

# OLUtool

Nutzenbewertung für Arzneimittel ausserhalb der regulären  
Leistungspflicht

---

## Grundlagen für die Praxis



SGV  
SSMC

Schweizerische Gesellschaft der Vertrauens- und Versicherungsärzte SGV

2 0 2 2

## Impressum

© Copyright by SGV

OLUtool «Grundlagen für die Praxis», 1. Fassung 2020

OLUtool «Grundlagen für die Praxis», 2. Fassung 2022

Autor: Robert Fries

Korrespondenz: Sekretariat SGV, Markus Bonelli, Rudolf-Diesel-Strasse 5, 8404 Winterthur

[info@vertrauensaerzte.ch](mailto:info@vertrauensaerzte.ch)

## OLUtool Arbeitsgruppe

### **Dr. med. Robert Fries**

Facharzt für Allgemeine Innere Medizin (FMH) und Vertrauensarzt SGV  
Lehrgang Arbeitsmedizin/Betriebsarzt SGARM  
Mandatiertes Vertrauensarzt HELSANA sowie Beratungen in der OKP und in Tariffragen

### **Dr. med. Beat Seiler**

Facharzt für Allgemeine Innere Medizin (FMH) und Vertrauensarzt SGV  
Master of Health Administration Universität Bern  
CAS Genomic Counselor  
Vertrauensarzt tarifsuisse sowie Mandate bei verschiedenen Krankenversicherungen

### **Dr. med. Beat Kipfer**

Facharzt für Herz- und thorakale Gefässchirurgie (FMH), speziell Gefässchirurgie und Vertrauensarzt SGV  
CAS Medlaw UZH und CAS Gesundheitsökonomie ZHAW  
Lehrgang «Management im Gesundheitswesen», Universität Bern  
Leiter Vertrauensärztlicher Dienst KPT

### **Dr. med. Matthias Voisard**

Facharzt für Rechtsmedizin (FMH) und Vertrauensarzt SGV  
Master of Advanced Studies ZFH in Leadership & Management  
Leiter Vertrauensärztlicher Dienst CONCORDIA

Vertrauensärzte sind heute (noch) mehrheitlich männlich.  
Es wurde deshalb einfachheitshalber die männliche Schreibform gewählt

## Grusswort der Präsidentin SGV

Geschätzte Kolleginnen und Kollegen

Ärztenschaft, Pharmaindustrie und Versicherer stützen ihre Beurteilungen mittlerweile auf das 2018 erstmals präsentierte OLUtool der SGV ab. Die Modelle erhalten auch zunehmend Interesse und Wertschätzung durch das BAG. Die anfängliche Skepsis ist gewichen und OLUtool hat heute seinen festen Platz in der Nutzenbewertung gefunden. Die Verantwortung für die Pflege dieser Arbeitsgrundlage, für die Anpassung an den medizinischen Fortschritt und für die Weiterentwicklung aufgrund der täglichen Anforderungen in der vertrauensärztlichen Arbeit, ist gross. Die SGV nimmt diese Verantwortung gerne wahr und ist stolz darauf, damit einen wesentlichen Beitrag zur adäquaten Versorgung von Medikationen ausserhalb der Spezialitätenliste zu leisten.

Mit dem OLUtool-Update 2022 dürfen wir eine präzierte und damit qualitativ für alle Anwender verbesserte Version vorstellen. Wichtiger Meilenstein ist die Zuordnung von Studien aus allen Fachdisziplinen ausserhalb der Onkologie unter einem gemeinsamen System von sechs Studientypen, welche in der Beurteilung im übergeordneten Clinical Benefit münden. Mit diesem generischen Ansatz konnte die Heterogenität der Studien im NonOnko-Bereich überwunden werden. Neu wurden in den Modellen die Ergebnisse von Molekularpathologie und genetischen Analysen berücksichtigt, die therapeutisch immer mehr an Bedeutung gewinnen. Ebenfalls neu wurden Studien aus der Hämatologie differenzierter bewertet.

Das vorliegende Scriptum beinhaltet die Grundlagen zum Modellverständnis und erlaubt dem Anwender, die Methodik der Modelle zu verstehen. Ein grosser Dank geht an die OLUtool-Arbeitsgruppe, welche die Entwicklung verfolgt, die Modelle immer wieder aktualisiert und dieses Script erstellt hat. Jeder Abschluss ist aber wieder Neuanfang. Die Arbeit geht weiter, die Fortschritte in der Medizin sind rasant. Dies bedingt eine sorgfältige Beobachtung der Entwicklungen und erfordert in Zukunft weitere Modell-Anpassungen. Die SGV ist bereit, diese Modellpflege qualitativ hochstehend weiterzuführen.

Dr. med. Ursula Schafroth

# Inhalt

<b>1</b>	<b>Allgemeines zu OLUtool</b>	<b>6</b>
	1.1 Ausgangslage	
	1.2 Erwartbarer grosser therapeutischer Nutzen	
	1.3 Wann ist ein Nutzen gross?	
	1.4 Das Modell	
<b>2</b>	<b>Modellkriterien</b>	<b>13</b>
	2.1 Publikation	
	2.2 Studienphasen	
	2.3 Crossover Studien	
	2.4 Risikokennzahlen (p-Wert und CI)	
	2.5 Ansprechen (Response Rate)	
	2.6 Vergleich der Ereignisse: Hazard Ratio (HR)	
	2.7 Adverse Events (AE)	
	2.8 Quality of Life (QoL)	
	2.9 Studiendefizite	
	2.10 Expertenbeizug	
	2.11 Molekularpathologie	
	2.12 Durchlässigkeit der Modelle	
<b>3</b>	<b>Besonderheiten OLUtool-Onko</b>	<b>19</b>
	3.1 Kaplan-Meier-Kurve	
	3.2 Varianten	
	3.3 LL95%HR versus Punktschätzer-HR Wert	
<b>4</b>	<b>Besonderheiten OLUtool-NonOnko</b>	<b>22</b>
	4.1 Clinical Benefit	
	4.2 NNT als Bewertungskriterium	
<b>5</b>	<b>Vorgehen beim Rating</b>	<b>25</b>
<b>6</b>	<b>Rechtsprechung</b>	<b>26</b>
<b>7</b>	<b>Literatur</b>	<b>27</b>
<b>8</b>	<b>Anhang</b>	<b>28</b>
	A1 Übersteuerung eines SL Präparates (leicht gekürzt)	
	A2 Art. 71a-d KVV, BAG 2021	
	A3 OLUtool Version 3.4 (Update 2022)	

# 1 Allgemeines zu OLUtool

## 1.1 Ausgangslage

Im März 2011 wurden im Art. 71a-d der KVV die Ausnahmebestimmungen erlassen, wann ein Arzneimittel ausserhalb der Spezialitätenliste leistungspflichtig werden kann. Dem Vertrauensarzt obliegt es, diese Begehren hinsichtlich Wirksamkeit und Zweckmässigkeit im Einzelfall medizinisch zu prüfen. Im Zentrum der vertrauensärztlichen Arbeit stehen zwei Aufgaben:

- Kommt Art. 71a-d KVV zur Anwendung?
- Falls JA, Bestimmung des therapeutischen Nutzens in den Abstufungen von A – D (Rating)

Ab 2012 existierten zwei Modelle zur Nutzenbestimmung:

→ MediScore-Modell, erarbeitet vom vertrauensärztlichen Dienst HELSANA

→ 9-Felder-Modell, erarbeitet vom vertrauensärztlichen Dienst VISANA

Beiden Modellen gemeinsam waren die Nutzenbeurteilungsstufen A - D. Die Existenz von zwei Modellen gab jedoch Anlass zu Kritik, dass die vertrauensärztliche Beurteilung zu wenig einheitlich sei. Deshalb beauftragte die Schweizerische Gesellschaft der Vertrauens- und Versicherungsärzte SGV im Jahr 2018 eine Arbeitsgruppe, je ein Modell zur Nutzenbestimmung für den Bereich Onkologie und ausserhalb Onkologie, zu erarbeiten. Daraus ergaben sich die zwei Varianten OLUtool-Onko und OLUtool-NonOnko. Regelmässige Updates sollen die Aktualität der Modelle gewährleisten (1) (2).

**Mit der Einführung von OLUtool-Onko und OLUtool-NonOnko im Herbst 2018 wurden die Vorgängermodelle ausser Kraft gesetzt. Beibehalten wurden die Nutzenbeurteilungsstufen A - D. Die SGV empfiehlt ihren Mitgliedern, jede Beurteilung auf Basis von OLUtool durchzuführen. Der Vertrauensarzt kann kraft seiner unabhängigen Stellung nach Art. 57 KVG bei der Beurteilung vom Modell abweichen, soll dies aber begründet festhalten.**

### OLUtool Arbeitsgruppe 2018/19



Im Uhrzeigersinn: Ursula Schafroth, Beat Seiler, Robert Fries, Beat Kipfer

### Neubesetzung der Arbeitsgruppe ab 2020

Rücktritt von Ursula Schafroth nach der Wahl zur Präsidentin der SGV. Mit Matthias Voisard konnte die Arbeitsgruppe ab 2020 einen kompetenten Nachfolger finden.

