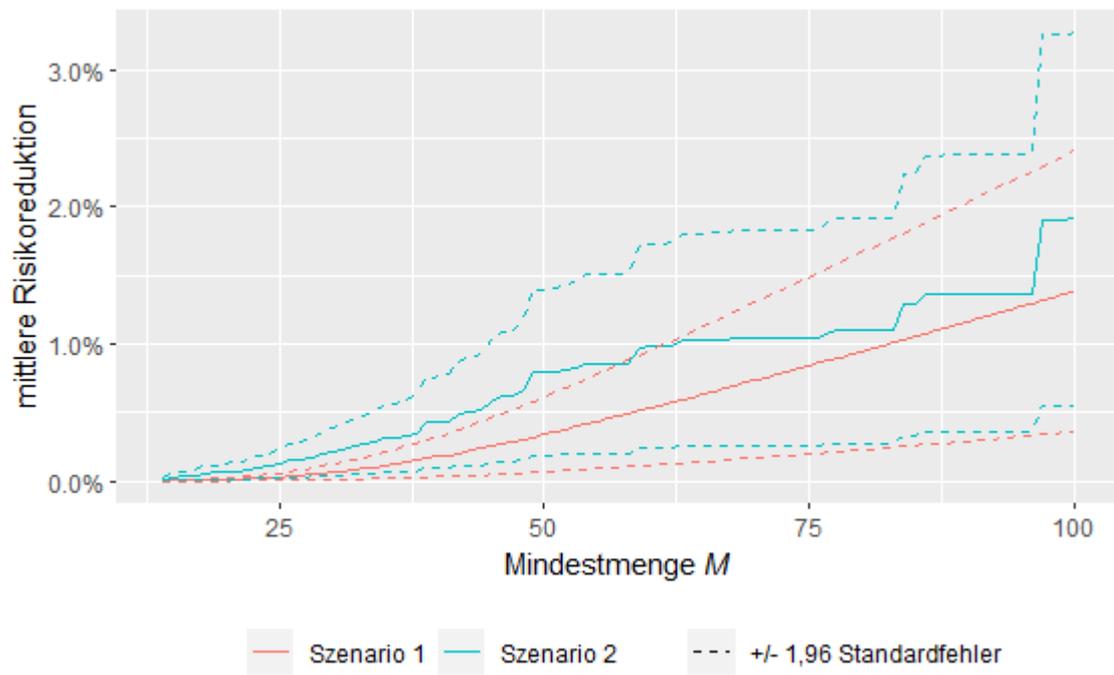


Abbildung 20: Geschätzte durchschnittliche Risikoreduktion $\delta_d(M)$ als Funktion von M .

Tabelle 11: Geschätzte Werte der Maßzahlen im Szenario 1

M	Übersterblichkeit pro Jahr (SE)	Risikoreduktion; Umverteilungen (SE)	Risikoreduktion; gesamt (SE)	NNT (SE)	Impact Number (SE)
15	0,12 (± 0,05)	0,053 % (± 0,024)	0,0021 % (± 0,00098)	1891 (± 863)	46598 (± 21256)
20	0,63 (± 0,28)	0,115 % (± 0,050)	0,0117 % (± 0,00512)	867 (± 379)	8546 (± 3740)
25	1,65 (± 0,71)	0,155 % (± 0,066)	0,0307 % (± 0,01316)	645 (± 277)	3258 (± 1397)
30	3,46 (± 1,46)	0,209 % (± 0,088)	0,0644 % (± 0,02715)	478 (± 201)	1552 (± 654)
35	6,08 (± 2,53)	0,271 % (± 0,113)	0,1132 % (± 0,04701)	368 (± 153)	883 (± 367)
40	9,51 (± 3,90)	0,344 % (± 0,141)	0,1770 % (± 0,07251)	291 (± 119)	565 (± 232)
45	13,59 (± 5,50)	0,417 % (± 0,169)	0,2530 % (± 0,10245)	240 (± 97)	395 (± 160)
50	18,27 (± 7,32)	0,508 % (± 0,204)	0,3401 % (± 0,13621)	197 (± 79)	294 (± 118)
55	23,36 (± 9,26)	0,583 % (± 0,231)	0,4348 % (± 0,17242)	172 (± 68)	230 (± 91)
60	28,64 (± 11,25)	0,686 % (± 0,270)	0,5331 % (± 0,20941)	146 (± 57)	188 (± 74)
65	34,08 (± 13,27)	0,793 % (± 0,309)	0,6344 % (± 0,24699)	126 (± 49)	158 (± 61)
70	39,66 (± 15,31)	0,890 % (± 0,344)	0,7383 % (± 0,28504)	112 (± 43)	135 (± 52)
75	45,22 (± 17,32)	1,015 % (± 0,389)	0,8418 % (± 0,32236)	99 (± 38)	119 (± 45)
80	50,85 (± 19,32)	1,084 % (± 0,412)	0,9465 % (± 0,35967)	92 (± 35)	106 (± 40)
85	56,53 (± 21,32)	1,170 % (± 0,441)	1,0523 % (± 0,39683)	85 (± 32)	95 (± 36)
90	62,54 (± 23,42)	1,232 % (± 0,461)	1,1642 % (± 0,43590)	81 (± 30)	86 (± 32)

