



DGOU Deutsche Gesellschaft für
Orthopädie und Unfallchirurgie

**Angeforderte Stellungnahme der Deutschen Gesellschaft für Orthopädie
und Unfallchirurgie (DGOU) e.V.**

**Methodenkritik
Entwurf für Version 4.2 vom 18.06.2014
„Allgemeine Methoden“
Institut für Qualität und Wirtschaftlichkeit im
Gesundheitswesen (IQWiG)**

Auftraggeber

Deutsche Gesellschaft für Orthopädie und Unfallchirurgie (DGOU) e.V.

Vertreten durch den Generalsekretär
Prof. Dr. med. Fritz-Uwe Niethard
per E-Mail am 11.07.2014

Auftragnehmer

Methodisches Beratungszentrum der Deutschen Gesellschaft für
Unfallchirurgie (DGU) e.V.

Vertreten durch
Priv.-Doz. Dr. med. Dirk Stengel, MSc(Epi)
Ärztlicher Leiter des Zentrums für Klinische Forschung,
Klinik für Unfallchirurgie und Orthopädie, Unfallkrankenhaus Berlin, Warener
Str. 7, 12683 Berlin,
Tel ++49 30 5681 4050, Fax ++49 30 5681 4053, dirk.stengel@ukb.de

Inhalt

1. Position der DGOU	3
2. Gegenstand des Kommentars.....	4
2.1. Allgemeines.....	4
2.2. Ergebnissicherheit.....	5
2.3. Fristen	5
2.4. Diagnostische Verfahren	7
2.5. Potenzialbewertung	9
2.6. Kosten-Nutzen-Bewertung.....	10
2.7. „Rangordnung verschiedener Studienarten / Evidenzgrade“	11
3. Empfehlungen der DGOU	12
4. Literatur.....	13

1. Position der DGOU

Die DGOU steht hinter einer transparenten und objektiven Nutzenbewertung im Einklang mit den Empfehlungen des G-BA und IQWiG

Das Methodenpapier des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) bildet (in allgemein verständlicher Auslegung der gesetzlichen Vorgaben) die Grundlage für die Bewertung des

- i. *Nutzens* etablierter und auf dem deutschen Markt zugelassener Diagnostika, Pharmaka, Medizinprodukte und sonstiger medizinischer Interventionen, welche der (frühzeitigen) Erkennung von Gesundheitsstörungen und deren Behandlung im Interesse der Betroffenen, der Behandler, aber auch der Solidargemeinschaft dienen sollen
- ii. *Zusatznutzens*, welcher sich aus einem (offensichtlich) günstigen Verhältnis zwischen Wirkung und Wirksamkeit, Sicherheit und Kosten neuer im Vergleich zu etablierten Methoden ergeben kann.

Nutzen und Zusatznutzen sind schwer zu trennen. Legt man den von Prof. Dr. Jürgen Windeler vorgeschlagenen Minimalkonsens zugrunde, ergibt sich der medizinische Nutzen einer Gesundheitsmaßnahme aus deren i. medizinischem, ii. positivem und iii. kausalem Effekt (i.S. einer günstigen Beeinflussung des Krankheitsverlaufes) in mehr als geringfügigem Ausmaß. Einfach kann man auch von Netto- und Brutto-Nutzen sprechen, je nachdem, ob unerwünschte Wirkungen in die Kalkulation einbezogen werden oder nicht.

Der Nutzenbegriff wird von Konsumenten, Anbietern und Kostenträgern unterschiedlich definiert. Hinzu kommen verschiedene klinische und methodische Perspektiven und Sozialisationen, unterschiedliche Definitionen von Relevanz, differierende Annahmen über akzeptable Belastungen des Gesundheitsbudgets, methodische Ansätze und neue methodische Strömungen, Umwälzungen im Sozial- und Gesellschaftsgefüge usw. Jeder Methodenvorschlag muss und sollte daher Debatten anregen, ohne dass als Endprodukt ein für alle Beteiligten akzeptables Gesamtwerk erwartet werden darf.

Die Deutsche Gesellschaft für Orthopädie und Unfallchirurgie (DGOU) e.V. geht mit den o.g. Definitionen des Nutzens bzw. Zusatznutzens medizinischer Maßnahmen grundsätzlich konform.