## Bemerkung zur Semantik des Begriffs «OLU»

OLU als Kürzel von Off-Label-Use bedeutet im engeren Sinn, dass ein Arzneimittel ausserhalb der zugelassenen Indikation (Label) eingesetzt wird. Im allgemeinen Sprachgebrauch bezeichnen wir jedoch Arzneimittel ausserhalb der Spezialitätenliste (SL) als OLU-Medikamente, weil sie damit ausserhalb der regulären Kassenpflicht stehen. Dadurch entsteht die Begriffsunschärfe, dass OLU-Medikamente im «off-label» und im «in-label» zur Anwendung kommen. Viele Medikamente sind nur deshalb «OLU», weil die Limitation nicht erfüllt ist oder weil die SL Aufnahme wegen gescheiterten Preisverhandlungen verweigert wurde. Sie sind jedoch swissmedic zugelassen und werden innerhalb des Labels (in-label) eingesetzt.

Präparate, die lediglich im Ausland (FDA, EMA) zugelassen sind, werden nur dann im Art. 71a-d KVV akzeptiert, wenn sie tatsächlich innerhalb des Labels, also innerhalb der im Ausland zugelassenen Indikation, angewendet werden. Nur swissmedic zugelassene Arzneimittel geniessen das Privileg, effektiv «off-label» eingesetzt zu werden.

**Rund die Hälfte aller Medikationen im Art. 71a-d KVV werden innerhalb des Labels angewendet. Trotzdem sprechen wir pauschal von OLU-Medikamenten und meinen damit, dass diese ausserhalb der SL eingesetzt sind.**

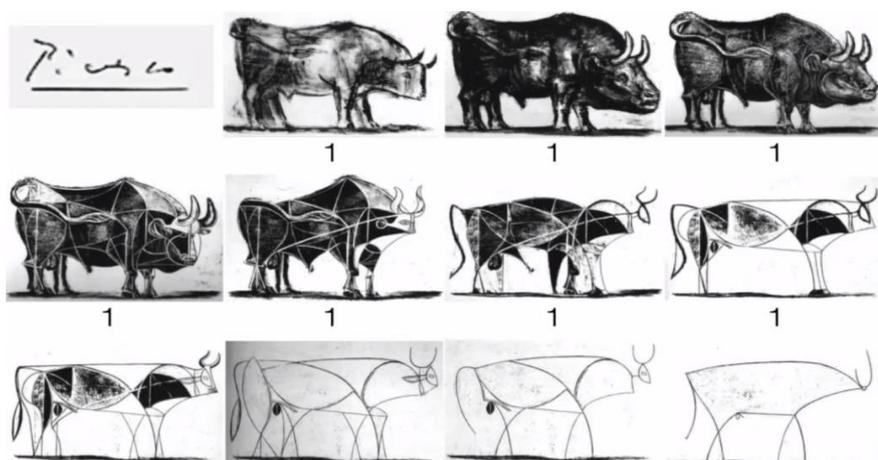
## OLUtool und Vertrauensarzt

OLUtool ist ein Modell zur Nutzenbestimmung einer OLU-Medikation. Der Zweck des Modells ist die rasche Nutzenbeurteilung durch den Vertrauensarzt (VA), um die grosse Menge an Gesuchen im Art. 71a-d KVV in nützlicher Frist zu bearbeiten, damit die Versicherten zu ihren Behandlungen kommen. Das Modell soll einfach und dennoch von guter Qualität in der Aussage sein. Dies ist möglich, weil sich OLUtool auf Erfahrungen der Vormodelle MediScore und 9-Felder-Modell abstützen kann. Zudem ist OLUtool-Onko an das valide Nutzenmodell der Europäischen Gesellschaft für Medizinische Onkologie ESMO angelehnt → «Magnitude of Clinical Benefit Scale 1.1» von 2017. (3)

Der Vorteil des Modells besteht darin, dass jede Studie mit den gleichen Kriterien beurteilt wird. Auch nichtärztliche Fachpersonen können ein Studienrating durchführen. Die Schlussbeurteilung muss ärztlich erfolgen. Nur der VA ist aufgrund seiner Qualifikation und klinischen Erfahrung befähigt, den therapeutischen Stellenwert einer Medikation angemessen zu beurteilen und nur der VA kann einschätzen, wann Experten beizuziehen sind. Art. 71a-d KVV verlangt daher zu Recht, dass die Nutzenbeurteilung durch den VA erfolgen muss.

→ Art. 71a-d KVV: «... übernimmt die Kosten des Arzneimittels nur auf besondere Gutsprache des Versicherers nach vorgängiger Konsultation des Vertrauensarztes oder der Vertrauensärztin .... ».

**Wir sind uns bewusst, dass ein Modell die Wirklichkeit nicht voll abbildet. Wir haben aber den Anspruch, mit OLUtool nahe an die Wirklichkeit heranzukommen → Welcher Picasso Stier könnte zu OLUtool passen?**



## 1.2 Erwartbarer grosser therapeutischer Nutzen

OLUtool bewertet den Nutzen im Art. 71a-d KVV als patientenrelevanten therapeutischen Effekt. Massgeblich sind die Verbesserung des Gesundheitszustandes, die Verkürzung der Krankheitsdauer, das verlängerte Überleben, die Verringerung von Nebenwirkungen und die Verbesserung der Lebensqualität.

Ein grosser therapeutischer Nutzen kann erwartet werden, falls aufgrund der Studienqualität und der Studiendaten es wahrscheinlich ist, dass das Medikament klinisch relevant ist und der Patient bezüglich Lebenserwartung und Lebensqualität profitieren kann. Surrogatendpunkte wie Parameter zu (patho)-physiologischen Funktionen, Bildgebungen, Laborwerten und anderes, werden berücksichtigt, falls diese Werte validiert und patientenrelevant sind. Beurteilt werden zusätzlich die Ansprechrate und die Nebenwirkungen.

Kein grosser therapeutischer Nutzen kann erwartet werden, falls im Vergleich zu einer zweckmässigen alternativen Therapie oder zum natürlichen Krankheitsverlauf, lediglich eine moderate Verbesserung durch die Studiendaten belegt ist, was sich insbesondere in der Verringerung von nicht schwerwiegenden Symptomen der Erkrankung oder in einer nur mässigen Abschwächung von Nebenwirkungen zeigt.

**Wenn wir vom therapeutischen Nutzen sprechen, meinen wir immer den Zusatznutzen (Mehrnutzen) für den Patienten gegenüber dem Spontanverlauf unter best care, gegenüber einer Plazebo- oder Standardtherapie → Der VA bestimmt dazu die Nutzendifferenz ( $\Delta$ ).**

### Duales Konzept von OLUtool

Der Vertrauensarzt (VA) bestimmt den Patientennutzen, der von einer Medikation im Einzelfall erwartet wird. Dies im Unterschied zu den Studien, wo der therapeutische Nutzen nicht für eine Einzelperson, sondern für die Studienpopulation beschrieben wird. Aufgrund dieser zwei Aspekte folgt eine Zweiteilung des Modells. Die Zweiteilung von OLUtool entspricht auch der aktuellen Gerichtspraxis (4).

- Beurteilung des therapeutischen Nutzens für die Studienpopulation = Studien Rating
- Beurteilung des therapeutischen Nutzens, erwartbar im Einzelfall = Abschliessendes Rating im Einzelfall

### Erwartbarkeit

Die Erwartbarkeit, dass der Nutzen eintritt, kann variieren von „möglich“ bis „sicher“. Zur Beurteilung dazu stützt sich der VA auf die Studiendaten. Je grösser die Evidenz der Studiendaten und je besser der Einzelfall zur Studienpopulation passt, desto wahrscheinlicher ist es, dass der erwartete Nutzen im Einzelfall eintritt. Zusammengefasst lässt sich sagen → Falls für die Studienpopulation ein grosser Nutzen ausgewiesen ist, kann erwartet werden, dass auch im Einzelfall der grosse Nutzen eintritt, wenn der Einzelfall zur Studienpopulation passt. Im Zweifelsfall kann eine Probestherapie eine fragliche Erwartbarkeit klären (siehe Kategorie C).

### Studienrating und Einzelfallrating

Die Rechtsprechung hat wiederholt festgehalten, dass der VA im Art. 71a-d KVV den Einzelfall zu betrachten hat. Ausgangsbasis ist immer die Datenlage der Studie, da die therapeutische Wirkung auf wissenschaftlichen Daten beruhen muss (WZW Art. 32 KVG). Dabei muss allerdings nicht ein umfassender wissenschaftlicher Nachweis vorhanden sein. Der VA weiss, dass die Studienpopulation eine ausgewählte Gruppe darstellt und im Alltag sich der Einzelpatient vielfältiger präsentiert, was zu Differenzen zwischen Studienpopulation und Einzelfall führen kann.