95	68,50 (± 25,46)	1,350 % (± 0,502)	1,2752 % (± 0,47403)	74 (± 28)	78 (± 29)
100	74,42 (± 27,47)	1,419 % (± 0,524)	1,3853 % (± 0,51127)	70 (± 26)	72 (± 27)

Tabelle 12: Geschätzte Werte der Maßzahlen im Szenario 2

M	Übersterblichkeit pro Jahr (SE)	Risikoreduktion; Umverteilungen (SE)	Risikoreduktion; gesamt (SE)	NNT (SE)	Impact Number (SE)
15	1,5 (± 0,6)	0,67 % (± 0,28)	0,027 % (± 0,011)	150 (± 63)	3692 (± 1556)
20	3,6 (± 1,5)	0,67 % (± 0,28)	0,068 % (± 0,028)	150 (± 62)	1478 (± 611)
25	7,0 (± 2,8)	0,65 % (± 0,27)	0,129 % (± 0,053)	153 (± 62)	773 (± 315)
30	11,3 (± 4,6)	0,69 % (± 0,28)	0,211 % (± 0,085)	146 (± 59)	474 (± 191)
35	16,6 (± 6,6)	0,74 % (± 0,29)	0,308 % (± 0,123)	135 (± 54)	324 (± 129)
40	23,2 (± 9,1)	0,84 % (± 0,33)	0,431 % (± 0,170)	119 (± 47)	232 (± 91)
45	30,9 (± 12,0)	0,95 % (± 0,37)	0,576 % (± 0,224)	105 (± 41)	174 (± 68)
50	42,7 (± 16,3)	1,19 % (± 0,45)	0,795 % (± 0,303)	84 (± 32)	126 (± 48)
55	46,0 (± 17,5)	1,15 % (± 0,44)	0,856 % (± 0,326)	87 (± 33)	117 (± 45)
60	52,9 (± 20,0)	1,27 % (± 0,48)	0,985 % (± 0,372)	79 (± 30)	102 (± 38)
65	55,3 (± 20,8)	1,29 % (± 0,49)	1,030 % (± 0,388)	78 (± 29)	97 (± 37)
70	56,2 (± 21,2)	1,26 % (± 0,48)	1,046 % (± 0,394)	79 (± 30)	96 (± 36)
75	56,2 (± 21,2)	1,26 % (± 0,48)	1,046 % (± 0,394)	79 (± 30)	96 (± 36)

80	58,9 (± 22,1)	1,25 % (± 0,47)	1,096 % (± 0,412)	80 (± 30)	91 (± 34)
85	69,2 (± 25,7)	1,43 % (± 0,53)	1,289 % (± 0,478)	70 (± 26)	78 (± 29)
90	73,5 (± 27,2)	1,45 % (± 0,54)	1,369 % (± 0,506)	69 (± 25)	73 (± 27)
95	73,5 (± 27,2)	1,45 % (± 0,54)	1,369 % (± 0,506)	69 (± 25)	73 (± 27)
100	102,8 (± 36,6)	1,96 % (± 0,70)	1,913 % (± 0,681)	51 (± 18)	52 (± 19)

7 Diskussion

Auf Grundlage der dargestellten Auswertungen und Ergebnisse lässt sich für Frühgeborene mit einem Aufnahmegewicht kleiner 1.250 g ein eindeutiger Zusammenhang zwischen Leistungsmenge und risikoadjustierter Sterbewahrscheinlichkeit feststellen. Der Zusammenhang ist dabei monoton fallend und auf Logitebene annähernd linear. Mit einer Zunahme der Fälle sinkt demnach das Sterberisiko der VLBW-I bei sonst identischer Fallkonstellation.

Dies bedeutet zugleich, dass kein Schwellenwert erkennbar ist bzw. anzunehmen wäre, dass eine Umverteilung in Perinatalzentren mit einer möglichst hohen Fallzahl mit einem optimalen Ergebnis einhergehen. Andererseits ist dies ggf. auch ein Ergebnis des statistischen Instrumentariums, das beim Volume-Outcome-Zusammenhang ein stetiges Modell generiert, das dem Instrumentarium entsprechend keine Stufen zulässt. Ein monoton fallender Zusammenhang wird aber durch die Bruchpunktanalysen bestätigt und auch die fachliche Interpretation kann einen solchen Bruchpunkt aus der Versorgung heraus nicht nachvollziehbar begründen. So kann das hier verwendete Instrumentarium zwar den Volume-Outcome-Zusammenhang klar erkennen, aber es bedarf der Kombination mit den Instrumenten der Folgeabschätzung, um tatsächlich Schwellenwerte anhand der Auswirkungen auf die regionale Versorgung festsetzen zu können. Zusätzlich sollte beachtet werden, dass die Aussagen des Modells für derart hohe Fallzahlen deutlich weniger präzise werden, da nur wenige Perinatalzentren derart hohe Fallzahl aufweisen. Darüber hinaus ist zu berücksichtigen, dass eine Behandlung aller VLBW-I in wenigen Perinatalzentren die hier nicht berücksichtigten Gefahren von Umverteilungen im Sinne von Transportrisiken und Ähnlichem erheblich erhöhen dürften. Vor diesem Hintergrund scheint eine Durchführung und Orientierung an Studien zur Folgeabschätzung dringend geboten und empfehlenswert.

