

3

Ermittlung der klinikspezifischen Ergebnisqualität der Behandlung von Früh- und Neugeborenen mit sehr niedrigem Geburtsgewicht (VLBW) auf der Basis von Routinedaten

Günther Heller

Abstract

In diesem Beitrag wird dargelegt, wie die Ergebnisqualität auf der Basis von Routinedaten erhoben und klinikspezifisch ausgewiesen werden kann. Dabei werden Analysen von insgesamt über 16 600 AOK-versicherten sehr untergewichtigen Früh- und Neugeborenen (VLBW) der Jahre 2003 bis 2009 erstellt und ein Vergleich der aktuell noch an der Regelversorgung teilnehmenden Level-1-Kliniken durchgeführt. Die Vor- und Nachteile von routinedatenbasierten Messmethoden im Vergleich zu Analysen auf Basis der seit 30 Jahren in Deutschland existierenden Neonatalerhebung werden diskutiert und Möglichkeiten einer künftigen sektorenübergreifenden Qualitätssicherung für VLBWs skizziert.

The article discusses how outcome quality for individual hospitals may be determined and reported by using routine data. This analysis is based on a total of over 16,600 AOK-insured very low birth weight infants (VLBW) born between 2003 and 2009. So-called level 1 hospitals which are currently still participating in standard care are benchmarked. The article discussed the pros and cons of measures based on the use of routine data in comparison to analyses based on the survey of neonatal care which has been carried out in Germany for the past 30 years and outlines possible future cross-sectoral quality assurance methods for VLBWs.

3.1 Einführung

Die Ermittlung der Behandlungsqualität von Frühgeborenen ist international wie auch national seit Jahrzehnten von großem wissenschaftlichen und öffentlichen Interesse. Um die Qualität der Versorgung von Frühgeborenen zu sichern, wurde in Deutschland bereits zu Beginn der 80er Jahre des letzten Jahrhunderts zunächst in Niedersachsen und Bayern eine Neonatalerhebung entwickelt und eingeführt, die

rasche Verbreitung fand und seit den 90er Jahren praktisch bundesweit flächendeckend eingesetzt wurde. Ausgehend davon hat die Bundesgeschäftsstelle für Qualitätssicherung (BQS) eine bundesweite Neonatalerhebung entworfen, die das AQUA-Institut nach Übernahme des sektorenübergreifenden Verfahrens nach § 137 SGB V aktuell umsetzt und weiterentwickelt (AQUA-Institut 2010).

Der zentrale Indikator für die Ergebnisqualität in der Versorgung von VLBWs ist die risikoadjustierte Neugeborenensterblichkeit oder neonatale Mortalität (Sterblichkeit von Lebendgeborenen innerhalb von vier Wochen nach der Geburt). Dabei existieren zahlreiche Ansätze, die Sterbe- bzw. Überlebenswahrscheinlichkeit der Neugeborenen in Abhängigkeit von unterschiedlichen Prognosevariablen zu einem Score zusammenzufassen und diesen Score anschließend als Basis für risikoadjustierte Modellbildungen zu nutzen.

Bekannte Beispiele hierfür sind z. B. der CRIB- bzw. CRIBII-Score (International Neonatal Network 1993; Parry et al. 2003), oder die SNAP- bzw. SNAPPE-Scores (Richardson et al. 1999; Richardson et al. 2001). Allerdings werden für die Berechnung der genannten Scores einige schwer zu erhebende physiologische Messwerte benötigt.

Ziel dieser Arbeit ist es, ein Prognose- bzw. ein Risikoadjustierungsmodell zu erläutern und zu diskutieren, das allein auf GKV-Abrechnungsdaten basiert. Nach Analyse des Modellfits werden zusätzlich klinikspezifische Ergebnisse dieser Berechnungen für die aktuell als Level-1-Kliniken noch an der Versorgung teilnehmenden Krankenhäuser dargestellt. Anschließend sollen Vor- und Nachteile dieses Modells gegenüber den surveybasierten Modellen und seine potenzielle Bedeutung bei der Weiterentwicklung neonatologischer Qualitätssicherungsinstrumente in Deutschland diskutiert werden.

