

Formblatt Stellungnahme zum Vorbericht

Auftrag des IQWiG/Projektbezeichnung mit Auftragsnummer:

Zusammenhang zwischen Menge der erbrachten Leistungen und der Ergebnisqualität für die "Perkutane Transluminale Coronare Angioplastie" – Vorbericht – Auftrag Q05-01B, Version 1.0

Name, Vorname; Titel und Funktion des/der Stellungnehmenden:

1. Klakow-Franck, Regina, Dr.med. M.A., Dezernentin, Dezernat Qualitätssicherung
2. Wetzel, Hermann, Dr.med. M.Sc., Referent, Dezernat Gebührenordnung
- 3.

Kontakt für weitere Nachfragen:

Institution: Bundesärztekammer
Anschrift: Herbert-Lewin-Platz 1, 10623 Berlin
Tel./Fax-Nr.: 030 / 400456 – 435 oder – 443
E-Mail-Adresse: Regina.Klakow-Franck@baek.de oder Hermann.Wetzel@baek.de

1. Folgende Originalstudien fehlen im Vorbericht

U.E. wurde die Studienlage zu möglichen relevanten Confoundern der Beziehung zwischen Leistungsmenge und Ergebnisqualität nicht hinreichend berücksichtigt. Dies bezieht sich insbesondere auf Struktur- und Prozessmerkmale von Krankenhäusern wie beispielsweise deren unterschiedliche Spezialisierung (siehe hierzu u.a.: Cram P, Rosenthal GE, Vaughan-Sarrazin MS (2005) Cardiac revascularization in specialty and general hospitals. N Engl J Med 352(14):1454-1462; diese Publikation wurde zwar zitiert, hinsichtlich ihrer potenziellen Relevanz aber nicht adäquat gewürdigt) oder unterschiedliche postinterventionelle Therapieansätze mit möglicher Bedeutung für Ergebnisqualitäts-Indikatoren (siehe hierzu z.B. Chen J, Radford MJ, Wang Y, Marcinak TA, Krumholz HM (1999) Do "America's Best Hospitals" perform better for acute myocardial infarction? N Engl J Med 340(4):286-292).

2. Folgende Bewertungen von Originalstudien im Vorbericht sind nicht korrekt (inkl. Begründung):

Datenextraktionsbögen mit detaillierten Einzelauswertungen

Die ausführlichen Auswertungsbögen für die ausgewerteten Studien sind leider nicht einsehbar. Die konsentierten Datenextraktionsbögen sollten um der Transparenz und der Nachvollziehbarkeit der Bewertungen willen – insbesondere im Hinblick auf die Beurteilungskriterien Datenqualität und –validität sowie Qualität der statistischen Analyse – zugänglich gemacht werden. Da die Einzelauswertungsbögen nicht veröffentlicht wurden, kann leider auch die Checkliste, in welcher aufgeführt wird, ob und welche Qualitätsmerkmale der Studien erfüllt sind, nicht eingesehen werden.

Einstufung von Studien bzgl. der Durchführungs- und Berichtsqualität

Eine Einstufung der Studien bzgl. der Durchführungs- und Berichtsqualität (siehe S. 32f) ist prinzipiell höchst wünschenswert; die vorgenommene dreistufige Rubrizierung ist für uns jedoch nur teilweise nachvollziehbar. Für die Einstufung einer Studie als "höchste Qualität" wäre es u.E. zusätzlich wünschenswert, dass die Leistungsmenge primär immer auch als kontinuierliche Variable modelliert wird, dass wechselseitig für Arzt- bzw. Krankenhaus-Fallzahlen adjustiert wird, dass auf Patientenseite Therapieparameter und auf Krankenseite Struktur- und Prozessvariablen bei der Analyse berücksichtigt werden, und dass eine angemessene Überprüfung von Datenqualität und –validität gewährleistet wurde.

Hinweise zu einzelnen Studien

Im Vorbericht wird auf S. 6 erwähnt, dass sich bei der Literaturrecherche lediglich retrospektive Datenauswertungen fanden. Rubartelli et al (2004) bezeichnen demgegenüber ihre Untersuchung als prospektive Studie.

