

Auftragsnummer: | V07-01

Zusammenhang zwischen Leistungsmenge und Ergebnis bei der Versorgung von Früh- und Neugeborenen mit sehr geringem Geburtsgewicht

Eingangsstempel

Formblatt zur schriftlichen Stellungnahme zum Vorbericht

Name, Vorname; Titel und Funktion des/der Stellungnehmenden

1. Dr.med. Regina Klakow-Franck, M.A.

Leiterin Dezernat 3 und 4, Bundesärztekammer, Herbert-Lewin-Platz 1, 10623 Berlin

2. Dr.med. Hermann Wetzel, M.Sc.

Referent Dezernat 4, Bundesärztekammer, Herbert-Lewin-Platz 1, 10623 Berlin

3.

Die Abgabe der Stellungnahme erfolgt

im Namen folgender Institution / Organisation: Bundesärztekammer

als Privatperson(en)

Kontaktdaten

Name: Dr.med. Regina Klakow-Franck, M.A.

Institution: Bundesärztekammer

Anschrift: Herbert-Lewin-Platz 1

Tel./Fax-Nr.: 030 / 400456-435 / 030 / 400456-378

E-Mail-Adresse: dezernat4@baek.de

Bitte strukturieren Sie Ihre Stellungnahme wie folgt

1. Benennung von zusätzlichen, im Vorbericht nicht genannten, relevanten Studien,
2. Bewertung und Interpretation der im Vorbericht in die Nutzenbewertung eingeschlossenen Studien.

1. Benennung von zusätzlichen, im Vorbericht nicht genannten, relevanten Studien

Über die im Vorbericht genannten Untersuchungen hinaus sind von unserer Seite keine weiteren Publikationen für eine Auswertung zu benennen.

Allerdings möchten wir darum bitten, auch die UK Neonatal Staffing (UKNS)-Studie von Tucker et. al, Lancet 2002, Literaturreferenz Nr. 40, vollständig auszuwerten. Bei dieser Untersuchung handelt es sich um eine Studie, bei der sich die Daten zu Ergebnisqualitätsparametern wie z.B. Mortalität zwar nicht nur auf Neugeborene mit sehr niedrigem Geburtsgewicht unter 1.500 Gramm (VLBW) beziehen, sondern auf Säuglinge aller Geburtsgewichte, doch erfolgte die Einteilung der Fallzahlklassen explizit mit Bezug auf VLBW-Neugeborene. Die Beauftragung des G-BA an das IQWiG vom 19.07.2007 begründet seinem Wortlaut nach keinen Ausschluss der Ergebnisdaten der UKNS-Studie, da im Auftrag in einem weiten Sinn von Ergebnisqualität gesprochen und diese nicht ausschließlich auf VLBW-Neugeborene bezogen wird. Umfassende Aussagen zur Ergebnisqualität von Neonatalintensivstationen (NICUs) auch über die zur Definition von Fallzahlklassen herangezogene Indexpopulation von VLBW-Neugeborenen hinaus wären jedoch unseres Erachtens notwendig, damit der G-BA seine Entscheidung zu einer potentiellen Mengenregulation für NICUs vor dem Hintergrund der erforderlichen breiten Informationsbasis über das Versorgungsgeschehen in NICUs insgesamt treffen kann. Die Studie von Hamilton et al. (2007) umfasst zwar die Subgruppe der Neugeborenen mit sehr niedrigem Geburtsgewicht oder mit weniger als 31 Wochen Gestationsalter aus der UKNS-Studie, doch bezieht sich die Ergebnisanalyse lediglich auf Teilaspekte der UKNS-Studie. Damit der G-BA ein vollständiges Bild der vorhandenen Datenlage erhält, sollte folglich auch die Publikation von Tucker et al., Lancet 2002, ausführlich ausgewertet werden; eine mögliche Mengenregulierung durch den G-BA könnte schließlich nach den Ergebnissen der UKNS-Studie nicht nur Ergebnisqualitätsparameter bei VLBW-Neugeborenen betreffen, sondern sich auf die Ergebnisqualität von NICUs bei Säuglingen insgesamt auswirken.