Beispiele von Differenzen der Studienpopulation versus Einzelfall

Differenz klein	Die Studienpopulation umfasst die Altersgruppe 40 – 75 und der Einzelfall ist 80 Jahre alt	Differenz tolerieren
Differenz gross	Die Studienpopulation basiert auf einem anderen genetischen Merkmal als der Einzelfall	Differenz nicht tolerieren und Antrag zurückweisen
Differenz unklar	Indikation ist korrekt, jedoch basiert die Studienpopulation auf einem anderen Messparameter als der beantragte Einzelfall	VA soll mit Facharzt Kontakt aufnehmen

**In der Praxis zeigt sich, dass wenige Abweichungen im Einzelfallrating vom Studienrating vorkommen. Mehrheitlich erfolgen Anträge, welche zur Studienpopulation passen und der Wirkungsnachweis wissenschaftlich fundiert ist. Abweichungen sind vom VA zu hinterfragen.**

**Beispiel: Einzelfall passt nicht zur Studie**

Bei einem Patienten mit Hodgkin Lymphom soll eine Therapie mit Panobinostat (Farydak®) erfolgen. Panobinostat ist swissmedic zugelassen zur Behandlung bei Multiplem Myelom. Es handelt sich also um eine echte off-label-Situation. Vorgelegt wird eine Studie zum Hodgkin Lymphom in der alle Probanden vorgängig eine Stammzelltransplantation erhielten.

**Das Gericht sagt:** Der allgemeine Wirkungsnachweis auf Basis der Studie könne nicht erbracht werden, da der Einzelpatient vorgängig keine Stammzelltransplantation hatte. Demzufolge wäre auch eine Probetherapie nicht zulässig. Das kantonale Gericht und später auch das Bundesgericht haben die Krankenversicherung gestützt, dass die Therapie mit Panobinostat in dieser Einzelsituation nicht leistungspflichtig und auch nicht zulässig im Art. 71a-d KVV ist, weil der Einzelfall nicht zur Studie passt (5).

Folgende Aspekte sind wichtig:

**Off-label:**

Der Einsatz eines Arzneimittels ausserhalb der regulär zugelassenen Indikation (off label) erfordert ein besonderes Mass an Sorgfalt, da nur freie Studien ohne behördliche Prüfung vorliegen, welche die Indikation begründen. Die Studie muss daher gut zum Patienten passen, ansonsten die Behandlung ethisch nicht zu vertreten ist.

**In-label:**

Erfolgt die Anwendung in-label, muss die Indikation der behördlich zugelassenen Indikation entsprechen. Der VA hat die Freiheit, zusätzliche Bedingungen zu verlangen, falls in der Zulassungsstudie relevante Besonderheiten vermerkt sind. Er soll zusätzliche Limitationen begründen können.

**Downgrading und Upgrading**

Der VA ist verpflichtet zu prüfen, ob der Versicherte zur freien Studie (off-label) oder zur Zulassungsstudie (in-label) passt. Die Erfahrung im Alltag zeigt, dass der Einzelpatient in der Regel gut zur Studienpopulation oder zum Label passt, sodass es wegen Divergenzen eher selten zu einem Up-/Downgrading des Einzelnutzens oder sogar zur Ablehnung des Antrages kommt. In unklaren Situationen wird der VA mit dem behandelnden Arzt Kontakt aufnehmen.

## 1.3 Wann ist ein Nutzen gross?

### Kriterium des grossen Nutzens

Eine Herausforderung für die Arbeitsgruppe war das Kriterium des «grossen Nutzens». In der Literatur und auch in den Nutzenmodellen von ASCO und ESMO fehlen klare Aussagen dazu. Der Verordnungstext des BAG verlangt aber vom VA, dass er eine «rote Linie» zieht und sich klar äussert, wann ein therapeutischer Nutzen gross sei. Bei der Festlegung zur «roten Linie» haben wir uns soweit möglich an der internationalen Literatur orientiert.

- Empirisch gilt in der Medizin eine Therapie als „gut“, wenn ein Drittel der Patienten profitiert.
- Im ESMO Modell fanden wir den „Drittel“ bestätigt, da eine Hazard Ratio von 0,65 – 0,70 (= 30 – 35%), kombiniert mit einem absoluten Benefit, mit den höchsten Scorepunkten bewertet wird (3).
- Weiter fanden wir, dass Ansprechraten von klinischen Effekten sich mit der Number-Needet-to-Treat (NNT) sehr gut darstellen lassen. Eine NNT von 3-4 wird als gute Therapie beurteilt. Dies im Gegensatz zu Screenings, wo eine NNT von 20 noch als Erfolg gilt (6).

Generell beruht OLUtool bei der Nutzenbeurteilung auf einem absoluten Benefit. Die Messung einer alleinigen relativen Verbesserung bewerten wir als ungenügend. Arbeitet eine Studie ohne absolute Ergebnisse, so ist der relative Benefit akzeptiert, jedoch verbunden mit einer Reduktion der Scorepunkte.

### Drittelsregel

Im Onko-Modell wird die Drittelsregel durchgehend angewendet, weil wir uns an das ESMO Modell anlehnen.

Im NonOnko-Modell wird die Drittelsregel modifiziert:

1. Bei den Typen ① - ④ erreichen 30% Verbesserung die verlangten drei Punkte für Kategorie B. Da es sich um direkt messbare «funktionelle Verbesserungen» handelt, greift hier die Drittelsregel gut und die Verbesserung entspricht gerade einem absoluten Benefit.

2. Bei Typ ⑤ (= Ansprechrate) wäre die Drittelsregel zu streng, weil es sich um schwerwiegende Krankheitsleiden handelt. Wir arbeiten deshalb auf Basis der Absoluten-Risiko-Reduktion ARR und damit mit der Number-needed- to-treat NNT. (Siehe Kapitel 4.2).
3. Bei Typ ⑥ (= Mortalität) wäre die Drittelsregel ebenfalls zu streng, getreu dem Grundsatz, dass Todesfälle möglichst zu reduzieren sind. Der grosse Nutzen muss trotzdem mit einer Grenze definiert sein, um die behördliche Vorgabe zu erfüllen. Wir halten uns deshalb an die Limiten von ESMO, da es sich wie in der Onkologie um Todesfälle handelt.

## 1.4 Das Modell

Aus vorgenannten Überlegungen ergeben sich die Kriterien, welche ein Modell enthalten muss, um den Nutzen im Sinne von Art. 71a-d KVV zu bestimmen:

- Studienevidenz / Studienqualität / Signifikanz der Daten (7)
- Daten zu Overall Survival (OS), Progression Free Survival (PFS), Hazard Ratio (HR), Clinical Benefit (8)
- Ansprechrate generell und als ORR = Complete Remission (CR) und Partial Remission (PR)
- Wirkdauer und Nachhaltigkeit der Behandlung
- Toxizität der Therapie → Adverse Events (AE)
- Daten zur Lebensqualität (QoL)

Zentral ist, ob die Therapie mit einer relevanten Kontrollgruppe verglichen wurde. Dies wird im Modell direkt mit  $\Delta$  abgebildet. Hinweise für eine besonders gute Wirkung oder für einen Mangel werden mit einem Bonus oder Malus bewertet. Der VA hat in der Betrachtung des Einzelfalls zudem die Möglichkeit zu einem abschliessenden Up- oder Downgrading um eine Nutzenkategorie. Dies ist erfahrungsgemäss aber selten der Fall und soll jeweils begründet werden.

### Einteilung der Nutzenkategorien A – D

Art.71a-d KVV erfordert für eine Leistungspflicht aus der OKP einen erwartbaren grossen Nutzen des Arzneimittels für den Patienten. Entsprechend wird im Modell eine beträchtliche oder erhebliche Nutzenhöhe verlangt (Nutzenkategorie A oder B). Der Nutznachweis soll mit guter Evidenz belegt sein.

Nutzenkategorie	Nutzenhöhe	Bemerkungen
A	Erheblicher Nutzen (= gross)	Nutznachweis erheblich und mit hoher Evidenz belegt
B	Beträchtlicher Nutzen (= gross)	Nutznachweis beträchtlich und mit guter Evidenz belegt
C / C*	Kein grosser Nutzen	Hinweise, dass im Einzelfall ein grosser Nutzen zu erwarten ist → Probetherapie empfohlen
D	Kein grosser Nutzen	Nutznachweis nicht belegt, da fehlende oder ungenügende Daten

### Kategorie C / Probetherapie

Kategorie C mit ihrer Probetherapie gibt immer wieder zu Diskussionen Anlass, ob dies überhaupt vereinbar sei mit dem KVG Art. 32 (post hoc ergo propter hoc). Unseres Erachtens wird oft verkannt, dass die Rechtsprechung für den Nutzen in Art. 71a-d KVV eine Beurteilung in allgemeiner Weise (Studie) und eine Beurteilung des Einzelfalls verlangt (4). Die Studie basiert bei C zwar auf einem tieferen Evidenzlevel an Wissenschaftlichkeit, doch bestehen Studienhinweise, dass im Einzelfall ein grosser Nutzen erwartbar ist. Die Probetherapie kann diesen erwartbaren grossen Nutzen aufzeigen, um dem Einzelfall gerecht zu werden. Die zwei Beispiele zeigen, wie sich diese Fälle im Praxisalltag präsentieren.