Im Vorbericht wird auf S. 31 angegeben, in der Studie von Tsuchihashi et al (2004) werde der Anteil von Patienten mit akutem Herzinfarkt nicht angegeben. Demgegenüber ist der Publikation und dem Vorbericht auf S. 20 zu entnehmen, dass offenbar nur Patienten mit einem akuten Myokardinfarkt in die Studie eingeschlossen wurden, so dass die entsprechenden Angaben in Tabelle 4 jeweils auf "100%" lauten müssten.

3. Anmerkungen zur projektspezifischen Methodik (bei Änderungsvorschlägen einschließlich Begründung unter Angabe entsprechender wissenschaftlicher Belege):

Abgrenzung "elektiv" / "planbar" versus "dringlich" / "notfallmäßig"

Es sollte berücksichtigt werden, dass eine – zunehmend häufiger durchgeführte – PTCA / PCI-Behandlung bei Patienten nach akutem Myokardinfarkt im strengen Sinne nicht mehr als planbar bzw. elektiv angesehen werden kann. Ein dringlicher, innerhalb von wenigen Stunden notwendiger Eingriff oder gar ein notfallmäßiger, unverzüglich und ohne Aufschub durchzuführender Eingriff stellen u.E. keine planbaren Eingriffe i.e.S. mehr dar. Für

die Ergebnisdarstellung des Evidenzberichtes sollten entsprechend der Fragestellung nur diejenigen Untersuchungen herangezogen werden, bei denen Zielgrößen bei einer planbaren PTCA / PCI erhoben und ausgewertet wurden, oder bei denen eine Stratifikation bzw. Subgruppenanalyse der Zielgrößen nach der Dringlichkeit des Eingriffs möglich wäre, so dass eine Darstellung der Ergebnisqualität für planbare Eingriffe ohne Vermischung mit anderen Dringlichkeitsgraden erfolgen kann. Auch bei Berücksichtigung des Dringlichkeitsgrads eines Eingriffs in einer Risikoadjustierung würden sich die Ergebnisse auf die Gesamtheit der Prozeduren und nicht nur auf die Unterkategorie der planbaren Eingriffe beziehen.

Notwendigkeit stratifizierter Analyse hinsichtlich bestimmter Diagnose- bzw. Prozeduren-Untergruppen

Im Bericht wird festgehalten, dass der Anteil von Patienten mit einem akuten Herzinfarkt je nach Studie zwischen 9 % und 100 % lag, wobei die Krankenhäuser mit niedrigen Fallzahlen einen höheren Anteil an Patienten mit akutem Herzinfarkt hatten. Außerdem erfolgte eine Verwendung von Stents bei einem Anteil von Patienten zwischen 19 % und 86 %, woraus möglicherweise im Vergleich zu einer PTCA (ohne Einlegen eines Stents) eine unterschiedliche postinterventionelle Medikation resultiert. Daraus könnte sich ein möglicher Bias bei der Abschätzung der Ergebnisqualitäts-Indikatoren ergeben, wenn diese Unterschiede nicht mittels stratifizierter Analyse berücksichtigt wurden. Eine Berücksichtigung i.S.e. Confounders als Merkmal für eine Risikoadjustierung wäre u.E. aus den o.a. Gründen nicht ausreichend. – Zumindest sollten für die o.g. Diagnose- bzw. Prozeduren-Untergruppen (sowie ggf. auch hinsichtlich der Durchführung einer Mehrgefäßprozedur, welche in bis zu 62% der Fälle erfolgte) Sensitivitätsanalysen durchgeführt werden. Die auf S. 74 mit vier Zeilen geschilderte Sensitivitätsanalyse bzgl. des Merkmals unterschiedlicher Prozeduren PCTA vs. PCI sollte u.E. hinsichtlich Methodik und Ergebnissen gründlicher dargestellt werden.