2. Bewertung und Interpretation der im Vorbericht in die Nutzenbewertung eingeschlossenen Studien

Insgesamt handelt es sich bei dem vorliegenden Evidenzbericht um eine wissenschaftliche Analyse von hoher Qualität, die mit großer Sorgfalt durchgeführt wurde. Wir stimmen dem Fazit des Berichts, wonach – bei nicht einheitlichen Studienergebnissen – die Daten in der Gesamtschau bezüglich der primären Zielgröße Mortalität

Hinweise auf eine Risikoreduktion mit steigender Leistungsmenge bieten, zu, bitten jedoch hier um eine Präzisierung der Formulierung, wonach sich diese Hinweise auf eine statistische Assoziation beziehen. Da mit diesem Evidenzbericht im Kern die Frage beurteilt werden soll, ob sich Leistungsmengen und Fallzahlgrenzwerte für NICUs als Surrogatparameter für Qualitätsindikatoren eignen könnten, sollte u.E. ergänzend – vor dem Hintergrund der Ergebnisse der Studie von Rogowski et al. (2004) – im Fazit klar gestellt werden, dass selbst eine hoch signifikante statistische Assoziation zwischen Leistungsmenge und Ergebnisqualität eine ausreichende prädiktive Valenz einer Mindestmenge bei der Unterscheidung von guter und schlechter Ergebnisqualität offenbar nicht garantieren kann. Die Bewertung des Evidenzberichts, wonach eine eindeutige kausale Beziehung zwischen Fallzahl und Ergebnisqualität oder ein eindeutiger Schwellenwert, ab dem die Ergebnisqualität nicht weiter zunimmt, sich aus den Daten nicht ableiten lassen würden, halten wir für uneingeschränkt zutreffend. Zusammenfassend kann also eine evidenzbasierte Mindestmenge aufgrund der ausgewerteten Studien nicht festgelegt werden. Wir stimmen ausdrücklich der Einschätzung des IQWiG zu, dass die vorhandenen Hinweise auf eine statistische Assoziation zwischen Leistungsmenge und Ergebnisqualität bei VLBWs in NICUs keine Aussagen über die Ergebnisse einer Einführung einer Mindestmengenvereinbarung für den deutschen Versorgungskontext zulassen, und dass zur Klärung der Frage eines kausalen Zusammenhangs eigentlich Interventionsstudien zu dieser Fragestellung nötig wären.

Wir bitten die nachfolgenden Hinweise und Vorschläge zu berücksichtigen:

Einstufung der Studienqualität

Im Evidenzbericht wird eine Rubrizierung der ausgewerteten Studien hinsichtlich ihrer Qualität vorgenommen; höherwertigen Studien wird ein niedriges Verzerrungspotenzial zugeschrieben. Diese Studienqualitätseinstufung bezüglich des Verzerrungspotenzials in die Kategorien „hoch“ und „niedrig“ wird im Berichtsplan nicht erwähnt. Wir hatten in unserer Stellungnahme zum Berichtsplan gebeten, dort a priori Kriterien für eine Evidenzklassifikation oder für eine Beurteilung der Studienqualität zu benennen.

Explizite Ein- oder Ausschlusskriterien oder Regeln für eine polythetische Verknüpfung von Kriterien für diese ad hoc-Einstufung der Studienqualität fehlen auch im jetzigen Vorbericht.

U.E. ist die Bewertung der Studienqualität der einzelnen Publikationen nicht schlüssig. Beispielsweise wird der Untersuchung von Phibbs et al. (2007) eine hohe Studienqualität

bescheinigt, obwohl offenbar keine vollgültige Mehrebenenanalyse durchgeführt wurde, und obwohl die Leistungsmenge nicht als kontinuierliche Variable analysiert wurde. Vor dem Hintergrund der vom IQWiG selbst vorgenommenen Schwellenwertberechnungen zur Knie-TEP und zu koronarchirurgischen Eingriffen, in denen sich teils ein U-förmiger Kurvenverlauf der Mengen-Qualitäts-Beziehung und teils eine Abhängigkeit der Ergebnisse von der Durchführung einer Mehrebenenanalyse zeigten, erscheint es schwerlich begründbar, Studien ohne kontinuierliche Analyse der Leistungsmenge und ohne Mehrebenenanalyse eine hohe Studienqualität und ein niedriges Verzerrungspotenzial zu bescheinigen.