Die Probetherapie selbst kann nicht vergütet werden, da erst deren Erkenntnisse zeigen, dass im Einzelfall ein grosser therapeutischer Nutzen erwartbar ist.

#### Beispiel 1

Eine neue Medikation erreicht aufgrund der wissenschaftlichen Daten einer Studie in OLUtool 3 Punkte = grosser Nutzen (B). Allerdings beträgt die Ansprechrate nur 25%. Dies führt zu einem Maluspunkt. Es resultieren dann 2 Punkte = C.  
→ Die Probetherapie soll nun zeigen, ob der Einzelfall zu dieser 25%-Gruppe gehört. Falls ein erfolgreiches Ansprechen vorliegt, kann ein «grosser Nutzen» (B) erwartet werden

## Beispiel 2

Eine Studie analysiert die Wirkung einer neuen Medikation. Die Studienpopulation reagiert nicht völlig gleich, sondern in einem range von (a-b). Der Durchschnitt von (a-b) erreicht in OLUtool nur die Nutzenkategorie C (= Beurteilung in allgemeiner Weise). 20% der Probanden im range (a-b) zeigen jedoch eine Wirkung, welche die Kriterien des grossen Nutzens erfüllen.  
→ Die Probetherapie soll nun aufzeigen, ob der Einzelfall zu diesen 20% gehört und ein «grosser Nutzen» (B) zu erwarten ist.

## Weiterbehandlung

Falls die Probetherapie einen guten klinischen Effekt zeigt, kann auf einen grossen, erwartbaren Nutzen im Sinne von Art. 71a-d KVV geschlossen und dem Krankenversicherer die Weiterbehandlung empfohlen werden. Laut OLUtool käme es zu einem Upgrading um eine Kategorie (= B). Die Vergütung ist aber Sache des Versicherers und der VA wird lediglich die Weiterbehandlung als Pflichtleistung empfehlen. Ein nächster Einzelfall müsste erneut die Probetherapie durchlaufen, da das Studienrating in der Kategorie C verbleibt, weil sich an der «Beurteilung in allgemeiner Weise» nichts geändert hat.

## Kategorie C\* (Sprich: C-Stern)

Im Regelfall beinhaltet Kategorie C, dass die Kosten der Probetherapie von der Pharma übernommen werden. Die Probe dauert meist 1-3 Monate. C\* wird vom VA empfohlen, wenn die übliche C-Lösung voraussichtlich nicht umsetzbar ist, weil die Probetherapie länger als 3 Monate andauert oder es sich um eine Einmaltherapie handelt. Der VA kann abschätzen, ob so eine spezielle Situation vorliegt. Es liegt im Ermessen des Versicherers, ob er in diesen Härtefällen im Sinne von Risk-Sharing, sich an der Probetherapie finanziell beteiligen will. Der VA unterstützt auf Wunsch den Versicherer bei der Formulierung von Limitationen oder der Messung von Therapiezielen. Ohne Lösung für die Probetherapie entfällt die Leistungspflicht, da bei C/C\* der grosse Nutzen zwar im Einzelfall möglich, aber nicht generell zu erwarten ist. Die Probetherapie soll erst starten, wenn die Limitationen, die Finanzierung und die allfällige Weiterbehandlung geregelt sind.

## Spezialfall Immunonkologika

Immunonkologika entfalten ihre Hauptwirkung erst nach 6-9 Monaten. Trotzdem mündet eine allfällige Nutzenbewertung C nicht automatisch in ein C\*. Der Grund liegt in der besonderen Wirkung der Immunonkologika.  
→ Die Immunreaktion führt zu einer Auflockerung im Tumorgewebe und damit zu einer anfänglichen Vergrösserung des Tumors. Diese Grössenzunahme nach 1 – 2 Monaten ist ein Hinweis, dass die Immuntherapie greift. Die Probetherapie kann deshalb im Rahmen des «normalen C» vom Facharzt innert 3 Monaten beurteilt werden.

## Internationale Beurteilungen auf Niveau HTA

In unklaren Beurteilungssituationen verwendet der VA auch Erkenntnisse von internationalen Fachgremien auf Niveau eines Health Technology Assessment (HTA) wie zum Beispiel von G-BA, NICE oder auch von Zulassungsbehörden wie swissmedic, FDA oder EMA. Dabei berücksichtigt er nur Aussagen zum therapeutischen Nutzen ohne Bezug zur Kosten/Nutzen Relation.

Bewertung internationaler Fachgremien:

- Betrifft die Kritik die Studienqualität, kann dies der VA als Studiendefizit mit Malus bewerten.
- Betrifft die Beurteilung die Wirkung der Medikation als Ganzes, kann dies der VA im abschliessenden Einzelfallrating mit einem allfälligen Up- oder Downgrading berücksichtigen.

Im Alltag verwenden wir in unklaren Situationen gerne die HTA Beurteilungen des Gemeinsamen Bundesausschusses in Deutschland (G-BA), weil deren Bewertungsskala gut mit unserer OLUtool Nutzenbewertung übereinstimmt:

Beurteilungsraster G-BA versus OLUtool

Zusatz-Nutzen G-BA	Erheblich	Beträchtlich	Gering	Nicht belegt
Nutzen OLUtool	A	B	C	D

## Studienlevel und Studienrating

OLUtool berücksichtigt die veränderten Zulassungsverfahren und neuen Studiendesigns und ist auch anwendbar für Studien, welche einen tieferen Evidenzlevel als Phase-III-Studien aufweisen. Es ist nicht mehr so, dass ein Medikament nur mit einer Phase-III-Studie die behördliche Zulassung erhält. Basket Studien haben die klassischen Vorgehensweisen aufgebrochen. Auch drängen die Pharmafirmen zunehmend früher auf den Markt, oft vor dem regulären Studienabschluss, nur mit Zwischenauswertungen (Interimsanalysen) von laufenden Studien. Dazu kommt der «Unmet Need» von Patienten und deren Ärzte, die für schwierige Situationen möglichst rasch wirksame Therapieoptionen verlangen.

Viele Studien werden nur mit PFS Werten publiziert, ohne OS abzuwarten. Die Erkenntnis, dass bei Tumorstadium-Treibern genetische Mutationen im Tumorgewebe verantwortlich sind, führt dazu, dass bei einer Krankheitsprogression eine Genanalyse gemacht wird. Wird ein neuer und starker Treiber eruiert, richtet sich die Therapie auf den neuen Treiber. Somit verliert OS an Bedeutung für die Therapiewahl. Der Schrittmacher für die Therapieeinstellung wird immer öfter PFS aufgrund der molekularpathologischen Ergebnisse → Personalized Medicine.

→ OLUtool ist so aufgebaut, dass eine Beurteilung unabhängig vom Studienlevel möglich ist. Voraussetzung ist allein, dass das Arzneimittel behördlich zugelassen wurde. Werden der Raster und die Regeln des Modells beachtet, fallen die Studienratings unabhängig vom durchführenden VA in der grossen Mehrheit gleich aus.

**Das Studienrating ist so konzipiert, dass nur minimaler Spielraum in der Beurteilung übrigbleibt. Damit wird die Einheitlichkeit im Studienrating bezweckt.**

## 2 Modellkriterien

### 2.1 Publikation

Eine Studie ist akzeptiert, wenn diese in einer anerkannten, internationalen und peer-reviewten Fachzeitschrift publiziert ist. Ein Peer-Review erachten wir als ein wichtiges Qualitätsmerkmal. Eine publizierte Datenpräsentation wie ein Poster oder Abstract an einem internationalen Fachkongress wird akzeptiert, falls vorgängig ein Peer-Review durchlaufen wurde. Dies ist zum Beispiel im Bereich Onkologie an den grossen, internationalen Kongressen wie ESMO oder ASCO der Fall. Allerdings wird der VA einen Maluspunkt geben, falls die ganze Studie noch nicht einsehbar ist. Nicht akzeptiert sind Daten, die nur präsentiert, jedoch nicht publiziert sind oder lediglich Pharma intern erhobene Daten, die dem VA zugestellt werden.

### 2.2 Studienphasen

Studienphasen sind ausgerichtet hinsichtlich der behördlichen Zulassung. Für eine behördliche Zulassung wurde bisher eine Phase-III-Studie verlangt. Mit dem Aufkommen der Personalized Medicine (Molekulargenetik) gelangen zunehmend Phase-II-Studien zur Zulassung. Dies hat dazu geführt, dass OLUtool die Studienphase nicht direkt bewertet. Wichtig ist die Studienevidenz mit dem Kriterium, ob es eine Vergleichsgruppe gibt, ob es sich um eine Single-arm-Studie ohne Kontrollgruppe oder nur um einzeln oder aneinandergereihte Case Reports handelt.