Mögliche Periodeneffekte

Hinsichtlich des Vorgehens bei PTCA und PCI sollte geprüft werden, ob nicht mögliche Periodeneffekte und unterschiedliche Diffusionsraten bzgl. des Vorgehens zum Zeitpunkt der Datenerhebung mit Bezug auf Krankenhaus- und Arzt-Volumenkategorien die Gültigkeit der Ergebnisse und deren Übertragbarkeit auf den Versorgungskontext in Deutschland für den Zeitraum der Jahre ab 2007 einschränken könnten. Dies betrifft insbesondere auch die Häufigkeit der Verwendung von Stents ohne und mit Medikamentenbeschichtung.

Unterschiedliche Zielgrößen

Bei der Beurteilung der Zielgröße "Letalität" bzw. "Mortalität" sollten mögliche qualitative Unterschiede zwischen der 30-Tage- und der Krankenhaus-Letalität beachtet werden. Die nur in einer Studie (McGrath et al, 2000) angegebene 30-Tage-Letalität hat gegenüber der in den restlichen Untersuchungen verwendeten Hospital-Letalität den Vorteil, dass sie vermutlich weniger biasanfällig – etwa durch Verlegungen – ist.

Bei Verwendung von kombinierten Endpunkten (wie z.B. in Moscucci et al, 2005) sollte u.E. deren Validität und

klinische Wertigkeit im Vergleich zu nicht-zusammengefassten Zielgrößen gesondert kritisch diskutiert werden.

Mögliche Zielkonflikte zwischen verschiedenen Ergebnisqualitäts-Indikatoren – auch im Zeitverlauf – sollten gründlicher dargestellt und diskutiert werden (ggf. im Rahmen einer Sensitivitätsanalyse); sie sollten gleichfalls in der Zusammenfassung Erwähnung finden.

Adjustierung für intervenierende Variablen / Risiko-Adjustierung / Case-Mix-Adjustierung

Im Hinblick auf die Risikoadjustierung wäre es u.E. wünschenswert, mögliche Confounder bezüglich Arzt- und / oder Krankenhausmerkmalen wie z.B. Spezialisierung von Arzt und Pflegepersonal oder Struktur- und Prozessqualitätsmerkmale des Krankenhauses wie z.B. Personalschlüssel hinsichtlich ihrer Wertigkeit stärker zu berücksichtigen. Auch im Hinblick auf Patientenmerkmale ist die Adjustierung nicht vollständig; ein wichtiger ergebnisrelevanter Parameter wie die postinterventionelle Therapie wird zumeist nicht oder nicht hinreichend spezifiziert (z.B. Canto et al (2000): In dieser Studie wurde zwar die medikamentöse Therapie berücksichtigt, aber nur für die ersten 24 h nach Krankenhausaufnahme). – Es wäre außerdem bei jeder Studie erforderlich gewesen, dass wechselseitig für Krankenhaus- bzw. Abteilungsfallzahl und Leistungsmenge des Arztes adjustiert wird.

Bezüglich der Risikoadjustierung sollte genauer dargestellt werden, in welcher der Studien für die in den jeweiligen Datensätzen ausgewertete Patientenpopulation validierte logistische Risiko-Scores angewandt wurden, wie sie bei vergleichbaren Indikationsbereichen für alternative Therapieoptionen (wie z.B. bei Koronarbypass-Eingriffen) als etabliert gelten können. Vor dem Hintergrund der für Mindestmengen-Studien bei vergleichbaren Prozeduren wie CABG mittlerweile etablierten Standards sollte die Güte der Risiko- bzw. Case-Mix-Adjustierung als zentrales methodisches Qualitätskriterium generell ausführlicher und kritischer bewertet werden – also nicht nur im Hinblick auf ihre "Angemessenheit" bzgl. der (möglicherweise insuffizienten) Datenquellen.