Gemäß den Ergebnissen dieses Berichts scheint es darüber hinaus sinnvoll, ein längeres Zeitintervall statt eines Eineinhalb-Jahres-Zeitintervalls, wie aktuell in der Mm-RL geregelt, für die Ermittlung einer Mindestmenge heranzuziehen, da sich für Fallzahlen auf Basis von fünf Jahren stärkere Volume-Outcome-Assoziationen ergaben. Umgekehrt stellen fünf Jahre einen sehr langen Zeitraum dar, sodass abgewogen werden sollte, wie lange ein Zeitraum für die Festlegung einer Mindestmenge sinnvollerweise sein soll.

Limitationen der vorliegenden Analyse

Neben der bereits angesprochenen Problematik fehlender Berücksichtigung von Risiken einer Umverteilung (z. B. durch ggf. drastische Kapazitätserhöhungen in Perinatalzentren oberhalb der Mindestmenge oder wegen längerer Transportzeiten) weist die hier durchgeführte Analyse eine Reihe von Limitationen auf, die im Folgenden angesprochen werden sollen.

Zunächst handelt es sich bei der Analyse um eine *retrospektive Beobachtungsstudie*, welche kausale Rückschlüsse nicht erlaubt. Dies spiegelt sich u. a. in dem Umstand wieder, dass eine vollständige Risikoadjustierung nicht möglich ist und ungleiche Verteilungen unberücksichtigter patientenbezogener Risikofaktoren die Analyse beeinflusst haben könnten.

Auf der Ebene der *statistischen Modellierung* konnten, wie bereits erwähnt, einige Aspekte nicht berücksichtigt werden (z. B. Clusterbildung unterhalb der Ebene der PNZ, Analyse von Interaktionen zwischen den patientenseitigen Risikofaktoren und den standortspezifischen Einflüssen). Sei dies aufgrund von Problemen der statistischen Schätzbarkeit oder auch aufgrund der Tatsache, dass interessierende Variablen (wie etwa die Qualifikation des Personals) nicht oder nicht in einer Form vorlagen, die eine seriöse Verwendung in den hier dargestellten statistischen Modellen sinnvoll erscheinen ließ. Im Vergleich zum Bericht zu den Volumen-Outcome-Beziehungen bei Revisionseingriffen in der Knieendoprothetik (IQTIG 2019e) erweist es sich dagegen als günstig, dass der hier betrachtete Analyseendpunkt vergleichsweise häufiger auftritt und Daten mehrerer Erfassungsjahre zur Verfügung stehen.

Insgesamt werden Volume-Outcome-Zusammenhänge im Kontext von kausalen Argumentationen oft als ein *Surrogatparameter* („indirect quality indicator“, Rogowski et al. 2004) betrachtet. Neben einer zugestandenen Übung bzw. Routine mit den zu analysierenden Erkrankungen werden gerne Faktoren wie die Anzahl des vor Ort verfügbaren Personals und dessen Qualifikation oder weitere Strukturmerkmale wie etwa apparative oder bauliche Ausstattungsmerkmale genannt.

In der hier vorgelegten Analyse wurde allerdings kein Versuch unternommen, den Effekt des Fallvolumens durch derartige *Strukturmerkmale* zu erklären oder davon zu trennen: Einerseits lagen diese Merkmale der Perinatalzentren teilweise nicht oder nicht in ausreichender Qualität in den zu verwendenden Daten vor, andererseits kann mit Blick auf die Beauftragung wie auch mit Blick auf die Formulierung in § 136b Abs. 1 SGB V die Frage gestellt werden, inwieweit eine Adjustierung nach struktur- oder krankenhausbazogenen Variablen für die Ermittlung einer Mindestmenge überhaupt im Sinne des Auftrages bzw. allgemeiner im Sinne der gesetzlichen Regelung ist.

In zahlreichen Arbeiten wurde die starke *Variabilität der PNZ-spezifischen Mortalitäten* jenseits des beschriebenen Volume-Outcome-Zusammenhanges beschrieben (Rogowski et al. 2004, Heller 2011, Kutschmann et al. 2012, Rochow et al. 2016). Auch in den hier vorgelegten Analysen zeigt sich, dass die Variabilität der Sterblichkeit bei PNZ vergleichbarer Fallzahl im Vergleich zur Variabilität, die auf Fallzahlunterschiede zurückzuführen ist, nicht unerheblich ist. Dieser Umstand weist darauf hin, dass es trotz einer Einführung bzw. der Anhebung von Mindestmengen unabdingbar bleibt, weiterhin beständige Qualitätsmessungen und Qualitätsverbesserungsprozesse durchzuführen, wie sie derzeit im QS-Verfahren *Neonatologie* und bei der verpflichtenden Ergebnisveröffentlichung auf perinatalzentren.org betrieben werden.

Schließlich sollte betont werden, dass die hier berechneten *absoluten Effektmaße*, dies wird am Beispiel der potentiell vermeidbaren Todesfälle am deutlichsten, keinesfalls den Anspruch haben, eine exakte Anzahl an vermeidbaren Todesfällen oder Ereignissen nach Einführung einer Mindestmenge zu prognostizieren. Beispielsweise fehlen sekundäre Umverteilungseffekte ebenso, wie eine Berücksichtigung von Verlegungs-, Transportrisiken oder zeitliche Trends. Ziel der Berechnung dieser Effektmaße ist es vielmehr lediglich, Größenordnungen von Effekten angeben zu können, die durch eine Einführung einer Mindestmenge erwartet werden können. Ein

weiteres Ziel dieser Berechnung ist es, ein allgemein verständliches Maß anzugeben (Gigerenzer et al. 2007).

