

# 1. Lebensqualitätsmessung in der frühen Nutzenbewertung

## 1.1 Lebensqualitätsmessung zwischen Theorie und Praxis

Ein Gastbeitrag von: Dr. Katharina Schmidt und Dr. Claudia Pütz, Ecker + Ecker GmbH

### 1.1.1 Bedeutung der Lebensqualitätsmessung im Arzneimittelsektor

Die Lebensqualität wird von der Weltgesundheitsorganisation (engl.: World Health Organization, WHO) definiert als „die Wahrnehmung von Individuen bzgl. ihrer Position im Leben im Kontext der Kultur und der Wertsysteme, in denen sie leben und in Bezug auf ihre Ziele, Erwartungen, Standards und Interessen.“<sup>1</sup> (*eigene Übersetzung der Autoren*). Eine Teilmenge der Lebensqualität ist die gesundheitsbezogene Lebensqualität, wobei zwischen diesen zwei Begrifflichkeiten häufig nicht unterschieden wird. Die gesundheitsbezogene Lebensqualität bezieht sich jedoch nur auf die Aspekte, die die Position im Leben, Ziele, Erwartungen, Standards und Interessen in Bezug auf die Gesundheit betreffen. Dennoch beeinflusst die gesundheitsbezogene Lebensqualität natürlich auch umgekehrt alle weiteren Aspekte des Lebens und der Lebensqualität. Im Folgenden beziehen wir uns hier auf die gesundheitsbezogene Lebensqualität, die im Weiteren zur einfacheren Lesbarkeit mit „LQ“ abgekürzt wird.

Die LQ ist ein mehrdimensionales und hochkomplexes Konstrukt. Daher gibt es auch keine abschließende Aufzählung, welche Aspekte die LQ tatsächlich beeinflussen, gerade weil die Einflussfaktoren von Person zu Person unterschiedlich sein können. Es ist an dieser Stelle sehr wichtig zu erwähnen, dass die LQ somit immer subjektiv beurteilt wird – auch wenn es Ansätze gibt, die LQ durch sogenannte Fremdbewerter (engl.: Proxy), wie z.B. nahe Angehörige, beurteilen zu lassen.

Dadurch, dass die LQ in engem Zusammenhang mit Gesundheit und Krankheit steht, ist sie ein wichtiger Endpunkt in klinischen Studien. Insbesondere fordern daher die Zulassungsbehörden (u.a. EMA, FDA, BfArM) im Rahmen der Arzneimittelzulassung regelhaft diese Daten, um nicht nur die Wirksamkeit und die Sicherheit zu beurteilen, sondern auch um Erkenntnisse zu erlangen, wie die Medikation die LQ der Patienten beeinflusst. Auch das deutsche

---

<sup>1</sup>World Health Organization Division of Mental Health and Prevention of Substance Abuse (Hrsg.) (1997)

Sozialgesetzbuch V (SGB V) sieht die LQ als einen patientenrelevanten Endpunkt an (§ 35 Abs. 1b Satz 5 SGB V).

Der Gemeinsame Bundesausschuss (G-BA) beurteilt den therapeutischen Zusatznutzen von Arzneimitteln mit neuen Wirkstoffen in Deutschland und bezieht hierbei ebenfalls maßgeblich neben Mortalität und Morbidität die LQ mit ein (§ 6 Abs. 2, 4. Kapitel VerFO). Über die Gewichtung der drei Endpunktkategorien untereinander entscheidet der G-BA von Fall zu Fall. Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), das vom G-BA mit der Nutzenbewertung von Arzneimitteln beauftragt werden kann, schreibt dazu in seinem aktuellen Methodenpapier etwas konkreter: „Bei besonders schwerwiegenden oder gar lebensbedrohlichen Erkrankungen ist es beispielsweise in der Regel nicht allein ausreichend, eine Verbesserung der LQ durch die Anwendung einer zu bewertenden Intervention nachzuweisen, wenn nicht gleichzeitig mit ausreichender Sicherheit ausgeschlossen werden kann, dass die schwerwiegende Morbidität oder gar Mortalität in einem nicht mehr akzeptablen Ausmaß ungünstig beeinflusst wird.“<sup>2</sup>

LQ-Erhebungen sind somit sowohl für die Zulassung und Nutzenbewertung als auch für den Therapiefortschritt und den Versorgungsalltag von hoher Relevanz. Chronische Erkrankungen nehmen anteilig im Vergleich zu Infektionskrankheiten weiter zu,<sup>3</sup> sodass klassische Mortalitätsendpunkte oder eine Heilung als Endpunkt erst deutlich später oder gar nicht erreicht werden. Somit wird bei vielen Erkrankungen heutzutage eine verbesserte LQ des Patienten als relevanter Therapiefortschritt gewertet.

Ein weiterer Aspekt, der die Relevanz von LQ-Daten herausstellt, ist das gesellschaftliche Bestreben nach selbstbestimmten Patienten (engl.: empowerment), die über ihre Therapiealternativen informiert sind und selbst mitentscheiden können und sollen. Patienten können jedoch leichter über Therapiealternativen entscheiden, die merkliche Auswirkungen auf ihren Alltag mit der Erkrankung und eben ihre LQ haben, als über Laborparameter oder durch Bildgebung gewonnene Befunde, die sie nicht unmittelbar spüren.