Die DGOU ist die Dachorganisation und berufspolitische Vertretung von Ärztinnen und Ärzten, welche sich der Prävention, nicht-operativen und operativen Therapie, Rehabilitation und Erforschung von Erkrankungen und Verletzungen des Bewegungsapparates verschrieben haben. Die DGOU sieht sich zunehmend mit Decken- und Bodeneffekten und den daraus resultierenden Schwierigkeiten in der Empfehlung bestimmter Diagnose- und / oder Therapieschritte (z.B. im Rahmen Evidenz- und Konsens-basierter Leitlinien) konfrontiert.

Typische Beispiele (willkürlich ausgewählt aus einer großen Bandbreite potentieller Innovationen) sind:

- i. Der (Zusatz-)Nutzen der minimal-invasiven und / oder Computer-assistierten (navigierten) Endoprothetik des Knie- oder Hüftgelenkes im Vergleich zum konventionellen Gelenkersatz
- ii. Der (Zusatz-)Nutzen einer bioaktiven Beschichtung extra- und intra-medullärer Implantate zur Stabilisierung diaphysärer Frakturen der langen Röhrenknochen im Vergleich zu unbehandelten Standardimplantaten

Die kritische Hinterfragung potenzieller und scheinbarer Innovationen unter dem Nutzenschirm, aber auch die Auseinandersetzung mit modernen Methoden der Nutzenbewertung gehören damit zu den vornehmlichen Aufgaben der DGOU.

2. Gegenstand des Kommentars

2.1. Allgemeines

Auf S. 3 (ii) des Methodenpapiers wird (in fetter Schrift) betont, dass „nur die (Anm.: mittels gelber Markierung bzw. Textlöschung) als Änderung gekennzeichneten Textteile Gegenstand des Stellungnahmeverfahrens sind.“ Obwohl dies auf einer rechtlichen Grundlage beruht und die Arbeit der methodischen Kommentatoren grundsätzlich erleichtern könnte, verbietet es gleichsam, antiquierten und überflüssigen Ballast (s. insb. die Ausführungen 1.2 Evidenzbasierte Medizin [EbM]) aus dem Methodenpapier zu entfernen. Dieser Kommentar wiederum bezieht sich ausschließlich auf *substanzielle* Änderungen und Ergänzungen im Methodenpapier.

Als solche werden definiert

- Streichungen, Änderungen, Ergänzungen oder der komplette Austausch von Passagen oder Kapiteln über mehrere Zeilen, Absätze oder Seiten
- Inhaltliche Änderungen, welche klar zu einer neuen oder andersartigen Interpretation von Tatsachen und / oder Konsequenzen im Vergleich zum vorherigen Report führen.

Nicht kommentiert werden Wort- oder einzelne Satzergänzungen, sofern sich hierdurch keine grundlegenden Änderungen der Aussagen ergeben.

Die DGOU akzeptiert ausschließlich durch wissenschaftliche Daten begründete Diagnose- und Therapiemaßnahmen und setzt damit ein klares Zeichen gegen Beliebigkeit und Willkür in der medizinischen Versorgung der Bevölkerung.

Gleichzeitig distanziert sich die DGOU von den apodiktischen, mittlerweile als veraltet und praxisfern geltenden Vorgaben der EbM, wie sie noch im kanadischen und britischen Raum, aber auch durch das deutsche Netzwerk EbM (DNEbM) gehandhabt, propagiert und gelehrt werden.

Die DGOU unterstützt und fördert hingegen Initiativen und Aktivitäten, welche dem Praxis- und damit Patienten-zentrierten Wissens- und Erkenntnisgewinn dienen. Dies umfasst explizit auch innovative biometrische bzw. biostatistische Ansätze, um das bisherige Arsenal klinisch-epidemiologischer Forschungsmethoden zu bereichern.