3.2 Material und Methoden

Bundesweit wurden alle AOK-versicherten Kinder analysiert, die ein Aufnahme-gewicht zwischen 300 g und 1499 g und ein Lebensalter von 0 oder 1 Tag bei Aufnahme aufwiesen und zwischen dem 01.01.2004 und dem 30.09.2009 entlassen wurden. Wurde ein Kind mehrfach stationär behandelt, wurde nur der erste Krankenhausaufenthalt gezählt. Als Analyseendpunkt wurde der Überlebensstatus des Kindes innerhalb von 30 Tagen nach Aufnahme gewählt. Dies geschah, weil dieser Endpunkt bereits für andere Zwecke aufbereitet in unseren Daten vorliegt (Heller 2010). Als Informationsgrundlage dienen die Angaben, die die Krankenhäuser gemäß § 301 SGB V im Rahmen der Abrechnung an die gesetzlichen Krankenkassen übermitteln, wie auch anonymisiert verknüpfte Personenbestandsdaten zur Ermittlung des Überlebensstatus. Neugeborene mit einer kongenitalen Anomalie als Hauptdiagnose wurden ausgeschlossen.

Prognosemodelle wurden anhand von Angaben zu Geschlecht, Aufnahme-gewicht und weiteren kodierten diagnosebasierten Risikoadjustierungsvariablen erstellt. Dazu wurden logistische Regressionen genutzt. Zur Evaluation der Modell-güte wurden Receiver-Operating-Characteristic-Analysen und Hosmer-Lemeshow-Tests durchgeführt (Hosmer und Lemeshow 2000; Iezzoni 2003) und die Ergeb-

nisse anschließend mit anderen publizierten Werten verglichen. Die graphische Illustration eines Volume-Outcome-Zusammenhangs wurde mit einer lokalen gewichteten Regression (Lowess) dargestellt (Schnell 1994). Alle statistischen Analysen wurden mit Stata 10.0 durchgeführt (StataCorp 2007).

Auf der Basis dieses Modells werden klinikspezifische Standardisierte Mortalitäts-Ratios (SMR) ermittelt, indem die Anzahl der beobachteten verstorbenen Kinder einer Klinik durch die Anzahl der gemäß dem Prognosemodell zu erwartenden verstorbenen Kinder dividiert wird.

3.3 Ergebnisse

Tabelle 3–1 zeigt die Charakteristika der analysierten Population. Insgesamt wurden 16 607 Früh- und Neugeborene aus Kliniken in die Analyse eingeschlossen. Das durchschnittliche Aufnahmegegewicht betrug 1 074 Gramm (Quartilsgrenzen 828/1120/1350). 21,7% waren Mehrlinge. In 19,4% der Fälle war eine intrauterine Mangelentwicklung oder fetale Mangelernährung dokumentiert. Die 30-Tage-Sterblichkeit betrug 10,4% (Tabelle 3–1)

In die logistische Regression wurden nach inhaltlicher und statistischer Prüfung zahlreicher potenzieller Einflussfaktoren folgende Variablen mit aufgenommen: Entlassungsjahr, kategorisiertes Aufnahmegegewicht, Geschlecht, übergewichtige oder zu schwere Neugeborene, Hypoxie unter der Geburt, intrauterine Mangelentwicklung und fetale Mangelernährung. Dabei weist die Fläche unter der Receiver-