Es wird demnach deutlich, dass LQ-Daten relevant für den Therapiefortschritt und den Versorgungsalltag sind. Daher befassen wir uns im Folgenden mit der Frage, wie die LQ-Daten im Rahmen der frühen Nutzenbewertung in Deutschland erhoben und aufbereitet sein sollten.

Als erstes werden wir einige Statistiken hinsichtlich LQ bezogen auf Entscheidungen des G-BA in der frühen Nutzenbewertung vorstellen und darauffolgend Ablehnungsgründe für einen Zusatznutzen sowie die Anforderungen des G-BA

---

<sup>2</sup> Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2017a)

<sup>3</sup> Robert Koch-Institut (Hrsg.) (2016)

zusammenfassen. Im Anschluss daran führen wir einen Vergleich mit den Anforderungen aus der internationalen LQ-Forschung durch. Abschließend ziehen wir ein Fazit und skizzieren die zukünftigen Herausforderungen.

### 1.2.1 Zahlen zur Lebensqualitätsmessung in der frühen Nutzenbewertung

Insgesamt liegen 345 abgeschlossene Verfahren mit Beschluss des G-BA vor (Stand: 15.01.2019), die in die Datenanalyse einbezogen werden. Von diesen hat der jeweilige pharmazeutische Unternehmer in insgesamt 262 Verfahren in seinen Dossiers angegeben, dass LQ-Daten erhoben wurden. Darunter werden letztendlich bei 46 Verfahren (18% der Verfahren mit LQ-Erhebungen) die LQ-Ergebnisse für die Bewertung des Zusatznutzens durch den G-BA herangezogen. Hierzu zählen auch nachteilige Effekte aus dem Bereich der LQ, die der G-BA berücksichtigt hat. Eine grafische Darstellung dieser Angaben findet sich in Abbildung 1.

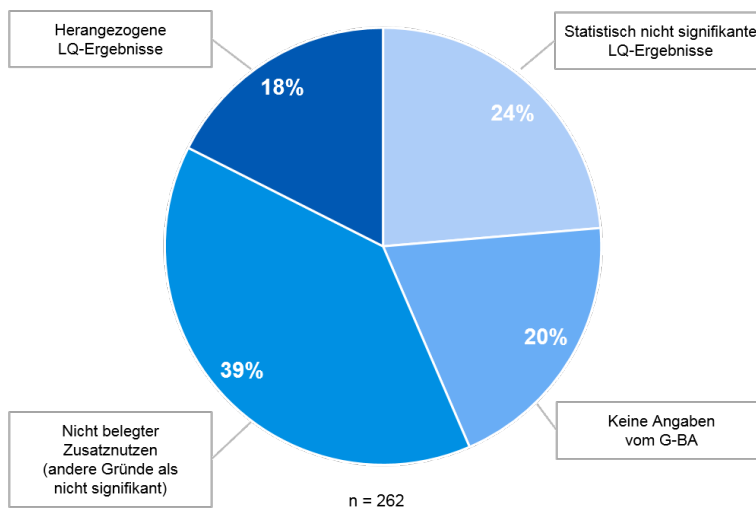


Abbildung 1: Übersicht der Verfahren mit LQ-Erhebungen in der FNB

Im Zeitverlauf von 2011 bis 2018 ist eine Zunahme an Verfahren mit LQ-Erhebungen zu erkennen (siehe Abbildung 2). Während in 2016 und 2018 die Anzahl der Verfahren mit LQ-Erhebungen besonders hoch waren (61 bzw. 58 Verfahren), gab es in 2017 einen kleinen Einbruch mit nur 35 Verfahren.

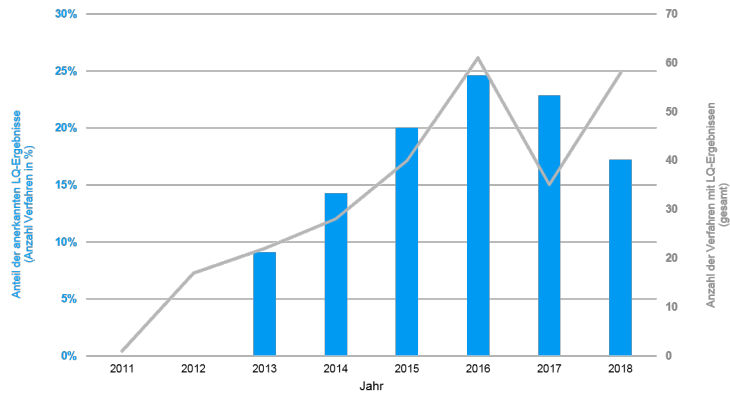


Abbildung 2: Übersicht der Verfahren mit LQ-Erhebungen im Zeitverlauf

Betrachtet man jedoch, wie erfolgreich die LQ-Ergebnisse im Rahmen der Anerkennung eines Zusatznutzens durch den G-BA waren, zeigen sich lediglich Erfolgsquoten von um die 20% in den vergangenen 4 Jahren.

Bei einer Betrachtung der Verfahren nach Indikationsgebiet zeigt sich, dass mit 109 von 262 Verfahren (42%) besonders viele LQ-Erhebungen in den Dossiers aus den onkologischen Bereichen berichtet werden. Die größte Erfolgswahrscheinlichkeit von Verfahren, die die LQ erhoben haben, fanden sich jedoch bei den Hauterkrankungen (LQ anerkannt in 6 von 8 Verfahren).

Dennoch sagen die bisher dargestellten Statistiken noch nicht viel über die Gründe von Anerkennung oder Ablehnung der LQ-Ergebnisse durch den G-BA aus. Dies soll in den folgenden Abschnitten näher beleuchtet werden.