2.2. Ergebnissicherheit

Die kurze Ergänzung in 1.2.5 (Anforderungen an die Präzision und Verlässlichkeit), Punkt 4 (S. 8), dass der G-BA die „Möglichkeit hat, selbst Studien zu fordern oder zu *initiiieren*, um die Evidenzlücken zu schließen“ ist von weitreichender Bedeutung. Wenn nämlich der G-BA Studien initiiert, muss er auch Sorge für deren Finanzierung tragen. Tatsächlich dürfen nicht i.S. einer selbsterfüllenden Prophezeiung klinische Studien gefordert werden (von denen es nie genug geben kann), ohne nicht auch deren finanziellen und logistischen Unterbau zu garantieren. *Die DGOU stützt daher diese kleine, aber wesentliche Ergänzung in jeder Hinsicht, erwartet aber konkrete und nachhaltige Entscheidungen über eine optimierte Finanzierung klinischer Studien aus öffentlicher Hand.*

Sowohl die Kosten-Nutzen-Bewertung nach §35b SGB V als auch die Potenzialbewertung sind gesetzlich vorgeschrieben und daher seitens der Fachgesellschaft kaum zu beeinflussen. Die Ergänzungen auf den Seiten 12 – 18 sind unkritisch und selbsterklärend.

In den in Abb. 4 (S. 23) illustrierten Ablaufplan der Kosten-Nutzen-Bewertung sollten speziell ausgewiesene und benannte Vertreter der betroffenen Fachgesellschaften integriert werden.

Der Passus in 2.1.4 (S. 22) „... Dabei werden regelhaft externe Sachverständige beteiligt ...“ ist zu vage. Erfahrungen der Vergangenheit zeigen, dass die Kriterien für die Auswahl der Sachverständigen und deren Qualifikation unbestimmt waren. Die DGOU fordert daher, im Falle einer Kosten-Nutzen-Bewertung nach §35b SGB V mindestens *zwei unabhängige, von einer Fachgesellschaft akkreditierte Sachverständige* in den Arbeitsprozess zu integrieren. Diese Forderung ergibt sich aus dem exponentiellen Wissenszuwachs und den Spezialisierungstendenzen innerhalb der verschiedenen Fachdisziplinen.

Der auf den Seiten 24 – 26 geschilderte Prozess, Auftragsinhalte zu konkretisieren und mit verschiedenen Stakeholdern abzustimmen, ist grundsätzlich klar.

Die Rolle der Vertreter der Fachgesellschaften sollte aber noch deutlicher herausgearbeitet werden. Eine allgemeine medizinische Expertise ist nicht mehr geeignet, um Empfehlungen des IQWiG und / oder des G-BA mit Einfluss auf die Versorgungslage zu beeinflussen.

Die Sachverständigendatenbank sollte in Abstimmung mit den Fachgesellschaften aktualisiert, angepasst und ggf. bereinigt werden.

2.3. Fristen

Die dreiwöchige Frist bis zur Abgabe einer Stellungnahme zum *Vorbericht* durch die „Öffentlichkeit“ (sind hiermit Konsumenten, Anbieter oder Kostenträger gemeint?) ist und bleibt indiskutabel. Aufgrund des üblichen Volumens der Vorberichte bei gleichzeitig nachhaltiger Bedeutsamkeit für die medizinische Versorgung muss aus Sicht der DGOU die Frist auf wenigstens sechs Wochen ausgedehnt werden, um adäquate Kommentare und Änderungsvorschläge zu unterbreiten.

Die auf S. 24 geschilderten Fristen sind etwas unklar. Der Öffentlichkeit soll eine Frist „von drei Wochen zur ... Stellungnahme (Anhörung) [zum Vorbericht] gegeben werden.“ Der Abschlussbericht wiederum soll dem G-BA „innerhalb von drei Monaten nach Einleitung des Stellungnahmeverfahrens zum Vorbericht“ übermittelt und muss innerhalb dieses Intervalls auch kommentiert werden.

Die DGOU empfiehlt dem IQWiG, ein zeitgenaues Ablaufschema der Einreichungs- und Kommentierungsintervalle zu entwerfen, um industriellen und nicht-industriellen Antragsstellern die notwendigen Zeithorizonte zu verdeutlichen.

Das Verfahren zur Einreichung des Abschlussberichtes (2.1.4) klammert die Fachgesellschaften bisher aus. Sollte dieses Verfahren gesetzeskonform sein, stimmt die DGOU dem Vorgehen zu. Eine unabhängige juristische Prüfung erscheint jedoch aufgrund der erheblichen Konsequenzen des Abschlussberichtes als Grundlage von und für Entscheidungen des G-BA zwingend notwendig.