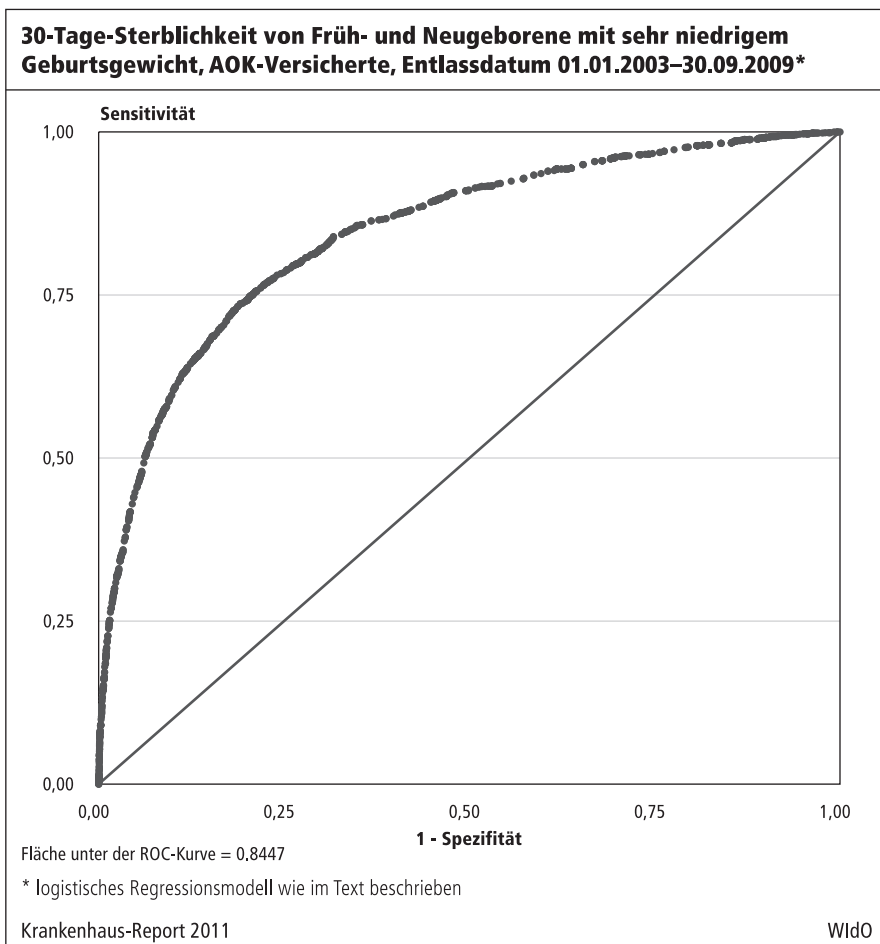
Tabelle 3–1

Charakteristika der Früh- und Neugeborenen*

	Anzahl	in %
Anzahl Früh- und Neugeborene	16 607	
Anzahl Kliniken	387	
Aufnahmegegewicht, Mittelwert	1 074	
Aufnahmegegewicht, Quartilsgrenzen	828/1 120/1 350	
Männliche Kinder	8 395	50,6
Mehrling	3 611	21,7
Oligohydramnion	291	1,8
Intrauterine Hypoxie	440	2,6
Hypoxie unter der Geburt	1 822	11,0
Intrauterine Mangelentwicklung und fetale Mangelernährung	3 230	19,4
Übergewichtige oder zu schwere Neugeborene	31	0,2
30-Tage-Sterblichkeit 1722	1 722	10,4

* AOK-versicherte Früh- und Neugeborene mit 300 bis 1 250 g Aufnahmegegewicht und einem Entlassdatum vom 01.01.2003 bis zum 30.09.2009