### 1.3.1 Gründe für eine Ablehnung der Lebensqualitäts-ergebnisse

Abgesehen von nicht signifikanten Ergebnissen aus LQ-Erhebungen gibt es einige weitere Aspekte, die häufig in den Nutzenbewertungen, Beschlussdokumenten oder Tragenden Gründen des G-BA zu finden sind und zu einem nicht belegten Zusatznutzen führen.

Diese Kritikpunkte des G-BA und des IQWiG lassen sich in drei verschiedene Phasen des Prozesses der Studienplanung beim pharmazeutischen Unternehmer einordnen:

- a. Planungsphase
  - b. Auswertungsphase
  - c. Berichtsphase
- a. In der Planungsphase wird ein wichtiger Grundstein für eine erfolgreiche Bewertung gelegt. Hier ist es wichtig, dass ein validierter LQ-Fragebogen ausgewählt und angewendet wird. Nach dem IQWiG und auch nach dem G-BA sollte möglichst sowohl ein generischer als auch ein krankheitsspezifischer Fragebogen genutzt werden.<sup>4</sup> Unter den verwendeten Fragebögen, die nicht für die Beurteilung eines Zusatznutzens herangezogen wurden, waren beispielsweise Fragebögen, die lediglich bestimmte Symptome oder Gesundheitszustände erhoben haben und keine LQ und stattdessen als Morbiditätspunkte anerkannt wurden (z.B. EQ-5D VAS, melanomspezifische Subskala des FACT-M, LCSS).<sup>5</sup> Andere Fragebögen waren nicht für die jeweilige (Sub-)Population der Erkrankung validiert und wurden daher vom G-BA nicht anerkannt (EORTC QLQ-C30 für Myelofibrose-Patienten<sup>6</sup> oder Patienten mit CLL<sup>7</sup>; IWQoL-Lite bei Diabetes mellitus Typ 2)<sup>8</sup>. Ebenfalls erkennt der G-BA auch keine Nutzwerte (z.B. vom EQ-5D) an, da diese nicht auf Basis der betroffenen Studienpopulation ermittelt wurden, sondern in Bezug auf eine Referenzbevölkerung.<sup>9</sup>

Des Weiteren spielt in der Planungsphase das Studiendesign eine relevante Rolle. Da dieser Punkt jedoch nicht spezifisch für LQ-Erhebungen gilt, erfolgt an dieser Stelle keine weitere Ausführung dieses Aspektes.

- b. Bei der Auswertung von LQ-Daten lehnt der G-BA Studien ab, die eine Fragebogen-Rücklaufquote von weniger als 70% aufweisen oder bei denen sich zwischen Interventions- und Kontrollgruppe die Rücklaufquote um mehr als 15%-Punkte unterscheidet.<sup>10</sup>

Des Weiteren werden LQ-Ergebnisse vom G-BA nicht herangezogen, wenn die gezeigten Unterschiede nicht klinisch relevant sind.<sup>11</sup> Hierfür gibt es für die meisten gängigen LQ-Fragebögen in Studien ermittelte Irrelevanzschwellen (engl.:

---

<sup>4</sup> Gemeinsamer Bundesausschuss (2012b); Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2017a)

<sup>5</sup> beispielsweise: Gemeinsamer Bundesausschuss (2016c, 2018a); Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2017b)

<sup>6</sup> Gemeinsamer Bundesausschuss (2012a)

<sup>7</sup> Gemeinsamer Bundesausschuss (2014a)

<sup>8</sup> Gemeinsamer Bundesausschuss (2018d)

<sup>9</sup> beispielsweise Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2016a)

<sup>10</sup> Gemeinsamer Bundesausschuss (2014c, 2016b)

<sup>11</sup> beispielsweise Gemeinsamer Bundesausschuss (2016a, 2016d)

minimal important difference, MID), die aufzeigen, wie viele Punkte Unterschied auf einer Skala für den Patienten einen merklichen Unterschied für die LQ darstellen.<sup>12</sup>

- c. In der letzten Phase des Berichtes der Ergebnisse im Nutzendossier können wiederum Fehler unterlaufen, die zu einer Nicht-Anerkennung eines Zusatznutzens im Endpunkt LQ führen können. Wenn beispielsweise statistische Tests so angewendet werden, dass die Interpretation der Ergebnisse nicht eindeutig ist, kann der G-BA diese aufgrund des Verzerrungspotentials nicht heranziehen.<sup>13</sup> Der G-BA wünscht außerdem die vollständige Darstellung der Ergebnisse und vermutet hohes Verzerrungspotential, wenn nicht alle Daten vollständig ausgeführt werden. Beispielsweise, wenn mehrere LQ-Fragebögen eingesetzt wurden, aber nur die Ergebnisse von einem berichtet werden.<sup>14</sup> Auch wenn die Ergebnisse selektiv dargestellt werden – z.B. nur einzelne Subgruppen – vermutet der G-BA Unsicherheiten und zieht die Erkenntnisse nicht in vollem Maße heran.

Aus diesen angesprochenen „Ablehnungskriterien“ lassen sich Anforderungen des G-BA ableiten, die im Folgenden dargestellt werden.

#### 1.4.1 Anforderungen des G-BA an Lebensqualitätsmessungen

Aus der Analyse der vorliegenden Beschlussunterlagen können vorrangig 4 Anforderungen des G-BA (A bis D) für LQ-Erhebungen identifiziert werden (siehe Abbildung 3).