Bei der Studie von Hamilton et al. (2007) handelt es sich u.E. nicht um eine Untersuchung mit hohem Verzerrungspotenzial; wir möchten daher anregen, das Verzerrungspotenzial neu zu bewerten, da auch bei dieser auf VLBW-Neugeborene beschränkten Subanalyse der UKNS-Studie als einziger prospektiver Erhebung zum Ergebnisqualitätsindikator Mortalität diagnostische und therapeutische Parameter zum Zeitpunkt der Geburt oder 12 Stunden nach Geburt umfangreich in die Risikoadjustierung eingingen. Bei der Studie von Hamilton et al. (2007) handelt es sich also nicht – anders als im Vorbericht dargestellt – um eine retrospektive Registerstudie. In der Mutterstudie (Lancet 2002) wird eine GEE-Analyse explizit erwähnt, vermutlich wurde eine solche Mehrebenenanalyse auch in der Untersuchung von Hamilton et al. (2007) durchgeführt. Wir möchten anregen, die Erstautorin zu dieser Frage zu kontaktieren.

Nach unserer Auffassung handelt es sich bei den Untersuchungen von Tucker et al. (2002) und Rogowski et al. (2004) um die Studien mit der höchsten Studienqualität.

Unveröffentlichtes Manuskript von Heller (2008)

Die Studie von Heller et al. (2007) aus dem Wissenschaftlichen Institut der AOK wurde wegen unzureichender Adjustierung für Risikofaktoren nicht in die Liste der aussagekräftigen bewertungsrelevanten Studien eingeschlossen. Das im März 2008 nachgereichte unveröffentlichte Manuskript (Heller 2008), dessen Datengrundlage sich offenbar auch auf die Stichprobe der Krankenhausabrechnungsdaten von AOK-Patienten bezieht, und das im Gegensatz zur vorgenannten Untersuchung für eine Auswertung akzeptiert wurde, liegt uns nicht vor und entzieht sich daher jedweder Beurteilung. Wir können folglich zur Bewertung dieser Analysen aus dem Wissenschaftlichen Institut der AOK im Vorbericht, der dort vorgenommenen Einschätzung des Verzerrungspotentials und der möglichen Bedeutung der Studienergebnisse für den deutschen Versorgungskontext nicht bzw. nur sehr eingeschränkt Stellung nehmen. Im Gegensatz zu den übrigen bewertungsrelevanten Studien besteht hier – zumindest für den für die Erstellung

des Evidenzberichts zentralen Abschnitt des Stellungnahmeverfahrens – eine auffällige Intransparenz. Transparenz und Nachvollziehbarkeit der Studienbewertungen im einzelnen und des Bewertungsprozesses insgesamt sind jedoch für einen Evidenzbericht von entscheidender Bedeutung; es erhebt sich daher die methodische Frage, ob und wie diese Intransparenz bei der Bewertung der AOK-Daten im Evidenzbericht ihren Niederschlag finden soll. Falls von Seiten des IQWiG geplant sein sollte, zu den eingegangenen Stellungnahmen zum Vorbericht V07-01 eine Anhörung durchzuführen, bitten wir darum, Herrn Dr. Heller und die Leitung des Wissenschaftlichen Instituts der AOK, in dessen Verantwortungsbereich die Studie entstanden ist, zu ersuchen, für die Anhörung die Manuskriptfassung der Studie interessierten Teilnehmern zur Einsichtnahme zur Verfügung zu stellen.