Abbildung 3–1

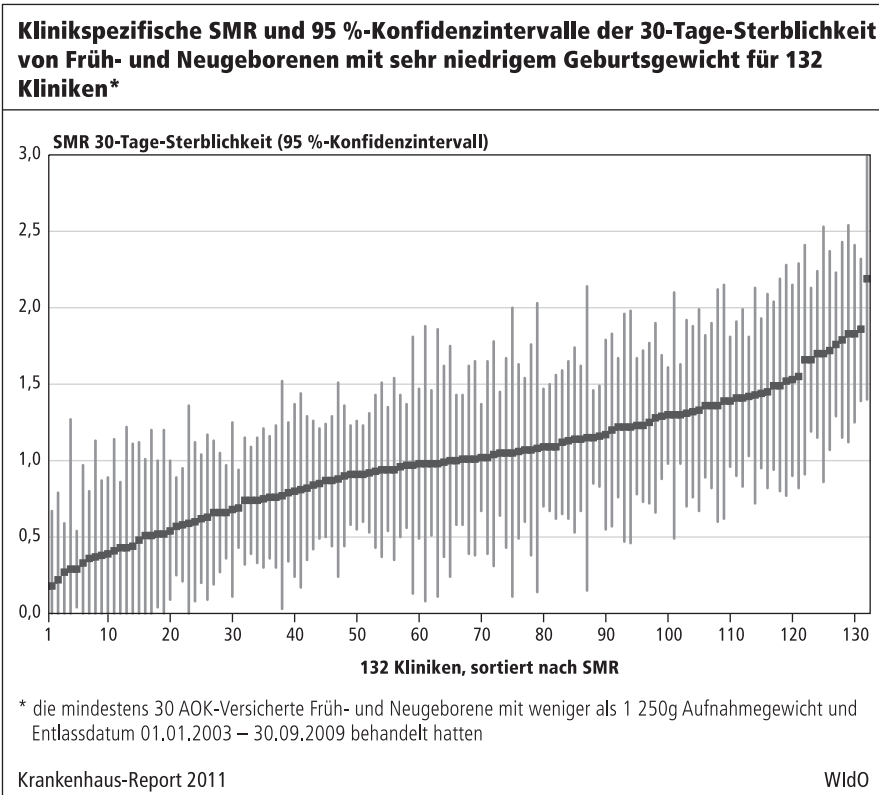


Operating-Characteristic-Kurve (ROC Kurve) mit etwas über 84% eine sehr gute Diskriminationsfähigkeit auf (Abbildung 3–1) und der Hosmer-Lemeshow Test mit einem p-Wert der Chi2-Statistik von 0,83 auf einen sehr guten Modellfit mit ebenso guter Übereinstimmung der vorhergesagten und geschätzten Werte hin.

Anschließend wurde für jede Klinik in Deutschland ein SMR berechnet. In dem hier analysierten Datensatz befanden sich 132 Kliniken, die im Untersuchungszeitraum mindestens 30 Kinder mit einem Aufnahmege­wicht von unter 1 250 g behandelt hatten, was in etwa 15 Kindern pro Jahr für alle VLBWs mit einem Gewicht unter 1 250 g entsprechen dürfte¹. Bei den Kliniken handelt es sich im Wesentlichen

1 In der hier analysierten Datenbasis befinden sich ohne Ausschlüsse etwa 16 600 Fälle. In Deutschland werden pro Jahr zwischen etwa 8 000–8 500 VLBWs geboren. Für eine approximative Hochrechnung der hier analysierten Kinder auf alle behandelten Kinder in Deutschland wurde daher die Anzahl der Kinder pro Klinik in unserer Datenbasis durch zwei dividiert.