Sollte es im Ablaufschema zwischen i. erster Vorberichtsphase, ii. überarbeitetem Vorbericht dank Stellungnahme der Fachgesellschaften, iii. erster Abschlussberichtsfassung und iv. Entscheidungsfindung des G-BA, des Stiftungsvorstandes, -rates und / oder -kuratoriums zu Verzögerungen oder diskrepanten Stellungnahmen kommen, muss erneut ein Vertreter / eine Vertreterin der jeweiligen Fachgesellschaft als Moderator/in, Schlichter/in oder Ombudsmann / Ombudsfrau benannt werden, um den Konflikt zu lösen.

Die institutseigene Sachverständigendatenbank sollte mit der Fachgesellschaft abgestimmt werden. Zu benennen sind vorzugsweise Sachverständige, welche sich durch aktuelle klinische oder sonstige Expertise auf dem entsprechenden Fachgebiet auszeichnen.

Die in 2.1.5 geschilderte Potenzialbewertung kommt im Vergleich zu anderen Passagen des Methodenpapiers etwas kurz. Sofern der interne Qualitätssicherungsprozess dargestellt werden darf, sollte er in das überarbeitete Dokument aufgenommen werden.

Die auf S. 39 genannten Zielgrößen Mortalität, Morbidität und gesundheitsbezogene Lebensqualität sind im SGB V verankert und werden durch die DGOU vertreten. Dennoch sind bei Verletzungen und Erkrankungen des Bewegungsapparates Surrogate wie radiologische Parameter (z.B. knöcherne Konsolidierung auf Röntgen- oder CT-Bildern) von erheblicher Relevanz. Die konditionale Wahrscheinlichkeit, dass z.B. im Falle einer inadäquaten Reposition von Gelenk-bildenden Fragmenten nach intra-artikulären Frakturen oder einer verzögerten bzw. ausbleibenden Knochenbruchheilung gute funktionelle Ergebnisse oder eine Wiederherstellung der gesundheitsbezogenen Lebensqualität auf das Ausgangsniveau erreicht werden können, ist gering.

Die Art der Gewichtung verschiedener patientenrelevanter Endpunkte (S. 49) ist unklar. Diese kann quantitativ, semi-quantitativ oder qualitativ sein, muss aber spezifiziert werden. Natürlich sind Nutzen und Schaden themenspezifisch zu definieren. Dennoch sollten die Gewichtungsprozeduren besser illustriert werden.

2.4. Diagnostische Verfahren

Die ab S. 64 ff unter 3.5 genannten Aspekte zu diagnostischen Verfahren sind inkonsistent und müssen überarbeitet werden.

Es sollte eine klare Trennung zwischen Efficacy (Accuracy), Effectiveness (Beeinflussung der ärztlichen Entscheidungsfindung) und Efficiency (günstige Beeinflussung des Krankheitsverlaufes) gemäß der Fryback-Thornbury-Hierarchie¹ erfolgen.

Die Aussage, dass „... die alleinige Gewinnung diagnostischer Informationen (ohne therapeutische oder präventive Konsequenzen) keinen sozialrechtlich relevanten Nutzen“ besitzt, ist schlichtweg falsch.

Ohne Kenntnisse über die diagnostische Genauigkeit eines Tests (d.h., ob der Test zwischen erkrankten und nicht-erkrankten Populationen unterscheiden kann) können keine Effectiveness- oder Efficiency-Studien geplant werden.

Für die meisten diagnostischen Tests liegen keine robusten Genauigkeitsinformationen vor. So sind z.B. Analysen des Traumaregisters, welche einen Überlebensvorteil von schwerverletzten Patienten durch die primäre Ganzkörper-CT suggerieren^{2, 3}, nur dann sinnvoll beurteilbar, wenn Informationen über die diagnostischen Genauigkeit der Methode⁴ und deren Konsequenz für ärztliche Entscheidungen berücksichtigt werden.