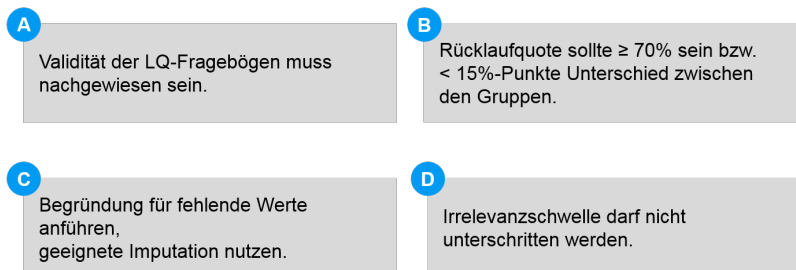


Abbildung 3: Anforderungen des G-BA an LQ-Erhebungen

- A. Die eingesetzten LQ-Fragebögen sollten in der jeweiligen Indikation, in der das zu bewertende Arzneimittel zugelassen

<sup>12</sup> King (2011)

<sup>13</sup> Gemeinsamer Bundesausschuss (2014b)

<sup>14</sup> Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2016b)

ist, validiert sein. Die Validierung von Fragebögen zieht nach sich, dass keine Veränderungen mehr an ihnen vorgenommen werden dürfen; es dürfen auch keine Fragen entfernt oder ergänzt werden. Es sollte also ausschließlich die validierte Version genutzt werden, keine Vorgängerversion, Folgeversion, gekürzte oder ergänzte Version.

- B. Als zweiter Anforderungspunkt sollen die Rücklaufquoten bei mindestens 70% liegen, da sonst Verzerrungen erwartet werden können. Diese könnten beispielsweise dadurch zustande kommen, dass die besonders kranken Patienten nicht geantwortet haben und somit deren sehr schlechten LQ-Ergebnisse nicht einbezogen werden. Ebenfalls könnten deutlich unterschiedliche Rücklaufquoten (>15%-Punkte) in Interventions- und Kontrollarm zu Verzerrungen der Ergebnisse führen. Der Grund hierfür ist, dass unterstellt werden könnte, dass nur die „gewünschten“ Fragebögen in die Auswertung eingeflossen sind und sich die Ergebnisse unvollständig darstellen.
- C. Ein weiterer Aspekt, den der G-BA als Anforderung stellt, ist, dass Begründungen ausfindig gemacht und angegeben werden sollten, wenn eine Häufung von fehlenden Werten zu erkennen ist. Es sollte hierbei auch eine geeignete Methode der Ersetzung von fehlenden Werten (Imputation) angewendet werden, um eine Verzerrung der Ergebnisse zu minimieren.
- D. Als letzte Anforderung ist vom G-BA häufig zu lesen, dass die Irrelevanzschwelle dargestellt und nicht unterschritten werden sollte. Wichtig ist, dass eine geeignete Irrelevanzschwelle für den jeweiligen Fragebogen genutzt wird, da diese je nach Skala unterschiedlich ausfallen kann. Es sollte daher die jeweilige MID für die Berechnungen der Effekte genutzt werden.

Zudem verlangt der G-BA in der letzten Zeit vermehrt, dass LQ-Erhebungen auch über den Krankheitsprogress und die Behandlungsphase der Arzneimittelstudie hinaus erhoben werden, da auch hier die weitere Entwicklung der LQ der Patienten von Bedeutung ist.

Nachdem die Anforderungen, die wir aus den vorliegenden G-BA-Beschlüssen ableiten konnten, nun zusammengefasst sind, widmen wir uns im Anschluss einem Abgleich der Erkenntnisse mit dem aktuellen wissenschaftlichen Standard.

### 1.5.1 Abgleich mit dem wissenschaftlichen Stand zur LQ-Erhebung

Für den Abgleich der Anforderungen des G-BA zu LQ-Erhebungen mit den Anforderungen aus der Forschung ziehen wir aktuelle Leitlinien bzw. Empfehlungen heran. Da speziell für Deutschland keine Leitlinien zur LQ-Erhebung existieren, werden internationale Empfehlungen der SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) Group<sup>15</sup> und der ISOQOL (International Society for Quality of Life Research) Group<sup>16</sup> herangezogen. Die SPIRIT-Gruppe setzt sich aus internationalen Experten von angesehenen Universitäten und Einrichtungen zusammen und veröffentlicht Empfehlungen, um die Vollständigkeit sowie Qualität von Studienprotokollen zu verbessern und hat mit der Erweiterung des SPIRIT-Statements einen zusätzlichen Schwerpunkt auf patientenberichtete Endpunkte gelegt (SPIRIT-PRO-Extension). Die ISOQOL-Gruppe ist multidisziplinär und ebenfalls international ausgestaltet (u.a. Ärzte, Patienten, Forscher, Professoren und Regierungsvertreter) und fördert qualitativ hochwertige Forschung in Bezug auf LQ und patientenberichtete Endpunkte (engl.: patient reported outcomes, PRO), u.a. mit jährlichen Konferenzen und Leitlinien. Im Folgenden werden die 4 Anforderungen des G-BA mit den internationalen Leitlinien verglichen.