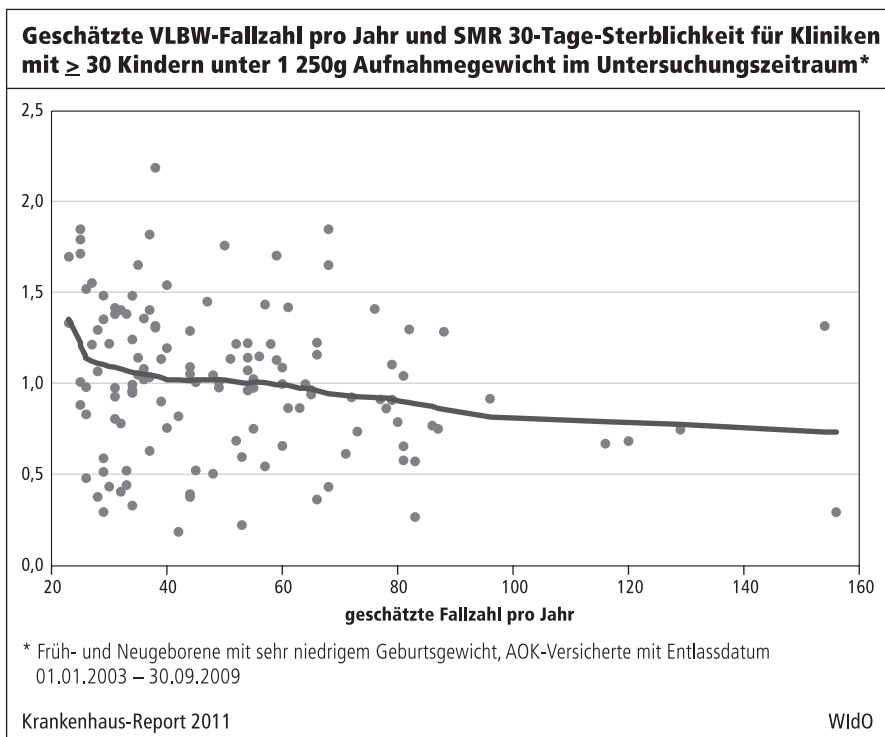
Abbildung 3–2



um die Kliniken, die bis Ende 2010 VLBWs behandeln dürfen (G-BA 2010) Dabei zeigt sich das kleinste SMR bei 0,18, das größte bei 2,19. 14 Kliniken weisen ein SMR von unter 1,5 auf, zehn Kliniken liegen mit ihrem 95 %-Konfidenzintervall oberhalb eines SMR von 1, wären also formal signifikant auffällig. Alle signifikanten auffälligen Kliniken weisen mindestens zehn Todesfälle auf (Abbildung 3–2).

In der Abbildung 3–3 ist ein Streudiagramm zwischen der hochgerechneten Fallzahl/Jahr und den klinikspezifischen SMR dargestellt. Ein möglicher Volume-Outcome-Zusammenhang wurde – analog zu dem Vorgehen bei Rogowski et al. 2004 – mittels einer Lowess-Regression analysiert. Dabei zeigt sich erwartungsgemäß eine deutliche Streuung der SMR mit der Fallzahl, aber auch ein Rückgang der geschätzten SMR mit steigender Fallzahl. Auffällig ist allerdings auch eine Klinik mit einer geschätzten Fallzahl von unter 150 VLBWs pro Jahr und einem SMR von 1,3.

Abbildung 3–3



3.4 Diskussion

In der vorliegenden Arbeit wurden erstmals klinikspezifische standardisierte Mortalitäts-Ratios (SMR) der 30-Tage-Sterblichkeit auf der Basis von Routinedaten und 16 607 AOK-versicherten Früh- und Neugeborenen mit sehr niedrigem Geburtsgewicht (VLBW) in Deutschland berechnet. Da die klinikspezifische Neugeborenensterblichkeit als der zentrale Ergebnisqualitätsparameter der Versorgung von Früh- und Neugeborenen mit sehr niedrigem Geburtsgewicht gilt, kann gleichzeitig ein zentraler Aspekt der Versorgungsqualität auf Basis von Routinedaten berechnet und ausgewiesen werden.

Dabei zeigt ein Blick auf die publizierten Gütekriterien anderer Prognosescores mitunter nochmals deutlich bessere Ergebnisse (International Neonatal Network 1993; Richardson et al. 1999; Richardson et al. 2001; Parry et al. 2003). Andererseits hatten frühere unveröffentlichte Arbeiten im Vergleich mit den Hessischen Neonataldaten der Jahre 2000–2005 gezeigt, dass ein verbesserter Modellfit des CRIBII-Scores in diesen Daten allein durch die Verwendung von Blutgasanalysewerten bei Aufnahme bedingt ist (Ergebnisse nicht dargestellt).