Die Gleichsetzung prognostischer und diagnostischer Indikatoren auf S. 64 des Papiers steht im eklatanten Widerspruch zu den Prinzipien des stellvertretenden Institutsdirektors, Herrn PD Dr. med. Stefan Lange, der zu Recht eine klare Abgrenzung zwischen diesen Informationen fordert. Sie steht allerdings auch im Widerspruch zu späteren differenzierten Äußerungen und muss dringend angepasst werden.

Im Folgenden sollen lediglich diagnostische, nicht prognostische Studien adressiert werden. *Randomisierte Studien stehen auf der finalen Stufe der diagnostischen Evaluationskaskade.* Sie sind *nur dann zulässig*, wenn sich der experimentelle Test im Vergleich zu einem akzeptierten oder offensichtlichen Referenzstandard als ausreichend genau erwiesen hat. Die Grenze, ab wann klassische Messgrößen wie Sensitivität, Spezifität, PPV, NPV, DOR, LR oder ROC des experimentellen Tests dem bisherigen Referenztest nicht-unterlegen oder äquivalent anzusehen sind, müsste durch das IQWiG festgelegt und benannt werden.

Die Aussage, dass Studien mit einer „idealerweise zufälligen Zuteilung von Patientinnen und Patienten zu einer Strategie mit bzw. ohne Anwendung der zu prüfenden diagnostischen Maßnahme oder zu Gruppen mit bzw. ohne Offenlegung der (diagnostischen) Testergebnisse“ der höchste „Evidenzgrad“ zugesprochen wird, wird durch die DGOU aus folgenden Gründen abgelehnt:

- i. Wie oben ausgeführt, erfordert eine randomisierte diagnostische Studie primär den Nachweis der diagnostischen Genauigkeit des interessierenden Tests. Dies ist häufig nicht gegeben. Ein inakkurater Test kann und darf aber keine therapeutischen Konsequenzen nach sich ziehen. Zeigt also ein

- hypothetisches RCT den Vorteil eines individuellen Tests oder einer diagnostischen Strategie im Vergleich zu einer Kontrolle, ohne dass verlässliche Daten zu deren Genauigkeit vorliegen oder evtl. im Rahmen des RCT erhoben wurden, sind die Ergebnisse nicht zu interpretieren.
- ii. Die Protokoll-konforme Nicht-Offenlegung von Testergebnissen kann insb. im Fall von invasiven oder Strahlen-exponierenden Methoden zu ethisch grenzwertigen Situationen führen. Ein konkretes und aktuelles Beispiel ist das EF3X Trial.⁵ Dieses soll den Nutzen einer intra-operativen 3D-Bildgebung für die chirurgische Therapie intra-artikulärer Frakturen nachweisen. In der Kontrollgruppe werden Patienten einem 3D-Röntgen-Scan unterzogen - die Bilder werden den Operateuren jedoch nicht gezeigt. Obwohl die Studie formal einem „hohen Evidenzgrad“ genügt, sind die ethischen Probleme offensichtlich. Die Kontrollgruppe unterläuft eine Testprozedur mit methodisch unvermeidbarer additiver Strahlenexposition. Ihr wird dabei jeder mögliche Gewinn, der aus den Testergebnissen resultieren könnte (insb. der unmittelbare intra-operative Nachweis einer unzureichenden Fragmentreposition oder suboptimalen Implantatpositionierung), bewusst vorenthalten. Das IQWiG kann und darf derartigen Studien nicht Vorschub leisten.
 - iii. Sollten die o.g. Annahmen die Verfahrensordnung des G-BA repräsentieren, kann die DGOU nur dringend eine kritische Reflexion des bisherigen Verfahrens und eine Neubewertung anregen.

Die Aussage, dass „derartige Studien ... mit vergleichsweise moderatem Aufwand durchführbar sind“ (S. 65), ist jenseits der Realität.

Tatsächlich gehört die schrittweise und hierarchische Evaluation des Nutzens diagnostischer Tests (insb. im Falle fehlender definitiver Referenzstandards) zu den kompliziertesten und aufwändigsten Unterfangen der klinischen und der Versorgungsforschung.

Auf S. 66 werden erneut prognostische und diagnostische Indikatoren entgegen der IQWiG-Direktive vermengt bzw. unklar voneinander abgegrenzt.