- Zu A. Die Forderung des G-BA nach der Nutzung von validierten Fragebögen lässt sich in den Empfehlungen der SPIRIT-PRO Group und der ISOQOL Group nicht auf diese Weise wiederfinden. Sie sprechen sich dafür aus, dass begründet Instrumente gewählt werden sollten, die nach Möglichkeit evidenzbasiert sind, idealerweise in der interessierenden Population.<sup>17</sup> Allerdings wird empfohlen, die Fragebögen anhand spezifischer Reliabilitäts- und Validitätskriterien selbst zu beurteilen.<sup>18</sup> Diese Abweichung von den G-BA-Anforderungen hängt mit der Perspektive zusammen, die in Kapitel 1.6.1 noch diskutiert wird.
- Zu B. Für die Rücklaufquoten wird im Gegensatz zu den G-BA-Anforderungen in den Leitlinien kein konkreter Schwellenwert genannt. Es wird jedoch angemerkt, dass fehlende Rückläufer ein spezielles Problem von Erhebungen zu patientenberichteten Endpunkten sind.<sup>19</sup> Die SPIRIT-Gruppe begründet dies damit, dass Patienten, denen es besonders schlecht geht, nicht zu geplanten Assessments kommen und die LQ-Erhebungen dann nicht

---

<sup>15</sup> Calvert et al. (2018)

<sup>16</sup> Reeve et al. (2013)

<sup>17</sup> Calvert et al. (2018)

<sup>18</sup> Reeve et al. (2013)

<sup>19</sup> Calvert et al. (2018)



nachgetragen werden können, da sie zeitpunktabhängig individuell berichtet werden müssen. Die Autoren beschreiben weiter, dass einige fehlende Werte auch nicht vermieden werden können, weil es das Recht jedes Patienten ist, die Befragung nicht (vollständig) abzuschließen.<sup>20</sup> Um diese Fälle zu reduzieren, sollten die Fragebögen die Patienten möglichst wenig belasten.<sup>21</sup> Daher sollten die Fragebögen möglichst kurz sein und in einer Sprache, die von 11- bis 12-jährigen Kindern verstanden werden kann, verfasst sein.<sup>22</sup> Es sollen außerdem die vermeidbaren fehlenden Rückläufer minimiert werden, die aufgrund organisatorischer Probleme im Studienablauf entstehen.<sup>23</sup>

- Zu C. Die Anforderungen zu fehlenden Werten in Studien aus den wissenschaftlichen Empfehlungen sind gleichzusetzen mit denen des G-BA. Hier wird ebenfalls gefordert, dass fehlende Werte begründet und mithilfe geeigneter statistischer Methoden (Imputation, Sensitivitätsanalysen) analysiert werden.<sup>24</sup>
- Zu D. Die Wissenschaft sieht das Kriterium des Nachweises von relevanten Veränderungen auf den Skalen ebenfalls als wichtig an.<sup>25</sup> Soweit verfügbar ist es wünschenswert, hierfür die MID zu nutzen.<sup>26</sup> Es wird in spezifischen Veröffentlichungen darauf hingewiesen, dass es nicht die eine universelle MID gibt, sondern die MID vom klinischen Kontext, dem Erhebungsinstrument und dem Startpunkt der Patienten auf der Bewertungsskala abhängt.<sup>27</sup> Ein ergänzender Artikel der ISOQOL-Gruppe zu ihrer Leitlinie empfiehlt einen möglichst validen Wert durch zusätzliche Befragungen zu ermitteln, falls keine MID für den jeweiligen Fragebogen veröffentlicht ist.<sup>28</sup>

Zusammenfassend lässt sich zu den Anforderungen des G-BA und der Wissenschaft sagen, dass der G-BA die Anforderungen strenger auslegt als die Forschung zur LQ und PROs.

Es bleibt jedoch zu berücksichtigen, dass die Empfehlungen aus der Forschung zum Teil einen anderen Fokus haben als die des G-BA. Die Leitlinien sollen (klinische) Studiendurchführungen im Bereich der patientenberichteten Endpunkte unterstützen und sowohl Studienplaner als auch Investigatoren informieren. Auch sollen

---

<sup>20</sup> Calvert et al. (2018)

<sup>21</sup> Reeve et al. (2013)

<sup>22</sup> Reeve et al. (2013)

<sup>23</sup> Calvert et al. (2018)

<sup>24</sup> Calvert et al. (2018)

<sup>25</sup> Calvert et al. (2018); Chan, Edwards, Haywood, Mikles, and Newton (2018)

<sup>26</sup> Reeve et al. (2013)

<sup>27</sup> King (2011)

<sup>28</sup> Chan et al. (2018)

Fragebögen anhand der wissenschaftlichen Empfehlungen weiterentwickelt werden. Dahingegen zielt die frühe Nutzenbewertung auf evidenzbasierte Bewertungen im Vergleich zum medizinischen State of the Art im Indikationsgebiet ab, die die Basis für Erstattungsfragen durch die GKV bilden. Die Zuverlässigkeit der Instrumente spielt somit bei der Nutzenbewertung eine größere Rolle als bei LQ-Studien, die mitunter einen explorativen Charakter haben (sollen).

Es stellt sich nun jedoch die Frage, warum die Arzneimittelhersteller die Anforderungen des G-BA in der frühen Nutzenbewertung nicht vollumfänglich erfüllen (können). Der Grund ist ein Spannungsfeld, in dem sich die Pharmaunternehmen mit ihren LQ-Erhebungen befinden, das im folgenden Abschnitt vertiefend dargestellt wird. Außerdem werden auch Lösungsansätze präsentiert.