Dabei sollte bedacht werden, dass eine valide Dokumentation von Blutgaswerten bei Aufnahme ggf. ein kritisches Datum darstellt, weil sie einen hohen appa-

rativen Aufwand und eine sorgfältige zeitgerechte Messung voraussetzt und diese Messung in problematischen klinischen Situationen ggf. schwer zu bewerkstelligen ist. In der Konsequenz ist zu befürchten, dass diese Werte in den Situationen, wo sie mutmaßlich besonders prädiktiv sind, ggf. nicht valide und nicht homogen abgebildet sind. Daher kann gefragt werden, ob eine Verwendung von Blutgaswerten zur Prognosebestimmung, genauer zur Beurteilung einer klinikspezifischen Sterblichkeit, wirklich benötigt wird. Dies gilt umso mehr, als Bühler et al. 2008 unlängst demonstrierten, dass eine alleinige Verwendung des Geburtsgewichts im Prognosemodell dem CRIBII-Score ggf. auch überlegen sein kann. Dieses Ergebnis ist deswegen von Bedeutung, weil die genannten Angaben zu Blutgaswerten in den Routinedaten nicht enthalten sind. Werden unsere Ergebnisse mit den Angaben von Bühler et al. 2008 verglichen, bleibt auch anzumerken, dass die hier ermittelten Gütemaße des Modells insgesamt etwas besser ausfallen, wenn ein vergleichbares Analyseintervall (bei Bühler et al. 2002 bis 2006) betrachtet wird. Zusätzlich sei angemerkt, dass auch die unlängst veröffentlichten Daten des NICHD Neonatal Research Network (NRN) für Neugeborene unter 1 000g sich nur auf Geburtsgewicht, Gestationsalter, Mehrlingsstatus, Geschlecht und antenatale Kortikosteroidtherapie stützen (Tyson et al. 2008), also ebenfalls ohne Angaben zu Blutgaswerten auskommen. Nur der Vollständigkeit halber sei erwähnt, dass die antenatale Gabe von Kortikosteroiden in unseren Routinedaten nicht verfügbar ist und der Mehrlingsstatus, unter simulatiner Risikoadjustierung mit den anderen Risikoadjustierungsvariablen, in unseren Daten keinen signifikanten Einfluss mehr aufwies. Grundsätzlich könnte diskutiert werden, ob einzelne weitere Risikoadjustierungsvariablen in unserem Modell nicht ggf. entbehrlich sind und aufgrund der seltenen Dokumentation mutmaßlich eher zu mehr Verzerrungen als zu einer besseren Vergleichbarkeit führen. Es kann gezeigt werden, dass die Ergebnisse mit oder ohne solche Risikoadjustierungsvariablen in Bezug auf die Gütekriterien der Modelle wie auch in Bezug auf klinikspezifische SMR sehr ähnlich sind (Ergebnisse nicht dargestellt). Insgesamt kann festgehalten werden, dass sich das hier vorgestellte Prognosemodell auf der Basis von Routinedaten mit international publizierten Scores durchaus messen kann.

Ungeachtet dessen bleibt die Frage, ob ausgewiesene klinikspezifische Sterblichkeiten ausreichend stabile oder reliable Ergebnisse liefern oder ob diese aufgrund von geringen Fallzahlen im Wesentlichen vom Zufall abhängig sind (Dimick et al. 2004; Heller 2010). Vor diesem Hintergrund war die Einführung einer Mindestmenge durch den gemeinsamen Bundesausschuss insbesondere auch insofern sinnvoll, als nunmehr künftig – zumindest für Level-1-Kliniken – ein relevanter Prozentsatz der Kliniken sachgerecht beurteilt werden kann. Wendet man z. B. die Berechnung von Dimick et al. (2004) auf die Früh- und Neugeborenen mit einem Gewicht von unter 1 250 g an (Kriterium für Level-1-Kliniken), dann ergibt sich eine minimal benötigte Fallzahl von 40 bis 50. Diese Fallzahl sollte die überwiegende Anzahl der Level-1-Kliniken innerhalb eines Jahres, aber faktisch alle innerhalb von zwei Jahren erreichen. Auch wenn die Berechnung Mehrebeneneffekte oder Risikoadjustierung nicht berücksichtigt, scheint mit dieser Vorgabe eine sinnvolle Qualitätsmessung für Level-1-Kliniken möglich, die über historische Betrachtungen hinausgeht. Schon jetzt sind auf Basis der hier vorgelegten Analysen einzelne Kliniken identifizierbar, deren Ergebnisse erläuterungsbedürftig sind. Dazu zäh-