Unterschiede Phase II und Phase III-Studien

Phase II- Studie	Dosis-Effekt Beziehung	0 - 100 Teilnehmer	Optimierung der Dosis/Wirkung steht im Vordergrund → mit/ohne random. Kontrollgruppe
Phase III- Studie	Wirksamkeit unter kontrollierten Bedingungen	100 - 1000 Teilnehmer	Annähernd praxisnahe Bedingungen → mit randomisierter Kontrollgruppe → mit oder ohne Verblindung → umfassende Beurteilung der AE

### 2.3 Crossover

Crossover bedeutet, dass die Verumgruppe so gut auf die neue Medikation reagiert, dass es ethisch nicht zu verantworten wäre, weiterhin das Verum-Medikament in der Kontrollgruppe nicht zu geben. Ab dem Zeitpunkt des Crossovers wird dadurch die Kontrollgruppe aufgehoben. Das Ergebnis der Studie entspricht dann nicht mehr dem zu Studienbeginn eingeplanten Vergleich.

OLUtool ist nicht so streng und empfiehlt, die Ergebnisse von Crossoverstudien zu akzeptieren und zwar so, wie sie präsentiert werden, auch im Wissen, dass die Ergebnisse «geschönt» sind und die ganze Studie nicht mehr den strengen Qualitätsanforderungen genügt. Das JA zu Crossover ist ein «goodwill» zu Gunsten der Patienten in schwieriger Krankheitssituation.

### 2.4 Risikokennzahlen (p-Wert und CI)

Die Kennzahlen stammen ursprünglich aus der Bevölkerungsstatistik (7) und basieren auf einer statistisch berechneten Basis. Um deren Merkmale zu erfassen, wäre eine Erhebung bei der ganzen Bevölkerung (Population) nötig, was jedoch meist nicht möglich ist. Deshalb bildete man eine repräsentative Stichprobe und diese Stichprobe wird auf die Bevölkerung übertragen. Der VA ist in der Regel kein Biostatistiker und nimmt diese Zahlen als «Anwender» einfach aus der Studie zur Kenntnis. Es ist aber hilfreich, dass er weiss, was hinter diesen Zahlen steckt, was sie bedeuten und dass es sich um errechnete Grössen handelt.

## p-Wert

Wenn wir zwei Gruppen haben und davon ausgehen, dass die Medikation in beiden Gruppen gleich gut wirkt, dann besteht eine Null-Differenz (= Nullhypothese). Mit der Studie soll bewiesen werden, dass die Null-Hypothese nicht stimmt und eine signifikante Differenz zu Gunsten der Verum-Medikation besteht.

Der britische Statistiker Sir David John Spiegelhalter sagt dazu: «Die Nullhypothese wird im Verlauf eines Experiments weder bewiesen noch aufgestellt, sondern möglicherweise widerlegt. Man kann sagen, dass jedes Experiment nur deshalb existiert, um den Fakten eine Chance zu geben, die Nullhypothese zu widerlegen».

Und zum p-Wert sagt er : «A p-value is the probability of getting a result at least as extreme as we did, if the null hypothesis (and all other modelling assumptions) were really true».

- Der p-Wert drückt aus, ob ein gefundener Unterschied zwischen zwei Gruppen statistisch signifikant ist. Er zeigt, mit welcher Wahrscheinlichkeit eine beobachtete Differenz der Messergebnisse zweier Gruppen zufällig auftritt.
- $p \leq 0,05$  bedeutet, dass die Wahrscheinlichkeit des zufälligen Auftretens maximal 5% beträgt. Diese Wahrscheinlichkeit gilt in der Medizin als Grenze zur Signifikanz. Je kleiner p, desto unwahrscheinlicher ist es, dass zufällige Ergebnisse vorliegen.
- Studien ohne Kontrollgruppe ergeben keinen p-Wert.

## Konfidenzintervall (CI)

Das CI wird auf Grundlage einer repräsentativen Stichprobe errechnet. Diese wiederum lässt auf die ganze Population schliessen. Das CI bezeichnet die Wahrscheinlichkeit mit der die erhobenen Werte der Zufallsstichproben aus der Population als repräsentativ gelten. Eine gut konzipierte Studie eines starken Arzneimittels muss zeigen, dass die erhobenen Streuwerte der Einzelprobanden einer Medikation zu 95% nicht zufällig entstanden sind. Je nach Fachgebiet sind andere CI möglich (z.B. 98%). In der Medizin ist das 95%-CI üblich.

- Das CI gibt die Streuung an, in welchem Bereich die «nicht zufälligen» Werte liegen. In OLUtool arbeiten wir mit der Hazard Ratio. Übersteigt das 95%-CI der Hazard Ratio den Wert 1,0 bedeutet dies, dass HR-Werte vorkommen, die zufällig im Intervall liegen. Die Signifikanz ist ungenügend.
- Je enger das CI, desto präziser die Studienaussage. Dazu sind ein potentes Arzneimittel und viele Probanden nötig (= hohe Studienpower).

All diese Zahlen basieren auf biostatistischen Berechnungen (9), welche die Studienmacher beim Studiendesign beachten. Der VA hat darauf keinen Einfluss. Zum Verhältnis von p-Wert und CI meint Spiegelhalter: «There is a precise correspondence between confidence intervals and P-values: if, say, the 95% interval excludes 0, we can reject the null hypothesis of 0 at  $p < 0.05$ ».

**OLUtool legt den Fokus auf die Signifikanz. Wir verlangen ein  $p \leq 0,05$  und ein 95%-CI für HR von  $< 1$**

## 2.5 Ansprechen (Response Rate)

### 1. Häufigkeit des Ansprechens

Anzahl Probanden, welche auf die Behandlung initial ansprechen mit Angabe in Prozenten.

### 2. Intensität (Wirkungsstärke) des Ansprechens

Complete Remission (CR) bezeichnet den vollen Rückgang der Krankheit, Partial Remission (PR) nur einen teilweisen Rückgang. Stable disease (SD) beschreibt lediglich das Stillstehen des Krankheitsprozesses. Im Bereich Onkologie verlangt OLUtool eine CR oder PR als Wirkung eines Arzneimittels. Es ist üblich, die Ansprechrate als ORR anzugeben  
→  $ORR = CR + PR$ .

Im NonOnko-Bereich kommt es immer wieder vor, dass durch eine Medikation lediglich das Fortschreiten eines chronischen Krankheitsprozesses gestoppt oder gebremst werden kann (analog SD in der Onkologie). In diesen Fällen kann der Nutzen anerkannt werden, falls das Aufhalten des Krankheitsprozesses nachhaltig und mit einem klinischen Nutzen verbunden ist.

### 3. Zeitliches Andauern des Ansprechens

In der Onkologie ist das Ansprechen im Rahmen der RECIST Kriterien definiert. So wird von einer CR verlangt, dass das volle Ansprechen mindestens 30 Tage anhält. Ausserhalb der Onkologie ist die Zeitdauer des Ansprechens nicht klar definiert.

- Im OLUtool-NonOnko resultiert für das Ansprechen über 12 Monate ein Bonus, falls  $CR \geq 30\%$
- Im OLUtool-Onko wird die Nachhaltigkeit mit OS, PFS oder DoR (Duration of Response) direkt mit Scorepunkten bewertet. Somit entfällt ein zusätzlicher Bonus. Ausnahme ist ein langer Verlauf des Ansprechens unterhalb der mOS (50%)-Marke. (siehe Kapitel 3.2)

## 2.6 Vergleich der Ereignisse: Hazard Ratio (HR)

Unter Hazard Rate (nicht Ratio) versteht man die Anzahl definierter Events in einer Gruppe innerhalb eines definierten Zeitraumes. Im Regelfall entspricht der definierte Zeitraum gerade der Studiendauer. Das Verhältnis der Hazard Rates von zwei Gruppen ergibt die Hazard Ratio (HR). Im Bereich Onkologie sind der Tod und die Krankheitsprogression die zwei dominierenden Events (HR-Werte für OS und PFS). Im NonOnko Bereich sind die Messparameter je nach Fachgebiet verschieden.

→ Mit dem Ausdruck HR ist ohne weitere Präzisierung immer Hazard Ratio gemeint. Eine HR gibt es nur, wenn es eine Kontrollgruppe gibt.

Der HR-Wert allein sagt zwar etwas über die Differenz der Therapiewirkung zwischen zwei untersuchten Gruppen aus, jedoch ist die Aussage allein ungenügend, weil eine Differenz auf hohem Niveau (über 50%) oder die gleiche Differenz auf tiefem Niveau (unter 50%) liegen kann. Deswegen wird im Modell der HR-Wert kombiniert mit einer absoluten Wirkdauer. Diese Kombination von relativem und absolutem Nutzen bildet dann den Gesamtnutzen einer Therapie. HR-Werte allein, ohne Verknüpfung mit einer absoluten Wirkdauer, sind möglich und im Modell akzeptiert, ergeben aber eine geringere Punktezahl.