Mehrebenenanalyse

Wir bitten um nochmalige Prüfung, ob in der Studie von Phibbs et al (2007) tatsächlich eine suffiziente Mehrebenenanalyse in dem Sinn durchgeführt wurde, dass die wesentlichen, sich aus der hierarchischen Datenstruktur ergebenden Informationen angemessen berücksichtigt worden wären. Bei Phibbs et al. (2007) wurde zur Datenanalyse die Cluster-Prozedur in STATA angewandt; für eine Mehrebenenanalyse i.e.S. sind nach unserer Kenntnis weitere bzw. andere Befehlsprozeduren nötig, nämlich xt-Kommandos wie xtreg, xtlogit, xtgee, xtmixed etc., oder – als Add-on – gllam.

Verweis auf deutsche Versorgungssituation

Im Vorbericht wird betont, dass für die deutsche Versorgungssituation die Studien mit deutschen Versorgungsdaten von Bartels et al. (2005, 2006, 2007a), Bartels et al. (2007b) und Heller (2008) eine besondere Bedeutung hätten. Es sollte dann allerdings zusätzlich darauf verwiesen werden, dass sich die Ergebnisse von Heller offenbar lediglich auf AOK-Patienten und die Daten von Bartels lediglich auf das Land Niedersachsen beziehen, weshalb deren Repräsentativität für den deutschen Versorgungsrahmen insgesamt kritisch hinterfragt werden sollte.

Im Evidenzbericht sollte im Hinblick auf die derzeitige deutsche Versorgungssituation außerdem darauf hingewiesen werden, dass seitens des G-BA seit Anfang 2006 Struktur- und Prozessqualitätsanforderungen für NICUs durch eine Richtlinie festgelegt wurden. Diese zwischenzeitlich eingeführten Struktur- und Prozessqualitätsverbesserungen sind wegen der jeweiligen Datenerhebungszeiträume in den Studien von Bartels gar nicht und in der Untersuchung von Heller (2008) offenbar nur rudimentär berücksichtigt.

Wir würden es begrüßen, wenn die AOK-Daten von Heller (2008) stratifiziert für die

Zeiträume vor und nach 2006 ausgewertet werden könnten, um mögliche Effekte der Einführung der G-BA-Richtlinie auf Ergebnisqualitätsparameter zumindest teilweise abbilden zu können.

Mögliche Periodeneffekte

Wir bitten bei der Bewertung zu berücksichtigen, dass alle Daten, die sich auf Erhebungszeiträume der Jahre 1990 bis ca. 2000 beziehen und insbesondere die Jahre 1990 bis 1995 betreffen, einem Periodeneffekt unterliegen könnten. Dies wurde in der auch im Bericht zitierten Studie von Horbar et al. (2002), Literaturreferenz Nr. 176, gezeigt. Es wäre daher von großem Interesse, hinsichtlich der Ergebnisse der Studien von Bartels et al. (2005, 2006, 2007a) und Phibbs et al. (2007) nachzuprüfen, ob sich die beobachteten Beziehungen zwischen Leistungsmenge und Ergebnisqualität auch singular für das späteste Datenerhebungsjahr sichern lassen, und ob zwischen dem jeweils ersten und letzten Jahr des Erhebungszeitraumes signifikante Unterschiede bezüglich der Mengen-Qualitäts-Beziehung bestehen.

Prognosekraft einer potentiellen Mindestmenge

Wir bitten bei der Bewertung der Studie von Rogowski et al. (2004) im Ergebnisteil und in der Diskussion stärker hervorzuheben, dass sich zwar eine hoch signifikante Assoziation zwischen Leistungsmenge und Mortalität fand, eine demgemäß in Höhe von 50 Fällen pro Jahr festgelegte Mindestmenge jedoch keine hinreichende positive prädiktive Valenz bezüglich der Ergebnisqualität (Mortalität) nachweisen konnte. Obwohl sich also anhand des indirekten „Qualitäts“-Indikators Leistungsmenge ein statistischer Zusammenhang mit der Ergebnisqualität zeigen ließ, hatte eine Einteilung von NICUs gemäß diesem Kriterium keine prognostische Kraft hinsichtlich der Ergebnisqualitäts-Zielgröße Mortalität. Dieser Befund stellt u.E. die Eignung der Festlegung einer Mindestmenge für die Verbesserung der Ergebnisqualität in NICUs in Frage.