Die Aussage „Insgesamt entscheidend ist weniger, inwieweit eine diagnostische oder prognostische Information einen aktuellen oder zukünftigen Gesundheitszustand feststellen kann, sondern dass diese Information auch prädiktive Bedeutung hat, also den höheren (oder geringeren) Nutzen einer Folgebehandlung vorhersagen kann“ ist inkorrekt. Dies wurde bereits detailliert im obigen Absatz begründet.

Einfach gesagt ist die diagnostische Information über den aktuellen Gesundheitszustand die unabdingbare Voraussetzung für alle weiteren medizinischen Maßnahmen.

Logisch fehlerhaft ist die Aussage, dass „ein Nutzen insgesamt sich in der Regel nur dann ergeben kann, wenn beide Interventionen ihr Ziel erfüllen: Sowohl bei unzureichender prädiktiver Diskriminationsfähigkeit der diagnostischen Intervention als auch bei fehlender Effektivität der therapeutischen Intervention wird eine Studie keinen Nutzen der diagnostischen Intervention zeigen können.“

Der Nutzen eines Tests kann sich nur dann ergeben, wenn dessen Diskriminationsfähigkeit nachgewiesen wurde. Es macht keinerlei Sinn, eine vergleichende Studie mit einem ungenauen Test durchzuführen. Auch der Nachweis

eines besseren Outcomes durch Anwendung eines experimentellen Tests ist bedeutungslos, sofern nicht vorher dessen Genauigkeit evaluiert wurde.

Unzutreffend ist auch die Aussage, dass „wenn der neue Test direkte patientenrelevante Vorteile aufweist, wie zum Beispiel geringere Invasivität oder Strahlungsfreiheit, es nicht immer notwendig sein (wird), erneut die gesamte diagnostisch-therapeutische Kette zu untersuchen, weil die therapeutischen Konsequenzen, die sich aus dem neuen Test ergeben, sich nicht von denen des bisherigen Tests unterscheiden.“

Gerade dann, wenn der neue Test mit einer geringeren Rate unerwünschter Ereignisse assoziiert ist, muss er beweisen, dass er genauso akkurat ist wie sein direkter Vergleichspartner.

Die auf S. 66 als „Anreicherungsdesign“ beschriebene Methode repräsentiert die Philosophie der individualisierten, personalisierten oder stratifizierten Medizin. Ihr sollte entweder ein eigenes Kapitel oder sogar ein separates Methodenpapier gewidmet werden. In der aktuellen Form und Fassung sind die Ausführungen schnell angreifbar. Beispielsweise wird das Problem eines *partial verification bias*, welches dann entsteht, wenn lediglich positive (nicht jedoch negative) Index-Test-Ergebnisse mittels eines akzeptierten Referenztests bestätigt werden, nur unzureichend und auch nicht im Hinblick auf therapeutische Konsequenzen adressiert.

Viele Referenztests sind entweder invasiv (z.B. Laparoskopie, Gewinnung biologischer Proben) oder exponieren Patient(inn)en gegenüber Röntgenstrahlung (z.B. CT). Die geforderten „Studien mit zufälliger Zuordnung der Reihenfolge der (voneinander unabhängigen und möglichst verblindeten) Testdurchführung bei denselben Patientinnen und Patienten“ wäre in diesem Falle aus ethischen Gründen unmöglich. Nach den Empfehlungen des IQWiG würden Patient(inn)en z.B. sowohl eine Röntgen- UND CT-Diagnostik, eine Normal- UND Niedrig-Dosis-CT oder eine CT UND eine invasive (operative) Diagnostik durchlaufen. Werden Patient(inn)en hingegen per Zufall der einen oder anderen Teststrategie (im Vergleich zu einem Referenzstandard?) zugeordnet, ist ein direkter Vergleich unmöglich.

Das 1. Brückenprinzip nach Albert lautet „Sollen impliziert Können.“⁶ Das IQWiG fordert hier theoretisch ideale Studiendesigns, welche praktisch schlichtweg nicht umsetzbar sind.

Auf S. 69 wird erstmals eine Trennung zwischen prognostischen und prädiktiven Wertigkeiten vorgenommen.

Aufgrund der o.g. nicht überzeugenden Darlegungen und der speziellen Probleme und Methoden empfiehlt die DGOU, für diagnostische Verfahren ein separates Dossier zu verfassen.