**Im OLUtool gilt, dass alleinige HR-Werte nur dann akzeptiert sind, wenn absolute Werte fehlen. Visualisiert ist der HR-Wert als Spreizung der Kurve von zwei Gruppen direkt sichtbar. Studien ohne Vergleichsgruppe haben keine HR-Werte.**

## 2.7 Adverse Events (AE)

Im Onko Bereich sind die AE Angaben mittels Gradeinteilung standardmässig aufgeführt. Die Studien nehmen meist Grad 3 + 4 zusammen. Grad 3 + 4 bedeutet, dass eine Dosisänderung oder aufwendige Zusatztherapien nötig sind. Grad 1 + 2 entsprechen milden Nebenwirkungen, welche ohne Einfluss auf die Hauptwirkung der Therapie sind. Deswegen berücksichtigen wir diese AE im Modell nicht. Grad 5 (Todesfälle) sind in den Kontext der ganzen Krankheits- und Studiensituation einzuordnen und spielen im Ratingprozess selten eine Rolle. Kommt dazu, dass bei zu vielen Todesfällen kein Arzt so eine Therapie anwenden wird.

Im NonOnko Bereich sind die Aussagen zu den AE sehr heterogen. Üblicherweise entsprechen die Grad 3 + 4 Nebenwirkungen pauschal den serious AE (sAE). Leichte AE, welche mit unwesentlichem Aufwand oder mit geringer Einschränkung im Alltag behandelt werden können, werden nicht gezählt. Dies analog zu Grad 1 + 2 im Bereich Onkologie.

In der neuen Version 3.4 von OLUtool haben wir die Limiten für den Maluspunkt verändert. Aus der Ärzteschaft kam oft der Vorwurf, dass wir uns immer in schwerwiegenden Krankheitssituationen befinden. Die Betroffenen würden Nebenwirkungen in Kauf nehmen, da es oft um Leben oder Tod gehe. Es sei übertrieben, dass eine Therapie vom Versicherer eventuell nicht übernommen würde, nur wegen etwas mehr Nebenwirkungen. Wir haben dieser Kritik Rechnung getragen und geben den Malus erst, wenn  $\Delta AE \geq 30\%$  Grad 3+4 oder sAE aufweist und bei Studien ohne Kontrollgruppe ab  $\geq 50\%$ .

### Spezialfall Hämatologie

Aus Hämatologenkreisen kam die Kritik, dass ein Maluspunkt resultiert, weil die hämatologischen AE regelmässig hoch sind. Dies sei nicht gerechtfertigt, da die «schlechten» Hämatologiewerte zum Krankheitsbild gehörten. Zusatzmedikationen wegen Anämie oder Neutropenie sollten daher nicht negativ bewertet werden. Wir tragen diesem Einwand Rechnung und bewerten die hämatologischen AE nicht mehr. Es zählen hinsichtlich Maluspunkt nur die AE ausserhalb der Hämatologie.

## 2.8 Quality of Life (QoL)

Die Messung der Lebensqualität ist schwierig objektiv zu erheben. Nur validierte Befragungen sind aber verwertbar. Damit QoL-Aussagen bewertet werden können, muss deshalb angegeben sein, mit welchen Fragebogen gearbeitet wurde. Der Vertrauensarzt muss überprüfen können, dass es sich um validierte Befragungen handelte. Im Zweifelsfall sind die Angaben wertlos. Damit eine QoL zu einem Bonus führt, muss die QoL signifikant besser sein als die QoL der Kontrollgruppe. Lediglich «non inferior» genügt nicht.

## 2.9 Studiendefizite

In der Regel ist die Studienqualität gut, weil wir nur Daten akzeptieren, welche ein Peer-Review durchlaufen haben. Dennoch soll der VA die präsentierten Studien kritisch durchsehen. Auch wenn er zurückhaltend Maluspunkte wegen Studiendefiziten vergibt, soll er offensichtliche Mängel mit einem Malus bewerten. Hilfreich sind Kritiken von HTA-Fachgremien oder Zulassungsbehörden wie G-BA, NICE sowie swissmedic/EMA oder FDA. Liegen relevante Kritiken zur Studienqualität vor, so kann sich der VA auf diese professionellen Institutionen berufen und einen Maluspunkt geben. In jedem Fall ist aber ratsam, die Begründung für den Maluspunkt schriftlich festzuhalten.

Es sind nicht immer statistische Fallstricke, sondern auch klinische Aspekte, welche auf die Qualität einer Studie hinweisen. Es ist z.B. wichtig, dass das Morbiditätsprofil der Verumgruppe und der Kontrollgruppe in etwa gleich verteilt sind. Diese Beziehung kann der VA mit seiner klinischen Erfahrung kompetent einschätzen.

### Beispiel

Eine Studie beschreibt die Wirkung einer Medikation bei einem Leiden, welches in die Schweregrade I – III eingeteilt wird. Das Prüfmedikament zeigt einen hohen Benefit. Der hohe Benefit in der Verumgruppe beruht auf dem Überhang von leichten Fällen (Stadium I). Auch hier hilft die Drittelsregel → Differiert der geringste Schweregrad um mehr als ein Drittel, ist ein Maluspunkt angebracht. In unserem Beispiel für Stadium I also 10% versus 50% →  $\Delta = 40\%$ . Im Einzelfall entscheidet der VA.

Stadium	Verteilung der Verumgruppe	Verteilung der Kontrollgruppe
I	50%	10%
II	30%	40%
III	20%	50%

## 2.10 Expertenbeizug

### Expertenmeinung bei unklarer klinischen Relevanz

Die VA sind nicht in jedem Fachgebiet gleichzeitig auch Spezialisten. Trotzdem ist es nicht so, dass immer ein Facharzt die Beurteilung vornehmen muss, um ein Studienrating durchzuführen.

Warum ist das so?

Der grosse Vorteil der Modelle besteht darin, dass OLUtool die relevanten Kriterien einer Studie bewertet. Das Fachwissen ist sozusagen aufgearbeitet und ins OLUtool transformiert. Der VA urteilt nicht über einen fachlichen Sachverhalt, sondern über die Studie. Deshalb kann er mit allen Fachärzten über die Studienergebnisse diskutieren ohne selbst Facharzt in der betreffenden Disziplin zu sein. Dennoch gibt es Situationen, für die es ratsam ist, Experten beizuziehen, wie dies auch die Gerichtspraxis verlangt.

### Beispiel 1

Bei Studienprobanden mit spinaler Muskelatrophie werden Atemvolumina gemessen. Dabei werden zwei Subgruppen definiert: Einmal eine Verbesserung um 100 ml, einmal mit 200 ml als relevante Grösse → Sind diese Volumina klinisch relevant? Oder nur bei einer Subgruppe? Der VA als Nicht-Neurologe wird die Expertenmeinung benötigen, sei es in Form eigener Literatur-Recherche oder im persönlichen Kontakt.

### Beispiel 2

Eine neue Medikation soll bei transfusionsbedürftiger Myelofibrose eingesetzt werden. Gegenüber der Standardtherapie kann die Bluttransfusion pro Monat um 1 Beutel Blut reduziert werden. Ist das klinisch relevant? → Der VA als Nicht-Hämatologe wird die Expertenmeinung benötigen, sei es in Form eigener Literatur-Recherche oder im persönlichen Kontakt.

## Stellenwert von Experten

Es ist einfach, nach Experten zu rufen, jedoch ist es schwierig, eine unabhängige Expertenmeinung zu finden und noch schwieriger ist es, einen Expertenkonsens zu finden, wann ein Nutzen gross sei. Der VA wird in einer ersten Phase selber recherchieren. Oft findet er kritische Berichte von unabhängigen Expertengremien im In- und Ausland, welche die Frage nach der klinischen Relevanz eines Parameters oder der ganzen Studie beantworten. Bleibt die Recherche unklar, wird er den persönlichen Kontakt mit universitären Fachstellen in der Schweiz suchen. Falls mindestens zwei UNI-Kliniken eine gleiche Empfehlung abgeben, kann dies als guter Standard betrachtet werden. Bleibt die Relevanz widersprüchlich und ungeklärt, wird der VA einen Maluspunkt geben oder die Studie mangels Relevanz zurückweisen.

## 2.11 Molekularpathologie

### Orphanisierung und Personalisierung der Medizin

Die Molekularpathologie nimmt immer breiteren Raum ein. Viele Studien beziehen sich explizit auf genetische Merkmale. Damit werden immer mehr Untergruppen von Indikationen geschaffen, die oft nicht mehr in der SL abgebildet sind. In der Folge steigen die Anzahl Fälle im Art. 71a-d KVV → Orphanisierung und Personalisierung der Medizin. Eine neue Herausforderung für OLUtool stellen die Basketstudien dar. Hier werden Medikationen in ganz verschiedenen Organsystemen getestet und hinsichtlich Ansprechrates ausgewertet. Basis der Medikation ist in der Regel eine genetische Mutation, welche als Tumorstadium-Treiber wirkt.