Bei der Schilderung der Ergebnisse der Studie von Rogowski et al. (2004) in der Diskussion wird korrekt dargestellt, dass die NICU-Leistungsmenge sowie die anderen verfügbaren Krankenhauscharakteristika nicht mehr als 16% der Mortalitätsvariation erklärten und maximal 5% der Variation der nächsten Jahre voraussagen konnten. Wir bitten zu ergänzen, dass für die Fallzahl allein die jeweiligen Werte für die Varianzaufklärung lediglich bei 9% bzw. für die Vorhersagekraft bei nur 1% [!] lagen.

Literaturdarstellung

Unter den Befunden zur Leistungsmengen-Ergebnisqualitäts-Relation sollten auch die Ergebnisdaten der auf hohem methodischem Niveau durchgeführten Schwellenwerts-

berechnungen des IQWiG zur Knie-TEP und zu koronarchirurgischen Eingriffen erwähnt werden. Zu Recht wird dort u.a. die Bedeutung einer stetigen Modellierung der Fallzahl im Rahmen geeigneter Regressionsmodelle betont. Die durch die Schwellenwertberechnungen des IQWiG gesetzten methodischen Standards sollten daher bei der Bewertung des Verzerrungspotenzials von Studien entsprechend Berücksichtigung finden.

Bedeutung weiterer Ergebniszielgrößen neben der Mortalität

Im Vorbericht wird zu Recht auf die Bedeutung weiterer Ergebniszielgrößen neben der Mortalität verwiesen. In diesem Zusammenhang ist von Bedeutung, dass sich hinsichtlich der Morbiditätsvariablen nur inkonsistente Hinweise für eine Mengen-Qualitäts-Beziehung ergaben. Grundsätzlich sollte vor dem Hintergrund der Ergebnisse der IQWiG-Schwellenwertberechnungen zur Knie-TEP auf mögliche Zielkonflikte bei der Festlegung von Fallzahlgrenzwerten hingewiesen werden.

Mögliche unerwünschte Wirkungen einer Fallzahlregulierung

Zu Recht wird im Evidenzbericht auf die kritische Rolle von z.B. Transportzeiten verwiesen. Mögliche unerwünschte Wirkungen müssen bei der Einführung einer Mindestmengenregelung bedacht werden, um im Falle einer zu starken Konzentration Nachteile für ländliche Versorgungsräume zu vermeiden.

Pragmatischer Ansatz einer Fallzahlregulierung

Wir unterstützen die an Geraedts (2004) anknüpfenden Überlegungen des IQWiG zu einem pragmatischen Ansatz bei der möglichen Festsetzung von Fallzahlgrenzwerten, wonach anhand eines noch festzulegenden Perzentils der Häufigkeitsverteilung der Leistungsmenge eine Trennung der NICUs in Deutschland vorgenommen werden könnte. Durch eine solche „Regularitätsmenge“ als Fallzahlanforderung würde gewährleistet, dass NICUs nicht nur gelegentlich VLBW-Neugeborene versorgen; gleichzeitig ließe sich eine für die Patientenversorgung verträgliche Konzentration und Regionalisierung des Leistungsgeschehens erzielen. Wir unterstützen die Forderung des IQWiG, dass durch eine entsprechende Begleitforschung evaluiert werden müsste, ob nicht eine mögliche Fallzahlregulierung in Flächenländern zu weite Entfernungen zu Versorgungszentren mit inadäquaten postnatalen Transportwegen oder generell zu lange Anfahrts- und Wartezeiten oder andere Zugangshindernisse zur Folge haben könnte. Auch die Frage einer möglichen Überlastung großer Zentren durch eine Fallzahlregulierung müsste vor dem Hintergrund der Ergebnisse der UKNS-Studie (Tucker et al. (2002) bzw. Hamilton et al. (2007)) im Rahmen einer Begleitevaluation geprüft werden.