### 1.6.1 Lebensqualitätserhebungen im multidimensionalen Spannungsfeld

Die primäre Schwierigkeit bei der Erfüllung der G-BA-Anforderungen ergibt sich dadurch, dass die Bewertungsgrundlage für die frühe Nutzenbewertung die Daten der Zulassungsstudien sind. Noch bevor die frühe Nutzenbewertung in Deutschland startet, nehmen für die international tätigen pharmazeutischen Hersteller somit die Zulassungsregulierungen und -anforderungen der amerikanischen Arzneimittelbehörde (engl.: Food and Drug Administration, FDA) und der Europäischen Arzneimittelagentur (engl.: European Medicines Agency, EMA) bei der Studienplanung eine hochrelevante Rolle ein. Da die Studien der Hersteller sowohl die Anforderungen der Zulassungsbehörden als auch die des G-BA erfüllen müssen, entsteht hier ein potentielles Spannungsfeld.

Die EMA sieht LQ-Endpunkte – genau wie der G-BA – als wichtige Ergänzung zu Wirksamkeits- und Sicherheitsendpunkten an.<sup>29</sup> Sie sagt, dass die LQ-Ergebnisse vorrangig relevant werden, wenn bei der Wirksamkeit und Sicherheit keine Unterschiede zwischen dem neuen Produkt und dem bisherigen Therapiestandard zu erkennen sind.<sup>30</sup> Die FDA warnt davor, LQ-Indexwerte konfirmatorisch zu benutzen, wenn in einzelnen Komponenten der LQ-Erhebung große Unterschiede erwartet werden.<sup>31</sup> Hier wird demnach ein großes Verzerrungspotential gesehen, wenn einzelne Komponenten der LQ losgelöst betrachtet werden.

---

<sup>29</sup> Committee for Medicinal Products for Human Use (CHMP) (Hrsg.) (2005)

<sup>30</sup> Committee for Medicinal Products for Human Use (CHMP) (Hrsg.) (2005)

<sup>31</sup> U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), and Center for Devices and Radiological Health (CDRH) (2009)

Die EMA und die FDA verlangen ebenso wie der G-BA den Einsatz von validierten Messinstrumenten und die Vorab-Definition des Umgangs mit fehlenden Werten.<sup>32</sup>

Sowohl die EMA als auch die Leitlinie für die Industrie der FDA besagen, dass unterschiedliche Response-Kriterien in Abhängigkeit des Instrumentes und des Kontextes der klinischen Studie berücksichtigt werden sollten.<sup>33</sup>

Trotz der Einigkeit der Zulassungsbehörden mit dem G-BA in vielen der Anforderungen, gibt es abweichende Empfehlungen, welche LQ-Fragebögen vorrangig genutzt werden sollten. Ein Spannungsfeld aus Sicht der pharmazeutischen Unternehmer ist es daher, dass sie meistens international agieren und sich an den jeweiligen Empfehlungen der Länder zu LQ-Fragebögen orientieren müssen. Als klassisches Beispiel ist hier sicherlich der EQ-5D zu nennen, der in England vom NICE als Messinstrument für die LQ präferiert wird,<sup>34</sup> vom G-BA als LQ-Fragebogen jedoch nicht akzeptiert, sondern nur die EQ-5D VAS als Morbiditätspunkt herangezogen wird. Hier muss ein Kompromiss gefunden bzw. eine Abwägung getroffen werden, auch vor dem Hintergrund, dass die Patienten nicht mehr als nötig durch zusätzlichen Aufwand durch eine Vielzahl an Fragebögen belastet werden.

Erhebungen mit dem primären Ziel der LQ-Erfassung werden weitaus aufwendiger konzipiert und sind umfangreicher als in Zulassungsstudien üblich, da sie (wie oben erwähnt) andere Ziele verfolgen. LQ-Studien setzen einen hohen zeitlichen Aufwand auf allen Seiten (u.a. pharmazeutischer Unternehmer, Patient, Arzt) und große Motivationsfähigkeit der Studienärzte und -schwestern sowie der Befragten voraus. Es handelt sich bei den meisten (vor allem krankheitsspezifischen) Fragebögen um mehrere Seiten mit zum Teil sehr persönlichen Fragen. Die Beantwortung eines Fragebogens setzt voraus, dass sich der Befragte mit seiner tagesaktuellen LQ auseinandersetzt und sich die Zeit nimmt, alle Fragen zu beantworten – auch wenn sie manchmal ähnlich klingen. Durch diesen Aspekt ergibt sich noch eine weitere Schwierigkeit: Der Patient muss sich über den gesamten Umfang des Fragebogens sehr stark mit seiner Gesundheit konfrontieren und sich konzentrieren. Und das obwohl die Patienten unter starken Beeinträchtigungen, Nebenwirkungen und ebenso psychischen und sozialen Folgen ihrer Krankheit leiden.

Daher ist in der Wissenschaft durchaus bekannt, dass z.T. niedrigere Rücklaufquoten auftreten können und dies auch nicht verhindert werden kann. Selbstverständlich ist aber auch der

---

<sup>32</sup> Committee for Medicinal Products for Human Use (CHMP) (Hrsg.) (2005); U.S. Department of Health and Human Services et al. (2009)

<sup>33</sup> Committee for Medicinal Products for Human Use (CHMP) (Hrsg.) (2005); U.S. Department of Health and Human Services et al. (2009)

<sup>34</sup> National Institute for Health and Care Excellence (2013)

Wunsch nach Vollständigkeit der Daten (nicht nur vom G-BA) in Hinblick auf Verzerrungen nachvollziehbar. Es entsteht hier ein weiteres Spannungsfeld, das nicht aufzulösen ist. Es wäre ggf. eine praktikable Möglichkeit, dass vermehrt Begründungen für fehlende Werte beim Patienten erfragt und in das Nutzendossier eingebracht werden. Dann könnte der G-BA besser beurteilen, ob die fehlenden Werte die Ergebnisse systematisch verzerren oder kein erhöhtes Risiko für Verzerrungen bieten.