**OLUtool hat in der neuen Version 3.4 auf diese Entwicklung eine Antwort gegeben. Einerseits mit der Erweiterung in Rubrik 5 im Onko-Modell. Andererseits meinen wir, dass Ergebnisse lediglich für eine Genmutation maximal Studienrating C hergeben kann, weil eine konkrete, organbezogene Evidenz fehlt. Zudem handelt es sich meist um Basketstudien ohne Kontrollgruppe, was ebenfalls maximal C ergibt.**

Ähnlich sieht es aus bei austherapierten Krebspatienten, welche aufgrund einer Analyse mit einem Grosspanel eine Therapieoption erhalten. Das Ergebnis ist ebenfalls eine Genmutation, welche einen Tumorstadium-Treiber aktiviert. Die Therapie wird aufgrund der genetischen Mutation ausgewählt, auch wenn vielleicht gar keine Studie dazu vorliegt. Wegen fehlenden Daten (ausser dem Genpanel) wären wir in der Kategorie D. Stammt der Antrag von einem interdisziplinären Tumorboard, kann der VA ein Upgrading von D zu C ermöglichen.

Die molekulargenetische Mutationsanalyse im Tumorgewebe mit Grosspanels bis zu 500 Mutationen wird zunehmend häufiger angewendet. Auch geht der Trend dahin, dass bei immer mehr Tumorentitäten diese Grosspanels schon vor dem ersten Therapieschritt eingesetzt werden.

OLUtool Onko	Studien mit einer Genmutation als 1° Endpunkt oder Genmutationen als Tumorstadium-Treiber ohne Studien (Grosspanel)	→ Pauschal Kategorie C → Upgrading von D zu C möglich, falls ein interdisziplinäres Tumorboard die Medikation befürwortet
Olutool NonOnko	Nur Case Reports oder Studien mit Kategorie D im Studienrating	→ Upgrading von D zu C möglich, falls ein Referenzzentrum für seltene Krankheiten die Medikation befürwortet

### Number-of-1-trial

Ziel einer N-of-1-Studie (Studie am Einzelpatient) ist es, die optimale oder beste Intervention für den einzelnen Patienten anhand objektiver datengesteuerter Kriterien zu bestimmen. Am Anfang einer Therapie steht das diagnostische Grosspanel zur Eruiierung von Genmutationen. Zusätzlich erfolgen individuelle Betrachtungen zur pharmakogenetischen Wirkung hinsichtlich Dosierung und AE. Damit wird sozusagen auf Basis der individuellen molekularbiologischen Daten eine personalisierte Studie für den Einzelpatienten erstellt mit dem Ziel, die bestmögliche Medikation zu bestimmen.

**«Jeder Patient erhält seine eigens auf ihn zugeschnittene Therapie». Wie weit dieses Konzept im Klinikalltag umgesetzt wird, bleibt heute noch offen.**

## 2.12 Durchlässigkeit der Modelle

Es gibt Situationen, in denen eine Studie mit dem dazugehörigen Modell nicht geratet werden kann. Typischerweise betrifft dies Studien im Bereich Hämatologie. Grundsätzlich wäre OLUtool-Onko das richtige Modell. Oft ist in der Hämatologie der 1° Endpunkt eine Ansprechrate, für welche Rubrik 5 vielleicht keine Antwort gibt (12). Trotzdem besteht eine klinische Relevanz für die Medikation und ein Rating ist notwendig. Der Ausweg bietet sich im Modellwechsel an. Die Scorepunkte werden nicht im Onko, sondern im NonOnko-Modell bestimmt, das sich sehr gut für Ansprechraten eignet. Für Bonus/Malus gilt wieder das Onko-Modell, weil die Hämatologie zur Onkologie gehört.

Umgekehrt könnte eine NonOnko-Studie auch über das Onko-Modell abgehandelt werden. Erfahrungsgemäss ist dies selten der Fall.

**Die Modelle sind so konzipiert, dass ein Wechsel möglich ist und dass sie adäquate Ergebnisse liefern. Es soll aber nur dann gewechselt werden, wenn mit dem angestammten Modell kein Rating erbracht werden kann.**

# 3 Besonderheiten OLUtool-Onko

## 3.1 Kaplan-Meier-Kurve

Bei den Studienangaben ist gut darauf zu achten,

- ob eine absolute Aussage gemacht wird → nach 18 Monaten leben noch 60% der Probanden
- ob es heisst, OS betrage 18 Monate. Damit ist immer mOS, also das mediane OS (50%) gemeint und dieses betrage 18 Monate.

OS/PFS-Werte werden mittels Kaplan-Meier-Kurve dargestellt. Es ist wichtig zu wissen, dass die Kurve durch eine Berechnung zustande kommt. → So stellt z.B. die Kaplan-Meier-Kurve für OS eine mathematisch errechnete Überlebenswahrscheinlichkeits-Kurve dar.

### Beispiel: OS beträgt 18 Monate

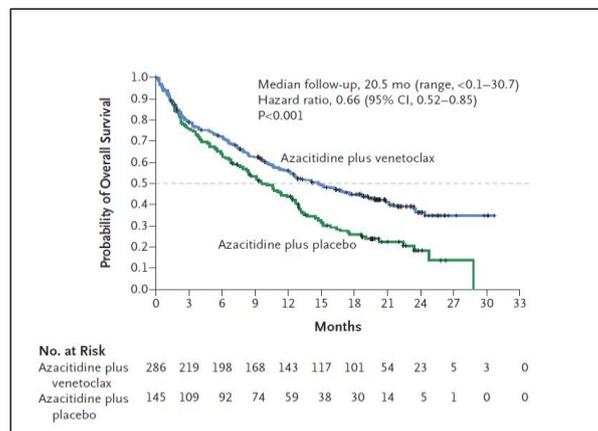
Was heisst das? → Wenn ein Proband diese Behandlung beginnt, kann er mit einer Wahrscheinlichkeit von 50% damit rechnen, dass er nach 18 Monaten noch am Leben ist.

Was heisst es nicht? → Dass nach 18 Monaten noch 50% der Probanden leben.

Wie viele Probanden der Studienpopulation zu welchem Zeitpunkt tatsächlich lebten, ist unterhalb der Kurve als «No. of patients at risk» vermerkt. Statistische Aussagen bis wann die Studie valide ist, müssen rechnerisch ermittelt werden. Als «Faustgrösse» verlangen wir VA mindestens 15 lebende Personen in der Verumgruppe und setzen dort die Grenze für die valide Studienaussage, falls die Studie selbst diese Grenze nicht angibt. Normalerweise ist die valide Studienaussage zeitlich kürzer als die Kurve, weil die Kurve meist bis zum letzten lebenden Probanden weitergeführt wird. Die gleichen Überlegungen gelten in Analogie auch für PFS.

### Beispiel

Die Studienautoren setzen die Validität der Studienaussage bei 20,5 Monaten fest (median follow-up) → Es leben noch rund 14 Probanden in der Placebogruppe und sogar 54 in der Verumgruppe. Ohne valide Studienangabe würde OLUtool in diesem Fall rund 25 Monate akzeptieren (= 15 lebende Patienten in der Verumgruppe).



### Kaplan-Meier-Kurve

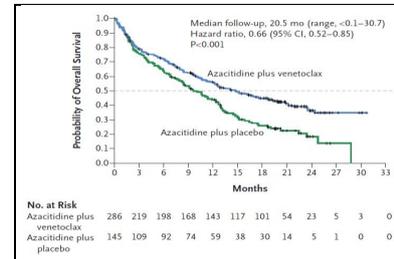
- Es handelt sich um eine errechnete Wahrscheinlichkeitskurve
- OLUtool-Onko akzeptiert die Zeitdauer der Kaplan-Meier-Kurve wie sie in der Studie definiert ist
- Falls eine Aussage zur Zeitdauer der Studiendauer fehlt, akzeptiert OLUtool bis 15 lebende Probanden in der Verumgruppe (No. at Risk) und setzt hier die Zeitgrenze.

## 3.2 Varianten

Im Alltag der Studienanalyse beobachten wir vier verschiedene Arten wie die Kaplan-Meier-Kurve verlaufen kann. Alle Situationen sind im OLUtool-Onko in den Rubriken 1 – 4 abgebildet.

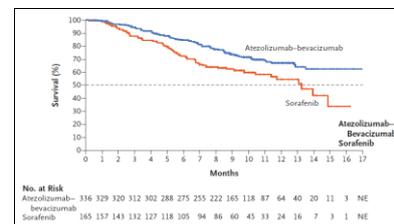
### 1. Standardsituation OS oder PFS

Es existieren mediane OS/PFS-Werte (50%) für das Prüfmedikament und die Kontrollgruppe in der Kaplan-Meier-Kurve. Der absolute Benefit bildet die Differenz ( $\Delta$ ) von mOS oder mPFS. Zusammen mit dem HR-Wert findet man die Score-Punktzahl im Modell.