Im Abschnitt 4.4.4, Sensitivitätsanalyse, wird betont, dass eine Adjustierung nach Krankenhaus-Strukturmerkmalen geplant gewesen sei. Dies hielten auch wir für substantiell notwendig. Es sollte dargestellt werden, in welcher der Studien welche Strukturmerkmale auf Seiten des Krankenhauses bei der Auswertung berücksichtigt wurden. Es sollte begründet werden, weshalb diese geplante Sensitivitätsanalyse hinsichtlich von Krankenhaus-Strukturmerkmalen nicht durchgeführt wurde, und welche Konsequenzen daraus für die Bewertung der Daten zu Ergebnisqualitäts-Indikatoren resultieren könnten.

Datenqualität / Datenvalidität

Im Bericht sollte generell mehr Augenmerk auf die Datenqualität (z.B. Vollständigkeit, Reliabilität, Datenclearing) und eine mögliche Datenvalidierung gelegt werden. Es sollte dargelegt werden, in welcher der ausgewerteten Studien eine Überprüfung von Datenqualität und Datenvalidität durchgeführt wurde. Auch sollte zur Problematik fehlender oder unplausibler Daten für jede der ausgewerteten Studien eine Einschätzung erfolgen. Fehlende Angaben hierzu rechtfertigen keinesfalls die Annahme, Datenqualität und Datenvalidität seien ausreichend, und

beschränken die Aussagekraft der betreffenden Studien in gravierender Weise.

Übertragbarkeit der Ergebnisse

U.E. sollte im Evidenzbericht zur Übertragbarkeit der vorwiegend aus den USA stammenden Daten auf den deutschen Versorgungskontext und zu potenziellen Ankerpunkten zu deren Beurteilbarkeit ausführlich Stellung genommen werden. Hierbei sollten auch mögliche Periodeneffekte berücksichtigt werden.

Beurteilung der Studienqualität insgesamt

Laut Evidenzbericht lassen sich bestimmte Studien anhand der Kriterien (1) Verwendung von Daten aus klinischen versus administrativen Registern, (2) ausführliche Adjustierung für relevante Diagnosen in der Vorgeschichte und für den Zustand des Patienten zum Zeitpunkt der Durchführung der Prozedur, (3) Berücksichtigung von Cluster-Effekten in der Analyse, (4) Angaben zur Modellgüte ohne Hinweise auf Mängel des Modells, und (5) gute Berichterstattung der Ergebnisse mit einer dreistufigen Einteilung klassifizieren. Eine Charakterisierung der Studienqualität anhand weiterer zusätzlicher Kriterien hielten wir für wünschenswert (siehe hierzu die Angaben unter Abschnitt 2); die Qualitätseinordnung sollte anhand der detaillierten Einzelauswertungsbögen im einzelnen nachvollziehbar sein. – Hierbei sollte auch berücksichtigt werden, inwiefern ein angegebener Schwellenwert tatsächlich zwischen schlechter und guter Qualität im Hinblick auf eine relevante Ergebniszielgröße diskriminieren könnte. Keine der ausgewerteten Studien untersucht die prädiktive Valenz eines potentiellen Mindestmengen-Kriteriums im Hinblick auf einen Ergebnisqualitätsindikator wie z.B. die Krankenhaus- oder die 30-Tages-Letalität.

Generell wäre anzuregen, dass eine ordinale Einordnung der Studien in Evidenzstufen gemäß der gültigen G-BA-Verfahrensordnung, den Oxford CEBM Levels of Evidence, dem SIGN 50 Grading System, oder gemäß GRADE vorgenommen wird.