Fazit

Man könnte versucht sein, auf Basis der hier durchgeführten Datenanalysen für eine hohe Mindestmenge in der Versorgung von VLBW-I zu votieren. Der größte Effekt wäre theoretisch bei der Konzentration aller VLBW-I in wenigen PNZ mit mehr als 100 VLBW-I zu erwarten. Dies wäre ein Ergebnis, das sich zuletzt in Arbeiten auf Basis von aktuellen Abrechnungsdaten wie auch auf Basis von Daten aus perinatalzentren.org insofern angedeutet hatte, als hier die stärksten Effekte für vergleichsweise große Fallzahlen berichtet wurden (Heller 2018, Rochow et al. 2016). Eine solche Schlussfolgerung würde jedoch übersehen, dass die statistische Unsicherheit im Bereich großer Fallzahlen sehr hoch ist und in den hier beauftragten Analysen die regionalen Aspekte von Umverteilungen völlig ausgeblendet bleiben. Regionale Erreichbarkeit, Umverteilungsrisiken und Transportrisiken sind jedoch die Aspekte, die eine zu hohe Konzentration der Versorgung auf wenige Einrichtungen verbieten. Abgesehen davon, dass es auch bei fallzahlstarken Einrichtungen große Qualitätsunterschiede gibt Kutschmann et al. (2012), würde der Vorteil großer Zentren durch die genannten Risiken möglicherweise wieder aufgehoben.

Die Ergebnisse der hier vorgelegten Analysen zeigen eine signifikante Volumen-Outcome-Relation, ohne dass ein distinkter Schwellenwert ableitbar ist. Detaillierte Folgeabschätzungen zu den Auswirkungen von Mindestmengen auf die Verfügbarkeit und Qualität regionaler Versorgung von Frühgeburten sind nun der nächste Schritt, um mögliche positive wie negative Folgen einer Umverteilung im Rahmen der Einführung einer Mindestmenge einschätzen zu können. Aus dieser Abwägung ist dann ein konkreter Schwellenwert als Kompromiss zwischen Zentrierung und wohnortnaher, regionaler Versorgung abzuleiten.

8 Literatur

- Agresti, A (2013): *Categorical Data Analysis*. Third Edition. (Wiley Series in Probability and Statistics). Hoboken, US-NJ: Wiley. ISBN: 978-0-470-46363-5.
- AQUA [Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen] (2015): Weiterentwicklung der Methodik Webseite www.perinatalzentren.org.
- Konzeptskizze zur Abbildung des Verlegungsgeschehens. Stand: 30. September 2015. Göttingen: AQUA. Signatur: 15-SQG-026.
- Bartels, DB; Wypij, D; Wenzlaff, P; Dammann, O; Poets, CF (2006): Hospital Volume and Neonatal Mortality Among Very Low Birth Weight Infants. *Pediatrics* 117(6): 2206-2214. DOI: 10.1542/peds.2005-1624.
- Bender, R (1999): Quantitative Risk Assessment in Epidemiological Studies Investigating Threshold Effects. *Biometrical Journal* 41(3): 305-319. DOI: 10.1002/(sici)1521-4036(199906)41:3<305::Aid-bimj305>3.0.Co;2-y.
- Birkmeyer, JD; Siewers, AE; Finlayson, EV; Stukel, TA; Lucas, FL; Batista, I; et al. (2002): Hospital Volume and Surgical Mortality in the United States. *The New England Journal of Medicine* 346(15): 1128-1137. DOI: 10.1056/NEJMsa012337.
- Bonita, R; Beaglehole, R; Kjellström, T (2006): *Basic epidemiology*. 2nd edition. Geneva, CH: WHO [World Health Organization]. ISBN: 978-92-4-154707-9. URL: https://apps.who.int/iris/bitstream/handle/10665/43541/9241547073_eng.pdf?sequence=1&isAllowed=y (abgerufen am: 09.06.2020).
- Budtz-Jørgensen, E; Keiding, N; Grandjean, P (2001): Benchmark Dose Calculation from Epidemiological Data. *Biometrics* 57(3): 698-706. DOI: 10.1111/j.0006-341x.2001.00698.x.
- Eilers, PHC; Marx, BD (1996): Flexible Smoothing with *B*-splines and Penalties. *Statistical Science* 11(2): 89-102.
- Fahrmeir, L; Kneib, T; Lang, S; Marx, B (2013): *Regression. Models, Methods and Applications*. Berlin [u. a.]: Springer. ISBN: 978-3-642-34332-2.
- G-BA [Gemeinsamer Bundesausschuss] (2019a): Beschluss des Gemeinsamen Bundesausschusses über die Veröffentlichung des Modells zur Risikoadjustierung gemäß § 7 Absatz 2 Satz 3 der Anlage 4 QFR-RL. 19.12.2019. Berlin: G-BA. URL: https://www.g-ba.de/downloads/39-261-4105/2019-12-19_QFR-RL_Veroeffentlichung-Modell-Risikoadjustierung.pdf (abgerufen am: 12.05.2020).
- G-BA [Gemeinsamer Bundesausschuss] (2009): Bekanntmachung eines Beschlusses des Gemeinsamen Bundesausschusses zur Versorgung von Früh- und Neugeborenen. 20.08.2009. Berlin: G-BA. BAnz. Nr. 195 (S. 4 450) vom 24.12.2009. URL: https://www.g-ba.de/downloads/39-261-875/2009-08-20-VB-NICU_BAnz.pdf (abgerufen am: 08.06.2020).