Aus Sicht der Arzneimittelhersteller ist prospektiv nur in den seltensten Fällen ersichtlich, wie erfolgsversprechend die jeweiligen LQ-Daten beim G-BA sein werden. Somit findet beim pharmazeutischen Unternehmer die Abwägung statt, wie viel Aufwand in die Planung, Auswertung und den Bericht von LQ-Daten eines jeden zu bewertenden Arzneimittels gelegt werden sollte. Unsere Analyse hat gezeigt, dass mit positiven LQ-Ergebnissen bisher in allen Fällen auch eine Verbesserung von weiteren Endpunkten einhergegangen ist, sodass kein Zusatznutzen allein aufgrund der LQ-Ergebnisse erreicht wurde.<sup>35</sup> Das ist auch nachvollziehbar, da LQ – wie eingangs beschrieben – ein multidimensionales Konstrukt und mit Morbiditätsparametern korreliert ist.

Aus unserer Analyse geht jedoch hervor, dass in bestimmten Indikationsgebieten LQ-Ergebnisse eine größere Rolle spielen als in anderen. Eingangs haben wir in Abschnitt 1.2.1 den hohen Anteil der LQ-Erhebungen im Bereich der Onkologie und die hohen Erfolgsaussichten zur Anerkennung der LQ bei den Hauterkrankungen erwähnt. Es ist plausibel, dass gerade bei diesen Erkrankungen die Nebenwirkungen für die Patienten große Einschränkung im Alltag und Sozialleben bedeuten (z.B. Haarausfall und Übelkeit/Erbrechen bei Chemotherapien oder starke Ausschläge bei Hauterkrankungen, die zu Schamgefühlen und sozialem Rückzug führen können). Somit gibt es zumindest Anhaltspunkte für den Hersteller, welche Relevanz LQ-Ergebnisse in dem jeweiligen Indikationsgebiet einnehmen, wenn bereits ein Arzneimittel in diesem Anwendungsgebiet bewertet wurde. Daher sollte in diesen Anwendungsgebieten darauf vermehrt Wert gelegt werden, bereits bei der Studienplanung neben den Zulassungsanforderungen auch die Anforderungen des G-BA zu berücksichtigen, wenn ebenfalls ein Launch in Deutschland geplant ist.

### 1.7.1 Fazit

Unsere Untersuchung zeigt, dass die Anforderungen des G-BA im Vergleich zur LQ-Forschung streng interpretiert werden.

---

<sup>35</sup> u.a. Gemeinsamer Bundesausschuss (2015, 2018b, 2018c)

Der G-BA fordert den Einsatz von ausschließlich validierten LQ-Instrumenten (A), Rücklaufquoten von mindestens 70% (B), Begründungen und Lösungsansätze für fehlende Werte (C) und erkennt LQ-Ergebnisse nur an, wenn sie oberhalb der Irrelevanzschwelle liegen (D).

Bisher haben die LQ-Ergebnisse eher einen unterstützenden Charakter in der frühen Nutzenbewertung. Das liegt einerseits daran, dass sich die Ergebnisse nur unter gewissen Unsicherheiten interpretieren lassen und andererseits daran, dass der G-BA Mortalitätsendpunkten scheinbar ein höheres Gewicht zuspricht. Ob das gerechtfertigt ist, lässt sich nicht pauschal sagen. Definitiv sollten LQ immer im Kontext der Erkrankung interpretiert werden und die Belastung durch die Einschränkungen vorzugsweise von Patienten oder ihren Ärzten beurteilt werden. Alternativ lässt sich auch diskutieren, ob die Gesellschaft über die Gewichtung der Endpunkte gegeneinander diskutieren sollte, da sie die Kosten der Therapien über die GKV trägt. Die Gewichtung der Endpunkte (Mortalität, Morbidität, LQ) sollte jedoch auf jeden Fall transparent und nachvollziehbar erfolgen.

Es gilt aber auch an die pharmazeutischen Unternehmer zu appellieren, vollständigere LQ-Daten zu erheben (wo sinnvoll und verhältnismäßig) und vermeidbare Fehler in LQ-Erhebungen zu unterlassen sowie die vorliegenden Daten bestmöglich einzubringen. Die LQ-Daten sind nämlich ergänzend zu Mortalitäts- und Morbiditätsdaten unverzichtbar – besonders im Kontext von chronischen Erkrankungen oder Erkrankungen, die den Alltag der Patienten stark beeinträchtigen.

Des Weiteren sollte bedacht werden, dass die heute vorliegenden LQ-Ergebnisse auf Basis der Bedingungen und Anforderungen der Vergangenheit geplant wurden, da die Zulassungsstudien langen Vorlauf und Rekrutierungszeit benötigen. Es bleibt daher abzuwarten, ob die hier zusammengefassten Anforderungen des G-BA in zukünftigen Studien besser umgesetzt werden und die Ergebnisse zur LQ in der frühen Nutzenbewertung entsprechend einen höheren Stellenwert erhalten.