len beispielsweise die Ergebnisse der oben genannten Klinik mit mehr als 150 AOK-Fällen im Beobachtungszeitraum und einem SMR von 1,3, auch wenn hier das Konfidenzintervall gerade noch nicht signifikant ist (Grenze unteres 95%-Konfidenzintervall = 0,98).

Nun mag eingewandt werden, dass dies (fast) genau die Ergebnisse sind, die auch im Rahmen der verpflichtenden Veröffentlichung der Ergebnisqualität bereits aktuell zur Verfügung stehen (GBA-2010). Diesem Argument ist entgegenzuhalten, dass die dort veröffentlichten Daten keiner kritischen Prüfung in Bezug auf die Vollständigkeit der dokumentierten Fälle standhalten, schon gar nicht bezüglich der in der Klinik aufgetretenen Todesfälle. Analysen auf einer solch dubiosen Datenbasis sind angesichts der sensiblen Thematik u. E. keinesfalls tragfähig. Umso mehr stellt sich die Frage, warum entsprechende (oder vergleichbare) Angaben nicht auf der Basis von Routinedaten erfolgen. Die DRG-Dokumentation ist ab dem Jahr 2005 als vollständig anzusehen – es gibt kaum eine andere Dokumentation auf der Welt, die mehrere zehntausend Euro als Anreiz für eine Qualitätsdokumentation anbietet und umgekehrt in einem relevanten Prozentsatz durch den MDK geprüft wird. Die Nutzung von Routinedaten bei der Weiterentwicklung des Verfahrens scheint für diesen Leistungsbereich also besonders sinnvoll. Dabei mag ergänzt werden, dass die Sterblichkeit nur einen sehr groben Qualitätsindikator darstellt (Gerber und Rossi 2010). Dennoch scheint eine Nutzung der Routinedaten aufgrund der Vollständigkeit der dokumentierten Fälle wie auch der Todesfälle usw. ein unabdingbarer Schritt hin zu einer sektorenübergreifenden Weiterentwicklung des Verfahrens.

Zusammenfassend lässt sich festhalten: Die Ermittlung von risikoadjustierten Sterblichkeiten für Früh- und Neugeborene mit sehr niedrigem Geburtsgewicht auf der Basis von Routinedaten ist sinnvoll möglich. Es konnte gezeigt werden, dass Gütekriterien des Modellfits sich mit anderen publizierten Scores durchaus vergleichen lassen. Zudem sind vergleichsweise reliable klinikspezifische SMR für alle Level-1-Kliniken nach der künftigen Mindestmengenregelung zu erwarten. Da das aktuelle Verfahren der Ergebnisberichterstattung einem erheblichen Reporting Bias unterliegen dürfte, sollte es durch das hier vorgeschlagene Verfahren ersetzt oder zumindest modifiziert und in Bezug auf die Vollzähligkeit der Fälle und der Verstorbenen ergänzt werden.

3.5 Literatur

- AQUA-Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH. Allgemeine Methoden im Rahmen der sektorenübergreifenden Qualitätssicherung im Gesundheitswesen nach § 137a SGB V. Version 2.0. Stand 30. Juni 2010. Auftraggeber: Gemeinsamer Bundesausschuss. http://www.sqg.de/sqg/upload/CONTENT/Hintergrund/Methodenpapier/AQUA_AllgemeineMethoden_Version_2-0.pdf (03. September 2010).
- Bührer C, Metzke B, Obladen M. CRIB, CRIB-II, birth weight or gestational age to assess mortality risk in very low birth weight infants? *Acta Paediatr* 2008; 97 (7): 899–903.
- Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *JAMA* 2004; 292 (7): 847–51.