- G-BA [Gemeinsamer Bundesausschuss] (2018): Beschluss des Gemeinsamen Bundesausschusses über die Veröffentlichung des Modells zur Risikoadjustierung gemäß § 7 Absatz 2 Satz 3 der Anlage 4 QFR-RL. 22.11.2018. Berlin: G-BA. URL: https://www.g-ba.de/downloads/39-261-3580/2018-11-22_QFR-RL_Veroeffentlichung-Risikoadjustierungsmodell_WZ.pdf (abgerufen am: 06.05.2020).
- G-BA [Gemeinsamer Bundesausschuss] (2019b): Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des IQTIG mit der Auswertung von vorhandenen Daten zur Versorgung von Frühgeborenen mit einem Aufnahmegegewicht von unter 1250g [Beschluss]. 20.06.2019. Berlin: G-BA. URL: https://www.g-ba.de/downloads/39-261-3818/2019-0-20_Mm-R_Beauftragung-IQTIG_Auswertung_Versorgungsdaten_Fr%C3%BChgeborene.pdf (abgerufen am: 13.01.2020).
- Gelman, A; Hill, J (2007): Data Analysis Using Regression and Multilevel/Hierarchical Models. (Analytical Methods for Social Research). Cambridge, GB: Cambridge University Press. ISBN: 978-0-521-86706-1.
- George, EI; Ročková, V; Rosenbaum, PR; Satopää, VA; Silber, JH (2017): Mortality Rate Estimation and Standardization for Public Reporting: Medicare's Hospital Compare. *Journal of the American Statistical Association* 112(519): 933-947. DOI: 10.1080/01621459.2016.1276021.
- Gigerenzer, G; Gaissmaier, W; Kurz-Milcke, E; Schwartz, LM; Woloshin, S (2007): Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest* 8(2): 53-96. DOI: 10.1111/j.1539-6053.2008.00033.x.
- Gortmaker, S; Sobol, A; Clark, C; Walker, DK; Geronimus, A (1985): The survival of very low-birth weight infants by level of hospital of birth: a population study of perinatal systems in four states. *American Journal of Obstetrics and Gynecology* 152(5): 517-524. DOI: 10.1016/0002-9378(85)90618-0.
- Grouven, U; Küchenhoff, H; Schröder, P; Bender, R (2008): Flexible regression models are useful tools to calculate and assess threshold values in the context of minimum provider volumes. *Journal of Clinical Epidemiology* 61(11): 1125-1131. DOI: 10.1016/j.jclinepi.2007.11.020.
- Heller, G; Richardson, DK; Schnell, R; Misselwitz, B; Künzel, W; Schmidt, S (2002): Are we regionalized enough? Early-neonatal deaths in low-risk births by the size of delivery units in Hesse, Germany 1990-1999. *International Journal of Epidemiology* 31(5): 1061-1068. DOI: 10.1093/ije/31.5.1061.
- Heller, G; Schnell, R; Richardson, DK; Misselwitz, B; Schmidt, S (2003): Hat die Größe der Geburtsklinik Einfluss auf das neonatale Überleben? *Deutsche Medizinische Wochenschrift* 128(13): 657-662. DOI: 10.1055/s-2003-38284.
- Heller, G; Günster, C; Misselwitz, B; Feller, A; Schmidt, S (2007): Jährliche Fallzahl pro Klinik und Überlebensrate sehr untergewichtiger Frühgeborener (VLBW) in Deutschland – Eine

bundesweite Analyse mit Routinedaten. *Zeitschrift für Geburtshilfe und Neonatologie* 211(3): 123-131. DOI: 10.1055/s-2007-960747.

Heller, G (2009): Auswirkungen der Einführung von Mindestmengen in der Behandlung von sehr untergewichtigen Früh- und Neugeborenen (VLBW). Eine Simulation mit Echtdateien. Teil II. Kapitel 13. In: Klauber, J; Robra, B-P; Schnellschmidt, H; Hrsg.: *Krankenhaus-Report 2008/2009. Schwerpunkt: Versorgungszentren*. Stuttgart: Schattauer, 183-199. ISBN: 978-3-7945-6500-9. URL:

https://www.wido.de/fileadmin/Dateien/Dokumente/Publikationen_Produkte/Buchreihen/Krankenhausreport/2008-2009/Kapitel%20mit%20Deckblatt/wido_khr20082009_kap13.pdf (abgerufen am: 11.09.2019).