**Literatur:**

- Calvert, M., Kyte, D., Mercieca-Bebber, R., Slade, A., Chan, A. W., King, M. T., . . . Groves, T. (2018). Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *Jama*, 319(5), 483-494. doi:10.1001/jama.2017.21903
- Chan, E. K. H., Edwards, T. C., Haywood, K., Mikles, S. P., & Newton, L. (2018). Implementing patient-reported outcome measures in clinical practice: a companion guide to the ISOQOL user's guide. *Qual Life Res.* doi:10.1007/s11136-018-2048-4
- Committee for Medicinal Products for Human Use (CHMP) (Hrsg.). (2005). Reflection Paper on the Regulatory Guidance for the Use of Health-Related Quality of Life (HRQoL) Measures in the Evaluation of Medicinal Products, CPMP/EWP/139391/04. Retrieved from <https://www.ema.europa.eu/en/regulatory-guidance-use-health-related-quality-life-hrql-measures-evaluation-medicinal-products>
- Gemeinsamer Bundesausschuss. (2012a). Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO. Wirkstoff: Ruxolitinib. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/38/#tab/nutzenbewertung>
- Gemeinsamer Bundesausschuss. (2012b). Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V. Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO. Wirkstoff: Ivacaftor. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/36/#tab/nutzenbewertung>
- Gemeinsamer Bundesausschuss. (2014a). Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO. Wirkstoff: Obinutuzumab. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/131/#tab/nutzenbewertung>
- Gemeinsamer Bundesausschuss. (2014b). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V -

- Radium-223-dichlorid. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/338/#tab/beschluesse>
- Gemeinsamer Bundesausschuss. (2014c). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Ruxolitinib. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/116/#tab/beschluesse>
- Gemeinsamer Bundesausschuss. (2015). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Ruxolitinib (neues Anwendungsgebiet). Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/167/#tab/beschluesse>
- Gemeinsamer Bundesausschuss. (2016a). Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V. Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO. Wirkstoff: Panobinostat. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/193/#tab/nutzenbewertung>
- Gemeinsamer Bundesausschuss. (2016b). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Nivolumab (neues Anwendungsgebiet). Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/186/#tab/beschluesse>
- Gemeinsamer Bundesausschuss. (2016c). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Trametinib. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/197/#tab/beschluesse>
- Gemeinsamer Bundesausschuss. (2016d). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Emtricitabin/Tenofovirafenamid. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/233/#tab/beschluesse>

- Gemeinsamer Bundesausschuss. (2018a). Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V. Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO. Wirkstoff: Velmanase alfa. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/374/#tab/nutzenbewertung>
- Gemeinsamer Bundesausschuss. (2018b). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Dupilumab. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/333/#tab/beschlusse>
- Gemeinsamer Bundesausschuss. (2018c). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Guselkumab. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/332/#tab/beschlusse>
- Gemeinsamer Bundesausschuss. (2018d). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Insulin glargin/Lixisenatid. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/353/#tab/beschlusse>
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). (2016a). IQWiG-Berichte – Nr. 379. Sacubitril/Valsartan – Nutzenbewertung gemäß § 35a SGB V. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/213/#tab/nutzenbewertung>
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). (2016b). Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V. Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO. Wirkstoff: Idebnon. Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/192/#tab/nutzenbewertung>
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). (2017a). Allgemeine Methoden Version 5.0 vom 10.07.2017. Retrieved from [https://www.iqwig.de/download/Allgemeine-Methoden\\_Version-5-0.pdf](https://www.iqwig.de/download/Allgemeine-Methoden_Version-5-0.pdf)
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). (2017b). IQWiG-Berichte – Nr. 471. Ceritinib (nicht



- kleinzelliges Lungenkarzinom) – Nutzenbewertung gemäß § 35a SGB V (Ablauf Befristung). Retrieved from <https://www.g-ba.de/informationen/nutzenbewertung/264/#tab/nutzenbewertung>
- King, M. T. (2011). A point of minimal important difference (MID): a critique of terminology and methods AU - King, Madeleine T. *Expert Review of Pharmacoeconomics & Outcomes Research*, 11(2), 171-184. doi:10.1586/erp.11.9
- National Institute for Health and Care Excellence. (2013). Guide to the methods of technology appraisal 2013. Retrieved from <https://www.nice.org.uk/process/pmg9/>
- Reeve, B. B., Wyrwich, K. W., Wu, A. W., Velikova, G., Terwee, C. B., Snyder, C. F., . . . Butt, Z. (2013). ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*, 22(8), 1889-1905. doi:10.1007/s11136-012-0344-y
- Robert Koch-Institut (Hrsg.). (2016). Gesundheit in Deutschland – die wichtigsten Entwicklungen. Gesundheitsberichterstattung des Bundes. Gemeinsam getragen von RKI und Destatis. Retrieved from [https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsGiD/2015/kurzfassung\\_gesundheit\\_in\\_deutschland.html?nn=2379316](https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsGiD/2015/kurzfassung_gesundheit_in_deutschland.html?nn=2379316)
- U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), & Center for Devices and Radiological Health (CDRH). (2009). Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims Retrieved from <https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf>
- World Health Organization Division of Mental Health and Prevention of Substance Abuse (Hrsg.). (1997). WHOQOL: Measuring Quality of Life. Retrieved from <http://www.who.int/iris/handle/10665/63482>