2.5. Potenzialbewertung

Der wesentliche Aspekt der Potenzialbewertung (3.8.) erscheint etwas abrupt nach den Themenfeldern Früherkennung und Screening (3.6.) und Prävention (3.7.). Insgesamt lässt diese Gliederung Tendenzen erkennen, dass der aktuelle Methodenreport unter Zeitdruck und ohne die erforderliche Tiefe erstellt wurde.

Leider offenbaren sich ab S. 73 alle Konflikte, welche sich aus den bisherigen Methoden einer externen Leistungsbewertung von medizinischen Interventionen mit Allgemeingültigkeitsanspruch ergeben.

Auf S. 74 heißt es:

Aufgrund der besonderen Zielsetzung sind bei der Potenzialbewertung im Vergleich zur Nutzenbewertung deutlich niedrigere Anforderungen an die Evidenz zu stellen. Letztendlich besteht erst das Ziel der Erprobung darin, eine adäquate Datengrundlage für eine zukünftige Nutzenbewertung herzustellen. Dementsprechend lässt sich ein Potenzial insbesondere auch auf der Basis nicht randomisierter Studien begründen.

Es handelt sich hierbei um eine fundamentale Aussage. Es ist völlig unerheblich, zu welchem Zweck eine klinische Studie durchgeführt wird – keine einzige klinische Studie kann aus biostatistischen Gründen beweisführend sein.⁷ Jede klinische Studie trägt zur Annahme über die Wirksamkeit, Nicht-Wirksamkeit oder den Schaden einer bestimmten Intervention im Vergleich zu einer Kontrolle bei. Wenn nunmehr gemäß IQWiG-Empfehlungen ein *Potenzial* aufgrund einer nicht-randomisierten Studie begründet werden kann, gilt dies aus einfachen logischen Gründen auch für den möglichen *Nutzen* einer Intervention. *Potenzial und Wirksamkeit sind in der Erkenntnistheorie nicht zu trennen.*

Die Definitionen von „geringer“, „sehr geringer“ und „minimaler“ Ergebnissicherheit nicht randomisierter Studien sind aus methodischer Sicht inakzeptabel - auch die Begründungen für diese Kategorisierung sind nicht überzeugend. Der letzte Absatz auf S. 75 springt ohne Überleitung von einer therapeutischen zu einer diagnostischen Fragestellung.

Aus methodischer Sicht hinterlassen die o.g. Punkte allenfalls Ratlosigkeit.

2.6. Kosten-Nutzen-Bewertung

Das Kosten-Nutzen-Modell des IQWiG wurde bereits in der Vergangenheit kritisiert. Insb. der Terminus „Effizienzgrenze“ konnte den oberen Gremien der DGOU bisher nicht verständlich erläutert werden.

Leider beruht auch der aktuelle Bericht vornehmlich auf qualitativen, bestenfalls semi-quantitativen (S. 76 – 112) Aussagen, welche auch einem Lehrbuch der Gesundheitsökonomie entnommen werden können.

Eine einfache, beantwortbare Frage wäre

„Wie viel darf eine bestimmte (operative / nicht-operative / rehabilitative / sonstige) Maßnahme in Orthopädie und Unfallchirurgie pro QALY / DALY kosten, um in der Bundesrepublik als effektiv zu gelten?“

Diese Frage wird aus Sicht der DGOU im IQWiG-Bericht nicht ausreichend adressiert.

2.7. „Rangordnung verschiedener Studienarten / Evidenzgrade“

Das IQWiG ist unverändert den „Levels of Evidence“ des Centre for Evidence-Based Medicine in Oxford / UK verhaftet (<http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/>). Spätestens seit dem berühmten Zitat von Bruce Charlton “The idea of a hierarchy of methods is amazing nonsense ... the validity of a piece of science is not determined by its method”⁸ hätte jedoch ein Umdenkprozess einsetzen müssen, der sich in der klinisch-methodischen Wissenschaftsgemeinschaft langsam durchsetzt. Eine immer noch stattfindende Hierarchisierung von Studiendesigns lehnt die DGOU ab und unterstreicht die Bedeutung der *methodischen Qualität* einzelner Untersuchungen zur Beantwortung einer *a priori* definierten wissenschaftlichen Frage. Einfach gesagt kann die klinische Bedeutsamkeit einer exzellent geplanten, durchgeführten und berichteten Fallserie, Kohorten- oder Fall-Kontroll-Studie diejenige einer schlecht geplanten, durchgeführten und berichteten randomisierten kontrollierten Studie überbieten. Es ist daher müßig, sich über Studiendesigns und deren „Hierarchie“ in der klinisch-wissenschaftlichen Erkenntniskaskade“ zu verständigen, solange nicht Einigkeit über die methodische Qualität besteht.