Heller, G (2011): Ermittlung der klinikspezifischen Ergebnisqualität der Behandlung von Früh- und Neugeborenen mit sehr niedrigem Geburtsgewicht (VLBW) auf der Basis von Routinedaten. Teil I. Kapitel 3. In: Klauber, J; Geraedts, M; Friedrich, J; Wasem, J; Hrsg.: *Krankenhaus-Report 2011. Schwerpunkt: Qualität durch Wettbewerb*. Stuttgart: Schattauer, 39-47. ISBN: 978-3-7945-2802-8. URL:

https://www.wido.de/fileadmin/Dateien/Dokumente/Publikationen_Produkte/Buchreihen/Krankenhausreport/2011/Kapitel%20mit%20Deckblatt/wido_khr2011_gesamt.pdf (abgerufen am: 16.06.2020).

Heller, G; Konheiser, S; Biermann, A (2015): Erhebung der Versorgungsqualität von Früh- und Neugeborenen mit sehr niedrigem Geburtsgewicht auf der Basis von Routinedaten. Teil II. Kapitel 16. In: Klauber, J; Geraedts, M; Friedrich, J; Wasem, J; Hrsg.: *Krankenhaus-Report 2015. Schwerpunkt: Strukturwandel*. Stuttgart: Schattauer, 249-264. ISBN: 978-3-7945-3091-5. URL:

https://www.wido.de/fileadmin/Dateien/Dokumente/Publikationen_Produkte/Buchreihen/Krankenhausreport/2015/Kapitel%20mit%20Deckblatt/wido_khr2015_gesamt.pdf (abgerufen am: 16.06.2020).

Heller, G; Bauer, E; Schill, S; Thomas, T; Louwen, F; Wolff, F; et al. (2017): Entscheidungs-Entbindungszeit und perinatale Komplikationen bei Notkaiserschnitt. *Deutsches Ärzteblatt* 114(35-36): 589-596. DOI: 10.3238/arztebl.2017.0589.

Heller, G (2018): Optionen der Regionalisierung der Frühgeborenenversorgung: Mindestmengen und Ergebnisqualität – Analysen zu Volume-Outcome-Zusammenhängen und zur Auswirkung einer Einführung von Mindestmengen im Vergleich zu einer Planung über Ergebnisqualität. Kapitel I. Abschnitt 6. In: Dormann, F; Klauber, J; Kuhlen, R: *Qualitätsmonitor 2018*. Berlin: MMW – Medizinisch Wissenschaftliche Verlagsgesellschaft, 85-103. ISBN: 978-3-95466-348-4. URL: https://www.gesundheitsstadt-berlin.de/fileadmin/user_upload/Downloads/Qualitaetsmonitor_2018_E-Book_V02_anGesundheitsstadt.pdf (abgerufen am: 11.09.2019).

- Hentschel, R; Guenther, K; Vach, W; Bruder, I (2019): Risk-adjusted mortality of VLBW infants in high-volume versus low-volume NICUs. *Archives of Disease in Childhood: Fetal & Neonatal* 104(4): F390-F395. DOI: 10.1136/archdischild-2018-314956.
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2019a): Geburtshilfe. Bundesauswertung zum Erfassungsjahr 2018. Qualitätsindikatoren und Kennzahlen. Stand: 23.07.2019. Berlin: IQTIG. URL: https://iqtig.org/dateien/dasiqtig/grundlagen/IQTIG_Methodische-Grundlagen-V1.1_barrierefrei_2019-04-15.pdf (abgerufen am: 05.05.2020).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2019b): Methodische Grundlagen V1.1. Stand: 15.04.2019. Berlin: IQTIG. URL: https://iqtig.org/dateien/dasiqtig/grundlagen/IQTIG_Methodische-Grundlagen-V1.1_barrierefrei_2019-04-15.pdf (abgerufen am: 05.05.2020).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2019c): Neonatalogie. Beschreibung der Qualitätsindikatoren und Kennzahlen nach QSKH-RL. Erfassungsjahr 2018. Stand: 25.04.2019. Berlin: IQTIG. URL: https://iqtig.org/downloads/auswertung/auswertung/2018/neo/QSKH_NEO_2018_QIDB_V03_2019-04-25.pdf (abgerufen am: 06.05.2020).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2019d): Umsetzungskonzept zur Abbildung des Verlegungsgeschehens bei Frühgeborenen. Abschlussbericht. Stand: 29.03.2019 Berlin: IQTIG. [unveröffentlicht].
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2019e): Volumen-Outcome-Beziehungen bei Revisionseingriffen in der Knieendoprothetik. Sonderauswertung des vormaligen Leistungsbereichs Knie-Endoprothesen-Wechsel und -komponentenwechsel. Stand: 20.12.2019. Berlin: IQTIG. [unveröffentlicht].
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2020): Technische Dokumentation zur Spezifikation für Mindestmengenregelungen. Version 2019. Stand: 07.01.2020. Berlin: IQTIG. URL: <https://iqtig.org/datenerfassung/spezifikationen/spezifikation-fuer-die-mindestmengenregelungen/2019/v03/> [Spezifikation der Mindestmengenregelung 2019, Version 03 > Komplettdownload Mindestmengenspezifikation > 2019_MMR_RB_XML_V03.zip > 2019_MMR_TechDok_V02.pdf] (abgerufen am: 16.06.2020).
- IQWIG [Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen] (2005): Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für Knie-Totalendoprothese. Abschlussbericht. (IQWIG-Berichte, Jahr: 2005 Nr. 2). Köln: IQWIG. Interne Auftragsnummer: B05/01a. URL: <https://www.iqwig.de/de/projekte-ergebnisse/projekte/medizinische-biometrie/b05-01a-berechnung-von-schwellenwerten-bei-mindestmengen-fuer-die-knie-totalendoprothese.1218.html> [Berichtsdokumente > Abschlussbericht] (abgerufen am: 25.02.2020).