Zusammenfassung der Ergebnisse

U.E. zeigt sich zwar in den meisten retrospektiven Untersuchungen zum Zusammenhang zwischen Krankenhaus-Leistungsmenge und insbesondere der Krankenhaus- bzw. 30-Tage-Letalität als Ergebnisqualitäts-Indikator eine statistisch signifikante negative Assoziation, doch ist dieser Befund aufgrund eines möglichen Bias durch nicht hinreichend berücksichtigte potenzielle Confounder mit einem beträchtlichen Maß an Unsicherheit behaftet. Auch war dieser Befund offenbar nicht konsistent, wenn neben kurzfristigen postinterventionellen Komplikationen auch mittelfristige Verlaufsbeobachtungen berücksichtigt wurden (Kimmel et al, 2002). Bezüglich des Zusammenhangs zwischen Arzt-Prozedurenmenge und der Krankenhaus- bzw. 30-Tage-Letalität als Ergebnisqualitäts-Indikator ergab sich kein einheitliches Bild. In einer Studie (McGrath et al, 2000) wurde bei allerdings nicht hinreichender Risiko- und Case-Mix-Adjustierung über einen signifikanten inversen Zusammenhang zwischen Arzt-Leistungsmenge und Notwendigkeit einer CABG während des gleichen Krankenhausaufenthalts berichtet.

Insgesamt sollte auch in Zusammenfassung und Fazit explizit Bezug genommen werden auf die vielfältigen, in den vorherigen Abschnitten des Berichts erwähnten Einschränkungen im Hinblick auf eine Vielzahl methodischer

Kritikpunkte, welche Bedeutung und Aussagekraft der gefundenen Assoziationen erheblich limitieren.

Hinsichtlich des Zusammenhanges zwischen Arzt-Leistungsmenge und Ergebnisqualität werden in den Studien unterschiedliche Prozedurenmengen-Kategorien verwendet, u. a. weil die Kategorienbildung auf Perzentilen basiert. Dieses Vorgehen erscheint insofern fragwürdig, als die Schwellenwertermittlung damit von Zufälligkeiten der Datenerhebung und der Datenverteilung abhängig gemacht wird und nicht nach dem möglicherweise trennschärfsten Schwellenwert gesucht wird. Die in diesen Studien gefundenen unterschiedlichen Schwellenwerte sollten deskriptiv dargestellt werden, um die inkonsistente Datenlage zu kennzeichnen.

U. E. sollten Daten im Ergebnisabschnitt und in der Zusammenfassung unbedingt getrennt für planbare Eingriffe und für die Gesamtheit der Eingriffe angegeben werden. In beiden Abschnitten des Berichts sollten die Ergebnisse und Bewertungen von Subgruppen- und Sensitivitätsanalysen ausführlicher dargestellt werden. Bei der Bewertung von kombinierten Arzt- und Krankenhaus-Leistungsmengen sollten die relevanten methodischen Einschränkungen hinsichtlich dieser wenigen Untersuchungen auch in der Zusammenfassung miterwähnt werden.

Der Einschätzung der Verfasser, dass durch keine der ausgewerteten Studien ein Schwellenwert ermittelt werden kann, und dass keine der Studien mögliche positive oder negative Effekte der Einführung einer Mindestmenge nachweisen kann, stimmen wir ausdrücklich zu. Die Gründe hierfür sollten jeweils noch näher spezifiziert werden. Im Kern wäre hierzu der Nachweis einer kausalen Beziehung zwischen Leistungsmenge und Ergebnisqualität durch eine prospektive Interventionsstudie nötig.

Angesichts der mit vielen Unsicherheiten belasteten Ergebnisse retrospektiver Studien vorwiegend aus den USA und deren allenfalls partieller Übertragbarkeit auf den aktuellen deutschen Versorgungskontext liegt es u. E. nahe, für den Fall, dass in Deutschland eine Mindestmenge beim planbaren Eingriff PTCA bzw. PTCA / PCI definiert werden soll, eine prospektive Datenerhebung für verschiedene patientenrelevante Ergebnisqualitäts-Indikatoren einschl. der Lebensqualität zu fordern, z. B. durch einen Vergleich von Ergebnisdaten unter hypothetischer Annahme eines (pragmatisch festgelegten) Mindestmengen-Kriteriums gegenüber dem Status quo im Rahmen einer Begleitforschung analog zu § 21 Abs. 4 der Verfahrensordnung des G-BA. Eine weitere Option bestünde darin, aus den Ergebnissen der BQS-Datenerhebung zur (elektiven) PTCA bzw. PCI aus den Jahren 2004 und 2005 (ggf. einschl. 2006) mittels der vom IQWiG entwickelten Methodik über eine Modellrechnung mögliche Schwellenwerte abzuleiten, wobei die BQS-Daten ggf. gegenüber der Datenbank zu Koronarinterventionen der "Deutschen Gesellschaft für Kardiologie" validiert werden könnten.