Auf S. 174 wird betont, dass „sich diese [qualitative] Graduierung auf das Verzerrungspotenzial (siehe Abschnitt 8.1.4) und nicht auf den Evidenzgrad einer Studie bezieht.“ Somit besteht Einverständnis zwischen IQWiG und DGOU. Sollten jedoch zukünftig weiterhin Begriffe wie „Evidenzklassifizierungen“ oder „Evidenzhierarchien“ in Empfehlungen und / oder Vorgaben des IQWiG bzw. des GBA auftauchen, wird die DGOU auf die Notwendigkeit einer einheitlichen und konsentierten Terminologie hinweisen.

3. Empfehlungen der DGOU

In der Zusammenschau zieht die DGOU folgende Rückschlüsse und Konsequenzen aus der aktuellen Version des Methodenreports des IQWiG:

1. Insg. wirkt der aktualisierte Report nicht „aus einem Guss“. Es gibt deutliche Diskrepanzen in den verschiedenen Aussagen, Interpretationen und Forderungen.
2. Letztere resultieren möglicherweise aus der sensiblen Schnittstelle des IQWiG zwischen der übergeordneten Instanz des G-BA, den Kostenträgern, Fachgesellschaften, Leistungserbringern und Konsumenten.
3. Ein „one-fits-all“ Methodenbericht ist nicht mehr zeitgemäß. Es müssen separate Berichte für diagnostische, therapeutische, gesundheitsökonomische und sonstige Evaluationen erstellt werden – ohne das Zielpublikum durch schiereres Volumen abzuschrecken. Methodisch belastbare, inhaltlich verständliche, politisch umsetzbare und dabei rechtskräftige Dossiers können und sollten auf weniger als 30 Seiten abgehandelt werden.
4. Viele der genannten methodischen Standards sind entweder veraltet, vage, offensichtlich kompromissbemerkt oder klinisch nicht umsetzbar. Die DGOU könnte besser mit klaren quantitativen Vorgaben oder Empfehlungen leben als mit Schaubildern ohne exakte Benennung der Einheiten auf x- und y-Achse.
5. Die DGOU bietet dem IQWiG und dem G-BA die klinische und methodische Expertise an, um für das gesundheitspolitisch und –ökonomisch relevante Aktionsfeld „Muskuloskeletale Erkrankungen und Verletzungen“ ein Methodenpapier mit unmittelbarer Versorgungsrelevanz zu erstellen.

4. Literatur

- (1) Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11(2):88-94.
- (2) Huber-Wagner S, Lefering R, Qvick LM et al. Effect of whole-body CT during trauma resuscitation on survival: a retrospective, multicentre study. *Lancet* 2009;373(9673):1455-1461.
- (3) Huber-Wagner S, Biberthaler P, Haberle S et al. Whole-body CT in haemodynamically unstable severely injured patients--a retrospective, multicentre study. *PLoS One* 2013;8(7):e68880.
- (4) Stengel D, Ottersbach C, Matthes G et al. Accuracy of single-pass whole-body computed tomography for detection of injuries in patients with major blunt trauma. *CMAJ* 2012;184(8):869-876.
- (5) Beerekamp MS, Ubbink DT, Maas M et al. Fracture surgery of the extremities with the intra-operative use of 3D-RX: a randomized multicenter trial (EF3X-trial). *BMC Musculoskelet Disord* 2011;12:151.
- (6) Albert H. *Traktat über kritische Vernunft*. 5 ed. Stuttgart: UTB; 1991.
- (7) Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2(8):e124.
- (8) Charlton BG. Fundamental deficiencies in the megatrial methodology. *Curr Control Trials Cardiovasc Med* 2001;2(1):2-7.