- IQWiG [Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen] (2006): Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für die Koronarchirurgie. Abschlussbericht. (IQWiG-Berichte, Jahr: 2006 Nr. 9). Köln: IQWiG. Interne Auftragsnummer: B05/01b. URL: <https://www.iqwig.de/de/projekte-ergebnisse/projekte/medizinische-biometrie/b05-01b-berechnung-von-schwellenwerten-bei-mindestmengen-fuer-die-koronarchirurgie.1217.html> [Berichtsdokumente > Abschlussbericht] (abgerufen am: 25.02.2020).
- IQWiG [Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen] (2008): Zusammenhang zwischen Leistungsmenge und Ergebnis bei der Versorgung von Früh- und Neugeborenen mit sehr geringem Geburtsgewicht. Abschlussbericht. [Stand:] 14.08.2008. (IQWiG-Berichte, Jahr: 2008 Nr. 35). Köln: IQWiG. Interne Auftragsnummer: V07-01. URL: <https://www.iqwig.de/de/projekte-ergebnisse/projekte/versorgung/v07-01-zusammenhang-zwischen-leistungsmenge-und-ergebnis-bei-der-versorgung-von-frueh-und-neugeborenen-mit-sehr-geringem-geburtsgewicht.1224.html> [Berichtsdokumente > Abschlussbericht] (abgerufen am: 03.06.2020).
- IQWiG [Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen] (2019): Zusammenhang zwischen Leistungsmenge und Qualität des Behandlungsergebnisses bei Lebertransplantation (inklusive Teilleber-Lebendspende) [Rapid Report]. Stand: 04.09.2019. (IQWiG-Berichte, Nr. 813). Köln: IQWiG. Auftrag: V18-04. URL: <https://www.iqwig.de/de/projekte-ergebnisse/projekte/versorgung/v18-04-zusammenhang-zwischen-leistungsmenge-und-qualitaet-des-behandlungsergebnisses-bei-lebertransplantationen-rapid-report.10904.html> [Berichtsdokumente > Rapid Report] (abgerufen am: 04.06.2020).
- Jochum, F; Schmidt, B; Schunk, K; Hesse, V; Distler, B; Laske, G; et al. (2008): Quality of care of VLBW neonates: relationship between unit volume and outcome is different between metropolitan and rural regions. *Acta Paediatrica* 97(10): 1470-1474. DOI: 10.1111/j.1651-2227.2008.00934.x.
- Kutschmann, M; Bungard, S; Kötting, J; Trümner, A; Fusch, C; Veit, C (2012): Versorgung von Frühgeborenen mit einem Geburtsgewicht unter 1 250 g. *Deutsches Ärzteblatt* 109(31-32): 519-526. DOI: 10.3238/arztebl.2012.0519.
- Larsen, K; Petersen, JH; Budtz-Jørgensen, E; Endahl, L (2000): Interpreting Parameters in the Logistic Regression Model with Random Effects. *Biometrics* 56(3): 909-914. DOI: 10.1111/j.0006-341x.2000.00909.x.
- Marra, G; Wood, SN (2012): Coverage Properties of Confidence Intervals for Generalized Additive Model Components. *Scandinavian Journal of Statistics* 39(1): 53-74. DOI: 10.1111/j.1467-9469.2011.00760.x.
- McCulloch, CE; Neuhaus, JM (2011): Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. *Statistical Science* 26(3): 388-402.