Bei einer prospektiven Untersuchung zur Definition von Mindestmengen beim planbaren Eingriff PTCA bzw. PTCA / PCI könnte vermutlich auf bereits etablierte Strukturen der externen Qualitätssicherung der "Deutschen Gesellschaft für Kardiologie" zurückgegriffen werden. Eine solche Studie könnte das jetzige Vorgehen beim Eingriff PTCA bzw. PTCA / PCI einschl. der Verwendung von (Drug-eluting) Stents adäquat abbilden und klären, ob möglicherweise Zielkonflikte bzw. trade-offs hinsichtlich verschiedener Ergebnisqualitätsindikatoren bestehen.

Kriteriengestützte Beurteilung für das Vorliegen eines Zusammenhangs "in besonderem Maße"

Im Evidenzbericht sollten nicht nur Begründungs- und Geltungsfragen analysiert, sondern auch Anwendungszusammenhänge berücksichtigt werden. Eine Bewertung des Zusammenhangs zwischen Leistungsmenge und Ergebnisqualität bei der PTCA / PCI erfordert folglich auch eine kriteriengestützte Beurteilung, ob die nach der Literatur mögliche Beziehung tatsächlich als Grundlage für eine in Deutschland für das Jahr 2007 zu beschließende Mindestmenge dienen kann. Es sollte also dargelegt werden, ob der Zusammenhang zwischen Leistungsmenge und Ergebnisqualität derart ausgeprägt ist und mit einem derart geringen Ausmaß an Unsicherheit angenommen werden kann, dass die nach SGB V notwendige Qualifikation erfüllt ist, wonach ein Zusammenhang "in besonderem Maße" besteht. Mögliche Kriterien hierfür wären z.B. Stärke und Konsistenz der Beziehung, Ausmaß der Varianzerklärung des Faktors Leistungsmenge für die Ergebnisqualität, Spezifität des Effekts, monoton verlaufende Mengen-Qualitäts-Kurve, Berücksichtigung möglicher intervenierender Variablen bzw. Ausschluss alternativer Erklärungen und anderer Bias-Faktoren sowie Beachtung von Periodeneffekten.

Substanzielle Stellungnahmen, die der Form genügen und rechtzeitig eingereicht werden, finden Berücksichtigung. Anhand folgender Checkliste können Sie prüfen, ob Ihre Stellungnahme die formalen Kriterien erfüllt. Bitte beachten Sie darüber hinaus die Hinweise im „Leitfaden zur Abgabe von Stellungnahmen“.

Der Umfang der Stellungnahme umfasst maximal **6 DIN A4-Seiten**.

Für alle Zitate sind **Originalpublikationen (Volltext)** beigelegt.

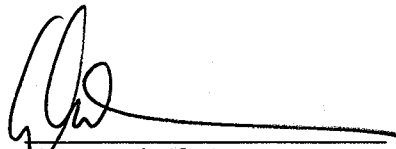
Das Formblatt zur Stellungnahme ist vollständig ausgefüllt und **von allen Stellungnehmenden** unterschrieben.

Das **Formblatt zur Darlegung potenzieller Interessenskonflikte** liegt **für alle Stellungnehmenden** einzeln **ausgefüllt** vor.

(Anm.: Dies gilt auch für Stellungnahmen von Institutionen / Gesellschaften)

Mir / uns ist bekannt, dass alle Stellungnahmen im Internet veröffentlicht werden können. Einer solchen Veröffentlichung stimme ich / stimmen wir mit meiner / unseren Unterschrift(en) ausdrücklich zu.

Berlin, den 13.03.06
Ort/Datum


Unterschrift 1


Unterschrift 2

Unterschrift 3

Vielen Dank für Ihre Stellungnahme.