- Gemeinsamer Bundesausschuss (G-BA). Vereinbarung des Gemeinsamen Bundesausschusses gemäß § 137 Abs. 1 Satz 3 Nr. 3 SGB V für nach § 108 SGB V zugelassene Krankenhäuser (Mindestmengenvereinbarung), geändert am 17. Juni 2010 (Anlage 1), veröffentlicht im Bundesanzeiger 2010 S. 2840, Inkrafttreten am 1. Januar 2011
http://www.g-ba.de/downloads/17-98-2889/Mindestmengenvb_2010-06-17.pdf (05. September 2010).
- Gerber A, Rossi R. Neonatologische Versorgung, Fallzahlregelung. Einfluss auf Qualität und Finanzierung pädiatrischer Einrichtungen. *Monatschr Kinderheilk* 2010; 158: 356–63.
- Heller G, Günster C, Misselwitz M, Feller A, Schmidt S. Jährliche Fallzahl pro Klinik und Überlebensrate sehr untergewichtiger Frühgeborener (VLBW) in Deutschland – Eine bundesweite Analyse mit Routinedaten. *Z Geburtshilfe Neonatol* 2007; 211 (3): 123–31.
- Heller G, Schnell R. Hospital Mortality Risk Adjustment Using Claims Data. *JAMA* 2007; 297 (18): 1984.
- Heller G. Auswirkungen der Einführung von Mindestmengen in der Behandlung von sehr untergewichtigen Neugeborenen (VLBWs). Eine Simulation mit Echtdateien. In: Klauber J, Robra BP, Schellschmidt H (Hrsg). *Krankenhaus-Report 2008/2009*. Stuttgart/New York: Schattauer 2009; 183–99.
- Heller G. Qualitätssicherung mit Routinedaten – Aktueller Stand und Weiterentwicklung. In: Klauber J, Geraedts M, Friedrich J (Hrsg.): *Krankenhaus-Report 2010*. Stuttgart/New York: Schattauer 2010; 239–54.
- Hosmer DW, Lemeshow S. *Applied logistic regression*. 2. Auflage. New York: Wiley 2000.
- Iezzoni LI (Ed) *Risk Adjustment for Measuring Health Care Outcomes*, 3. Auflage, Chicago. Health Administration Press 2003.
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) Zusammenhang zwischen der Leistungsmenge und Ergebnis bei der Versorgung von Früh- und Neugeborenen mit sehr niedrigem Geburtsgewicht. Abschlussbericht V07-01. Köln 2008.
http://www.iqwig.de/download/V07-01_Abschlussbericht_Menge_und_Ergebnis_bei_der_Versorgung_von_Fruehgeborenen.pdf (05. September 2010).
- The International Neonatal Network. The CRIB (Clinical Risk Index for Babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. *Lancet* 1993; 342: 193–8.
- Parry G, Tucker J, Tarnow-Mordi W. UK Neonatal Staffing Study Collaborative Group. CRIB II: an update of the Clinical Risk Index for Babies score. *Lancet* 2003; 361 (9371): 1789–91.
- Richardson D, Tarnow-Mordi WO, Lee SK. Risk adjustment for quality improvement. *Pediatrics* 1999; 103 (1 Suppl E): 255–65.
- Richardson DK, Corcoran JD, Escobar GJ, Lee SK. SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores. *J Pediatr*. 2001; 138 (1): 92–100.
- Rogowski UJ, Horbar JD, Staiger G, Kenny M, Carpenter J, Geppert J. Indirect vs Direct Hospital Quality Indicators for Very Low-Birth-Weight Infants. *JAMA* 2004; 291: 202–9.
- Schnell R. *Graphisch gestützte Datenanalyse*. München: Oldenbourg 1994.
- StataCorp 2007. *Stata Statistical Software: Release 10*. College Station, TX: StataCorp LP.
- Tyson JE, Parikh NA, Langer J, Green C, Higgins RD. Intensive care for extreme prematurity: Moving beyond gestational age. *N Engl J Med*. 2008; 358: 1672–81