- Miedaner, F; Langhammer, K; Enke, C; Göpel, W; Kribs, A; Nitzsche, A; et al. (2018): Volume, size, professionals' specialization and nutrition management of NICUs and their association with treatment quality in VLBW infants. *Journal of Perinatology* 38(4): 402-410. DOI: 10.1038/s41372-017-0036-0.
- Obladen, M (2007): Mindestmengen in der Versorgung sehr untergewichtiger Frühgeborener: Eine Literaturübersicht. *Zeitschrift für Geburtshilfe und Neonatologie* 211(3): 110-117. DOI: 10.1055/s-2007-960745.
- Panageas, KS; Schrag, D; Riedel, E; Bach, PB; Begg, CB (2003): The Effect of Clustering of Outcomes on the Association of Procedure Volume and Surgical Outcomes. *Annals of Internal Medicine* 139(8): 658-665, E-666. DOI: 10.7326/0003-4819-139-8-200310210-00009.
- Paneth, N; Kiely, JL; Wallenstein, S; Susser, M (1987): The Choice of Place of Delivery: Effect of Hospital Level on Mortality in All Singleton Births in New York City. *American Journal of Diseases of Children* 141(1): 60-64. DOI: 10.1001/archpedi.1987.04460010060024.
- Perperoglou, A; Sauerbrei, W; Abrahamowicz, M; Schmid, M (2019): A review of spline function procedures in R. *BMC Medical Research Methodology* 19:46. DOI: 10.1186/s12874-019-0666-3.
- Phibbs, CS; Baker, LC; Caughey, AB; Danielsen, B; Schmitt, SK; Phibbs, RH (2007): Level and Volume of Neonatal Intensive Care and Mortality in Very-Low-Birth-Weight Infants. *The New England Journal of Medicine* 356(21): 2165-2175. DOI: 10.1056/NEJMsa065029.
- R Core Team (2020): R: A Language and Environment for Statistical Computing [*Open Source Software*]. R version 3.6.3 (Holding the Windsock). Vienna, AT: R Foundation for Statistical Computing. URL: <https://cran.r-project.org/> [Source Code for all Platforms > Source code of older versions of R is available here >R-3/ > R-3.6.3.tar.gz] (abgerufen am: 10.06.2020).
- Rochow, N; Landau-Crangle, E; Lee, S; Schünemann, H; Fusch, C (2016): Quality Indicators but Not Admission Volumes of Neonatal Intensive Care Units Are Effective in Reducing Mortality Rates of Preterm Infants. *PLoS ONE* 11(8): e0161030. DOI: 10.1371/journal.pone.0161030.
- Rogowski, JA; Horbar, JD; Staiger, DO; Kenny, M; Carpenter, J; Geppert, J (2004): Indirect vs Direct Hospital Quality Indicators for Very Low-Birth-Weight Infants. *JAMA* 291(2): 202-209. DOI: 10.1001/jama.291.2.202.
- Rossi, R (2015): Qualität perinatologischer Versorgung im internationalen Vergleich und die konfliktträchtige Einführung der Mindestmengenregelung in der Neonatologie in Deutschland. *Pädiatrie & Pädologie* 50(Suppl. 2): S59-S65. DOI: 10.1007/s00608-015-0278-z.
- Rothman, KJ; Greenland, S; Lash, TL (2008): *Modern Epidemiology*. Third Edition. Philadelphia, US-PA [u. a.]: Lippincott Williams and Wilkins. ISBN: 978-1-4511-9005-2.

- Ruppert, D (2002): Selecting the Number of Knots for Penalized Splines. *Journal of Computational and Graphical Statistics* 11(4): 735-757. DOI: 10.1198/106186002853.
- Schröder, P; Grouven, U; Bender, R (2007): Können Mindestmengen für Knieprothesen anhand von Routinedaten errechnet werden? *Der Orthopäde* 36(6): 570-576. DOI: 10.1007/s00132-007-1066-7.
- Shah, PS; Mirea, L; Ng, E; Solimano, A; Lee, SK (2015): Association of unit size, resource utilization and occupancy with outcomes of preterm infants. *Journal of Perinatology* 35: 522-529. DOI: 10.1038/jp.2015.4.
- Stefoski Mikeljevic, J; Haward, RA; Johnston, C; Sainsbury, R; Forman, D (2003): Surgeon workload and survival from breast cancer. *British Journal of Cancer* 89(3): 487-491. DOI: 10.1038/sj.bjc.6601148.
- Stroup, WW (2013): Generalized Linear Mixed Models. Modern Concepts, Methods and Applications. (Texts in Statistical Science). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-4398-1512-0.
- Townsend, Z; Buckley, J; Harada, M; Scott, MA (2013): The Choice between Fixed and Random Effects. Part I. Chapter 5. In: Scott, MA; Simonoff, JS; Marc, BD: *The SAGE Handbook of Multilevel Modeling* London, GB [u. a.]: Sage Publications, 73-88. DOI: 10.4135/9781446247600.n5.
- Tyson, JE; Parikh, NA; Langer, J; Green, C; Higgins, RD (2008): Intensive Care for Extreme Prematurity – Moving Beyond Gestational Age. *The New England Journal of Medicine* 358(16): 1672-1681. DOI: 10.1056/NEJMoa073059.
- Ulm, K (1991): A Statistical Method for Assessing a Threshold in Epidemiological Studies. *Statistics in Medicine* 10(3): 341-349. DOI: 10.1002/sim.4780100306.
- Urbach, DR; Austin, PC (2005): Conventional models overestimate the statistical significance of volume-outcome associations, compared with multilevel models. *Journal of Clinical Epidemiology* 58(4): 391-400. DOI: 10.1016/j.jclinepi.2004.12.001.
- Varaganam, M; Hutchings, A; Black, N (2015): Relationship between Patient-reported Outcomes of Elective Surgery and Hospital and Consultant Volume. *Medical Care* 53(4): 310-316. DOI: 10.1097/MLR.0000000000000318.
- Wetzel, H (2006): Mindestmengen zur Qualitätssicherung: Konzeptionelle und methodische Überlegungen zur Festlegung und Evaluation von Fallzahlgrenzwerten für die klinische Versorgung. *Zeitschrift für ärztliche Fortbildung und Qualität im Gesundheitswesen* 100(2): 99-106.
- Wood, SN (2017): Generalized Additive Models. An Introduction with R. Second Edition. (Texts in Statistical Science). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-4987-2833-1.
- Wood, SN (2019): mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation [*Open Source Software*]. R package version 1.8-30. Published: 24.10.2019.

Vienna, AT: R Foundation for Statistical Computing. URL: <https://CRAN.R-project.org/package=mgcv> [Old sources > mgcv archive > mgcv_1.8-30.tar.gz] (abgerufen am: 10.06.2020).