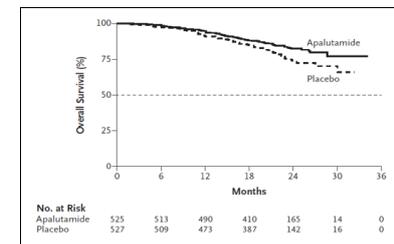
### 2. OS oder PFS-Kurve vom Prüfmedikament verläuft oberhalb der 50%-Marke, Kontrolle unterhalb 50%

Die Differenz  $\Delta$  des absoluten Benefits kann nicht bestimmt werden, weil für das Prüfmedikament der 50%-Wert fehlt. In der Studie findet man die Aussage «not reached» (NR). Die Zeitdauer entspricht der validierten Studiendauer oder der No. at Risk bis 15 lebende Probanden der Prüfsubstanz.



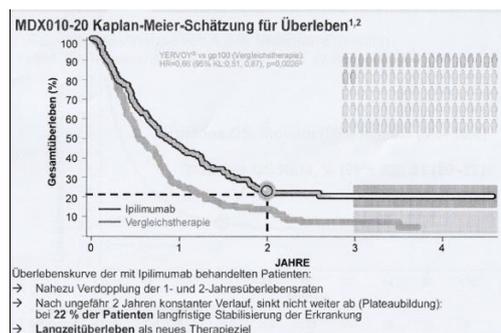
### 3. OS oder PFS-Kurve vom Prüfmedikament und Kontrolle verlaufen oberhalb der 50%-Marke

Diese Situation besagt, dass beide Therapien sehr gut wirken oder die Interimsanalyse so frühzeitig erfolgte, dass noch keine definitiven medianen Resultate verfügbar sind, welche den Wirkungsabfall zeigen. Beide Parameter sind «not reached» (NR). Die Zeitdauer für OLUtool entspricht der validierten Studiendauer oder der No. at Risk bis 15 lebende Probanden der Prüfsubstanz



### 4. Standardsituation, jedoch langes Überleben einer kleinen Gruppe

Die 50%-Marke wird von beiden Gruppen geschnitten und eine  $\Delta$ -Bildung für OS ist möglich (wie Fall 1), jedoch ist auffällig, dass auf tiefem Niveau bei der Gruppe vom Prüfmedikament sehr lange niemand mehr stirbt. Betrachtet man nur das  $\Delta$  des medianen OS, so würde man denken, dass das neue Medikament nur einen geringen Nutzen bringt. Gleichzeitig würde aber verkannt, dass eine kleine Gruppe langfristig profitiert. Diese Situation honorieren wir zu Gunsten der kleinen Patientengruppe mit einem Bonus  $\rightarrow$  Long term plateau in survival curve  $\geq 12$  Monate für den OS-Bereich 10 – 49 %, falls das Plateau waagrecht verläuft. Beispiel Ipilimumab bei Melanom (10).



Wir verlangen das Sockelniveau von 10%, weil aus Erfahrung bekannt ist, dass Einzelfälle aus einer natürlichen Immunitätslage heraus, auch ohne Therapie, lange überleben können. Verlaufen bei beiden Studiengruppen die OS-Kurven parallel und waagrecht auf tiefem Niveau (10 – 49%), verlangen wir eine Differenz im Plateaubereich von  $\geq 10\%$  zu Gunsten des Prüfmedikamentes.

**Zur Erinnerung:**

**Für den Bereich > 50% (not reached) vergeben wir Score-Punkte und deswegen gibt es trotz allfälligem Plateau keinen Bonus.**

### 3.3 LL95%HR versus Punktschätzer-HR

#### LL95% HR durch ESMO validiert

ESMO arbeitet mit dem unteren HR-Wert (= LL) des 95%-Konfidenzintervalls (95%-CI) und nicht mit dem gemittelten HR-Wert (= Punktschätzerwert). Damit erfolgt eine mildere Bewertung verglichen mit dem Punktschätzerwert. Grosse statistische Simulationen von ESMO haben gezeigt, dass mit LL95%HR Verzerrungen in Studien mit ganz grossen oder ganz kleinen Probandenzahlen ausgeglichen werden können.

Die SGV hat zur Prüfung, ob der HR-Punktschätzerwert oder der LL95%HR-Wert verwendet werden soll, eine Beurteilung durch einen neutralen Biostatistiker eingeholt. Das Fazit seiner Untersuchung lautete, dass man beide HR-Werte als Basis für die Nutzenbeurteilung nehmen kann. Es gibt kein «richtig oder falsch». Die Unterschiede seien klein. Tendenziell sei jedoch ein System strenger, falls der Punktschätzerwert als Basis genommen werde.

**Die Expertenaussage hat die SGV bewogen, im OLUtool-Onko den LL95%HR-Wert zu verwenden, falls gleichzeitig ein absoluter Benefit vorliegt. Damit rückt OLUtool-Onko nahe an das validierte ESMO Modell und kann in grossmehrheitlichem Teil als validiert gelten.**

#### Anwendung von LL95%HR

LL95%HR gilt	→ Nur innerhalb Onkologie in der Kombination des relativen (HR) und absoluten Benefits (Wirkdauer von mOS oder mPFS in Monaten).
LL95%HR gilt nicht	→ Falls der 50% Medianwert noch nicht erreicht ist (not reached). Hier verwenden wir den Punktschätzerwert. Ebenso bei Studien mit nur alleinigem HR Wert ohne Angaben zum absoluten Benefit → Bei Studien im adjuvanten Bereich, im ganzen NonOnko Bereich oder bei der Frage der Uebersteuerung eines SL Präparates, weil für all diese Situationen keine aufgearbeitete statistische Datenbasis vorliegt.

Vergleichende SGV eigene Auswertungen von Studienserien haben gezeigt, dass mit der Verwendung des LL95%HR-Wertes gegenüber dem Punktschätzerwert nur in wenigen Fällen eine höhere Nutzenkategorie resultiert, davon meist C anstelle von D.

# 4 Besonderheiten OLUtool-NonOnko

## 4.1 Clinical Benefit

Der NonOnko-Bereich behandelt alle medizinischen Fachdisziplinen ausserhalb der Onkologie. Die Studien sind dementsprechend heterogen. Die Herausforderung für OLUtool bestand darin, übergeordnete Parameter zu finden, damit nicht für jede Fachdisziplin ein eigenes Modell entworfen werden muss. Wir haben die Lösung darin gefunden, dass wir den Einheitsparameter «Clinical Benefit» einführen, der je nach Studiendesign modifiziert wird. International gibt es keine Modellvorbilder wie in der Onkologie. Die Arbeitsgruppe SGV ist mit dem OLUtool-NonOnko effektiv als Pionier unterwegs.

### Die 6 Studientypen

Die Heterogenität der Studien ist zwar gross, doch begegnen uns immer wieder die gleichen 6 Studientypen, so dass die Berechnung des Clinical Benefit an die Grundtypen von NonOnko-Studien angepasst werden kann. Die Herleitung der Bewertung mit Scorepunkten ist dem Kapitel 1.3 zu entnehmen.

#### Studientypen

Typ	1° Endpunkt Studie	Mess-Parameter von 1° Endpunkt	Bestimmung von Clinical Benefit
①	Funktionsänderung	%-Verbesserung ausgehend von Baseline	$\Delta\%$ Verum und Kontrolle → berechnen
②	Anzahl Ereignisse	%-Verbesserung ausgehend von Baseline	$\Delta\%$ Verum und Kontrolle → berechnen
③	Scorepunkte Messung	%-Verbesserung ausgehend von Baseline Total Scorepunkte = 100%	$\Delta\%$ Verum und Kontrolle von total Scorepunkte → berechnen
④	Surrogatwerte mit Studienrange (x - y)	%-Verbesserung ausgehend von Baseline Masszahl Studienrange (x - y) = 100%	$\Delta\%$ Verum und Kontrolle von Studienrange (x - y) → berechnen
⑤	Ansprechrate in %	%-Ansprechrate von 1°Endpunkt	$\Delta\%$ = Prozent Punkte (PP) = ARR → direkt aus Studie zu entnehmen
⑥	Mortalität	Mortalitätssenkung in Jahren oder HR	$\Delta$ direkt aus Studie zu entnehmen oder berechnen

### Beispiele zu den Studientypen und Berechnung von Clinical Benefit

#### ① Funktionsänderung

Kranke Studienpopulation, eingeteilt in die Verumgruppe und in die Kontrollgruppe → Im 6-Minuten-Gehtest erreicht das Kollektiv der Verumgruppe vor Studienbeginn 40m (= Baseline Verum), das Kollektiv der Placebogruppe 44m (= Baseline Kontrolle). Bei Studienende erreicht die Verumgruppe 60m, die Placebogruppe 50m.

$\Delta\%$  Verumgruppe = 20 m Verbesserung gegenüber der Baseline von 40m = + 50%

$\Delta\%$  Placebogruppe = 6 m Verbesserung gegenüber der Baseline von 44m = + 14%

→ Clinical Benefit = 50% - 14% = 36 %

#### ② Anzahl Ereignisse

Kranke Studienpopulation, eingeteilt in die Verumgruppe und in die Kontrollgruppe → Die Verumgruppe erleidet im Durchschnitt alle 3 Monate 10 Epilepsieanfälle (Baseline), die Placebogruppe 11 Anfälle (Baseline).

Bei Studienende → Verum-Medikament reduziert die Anfälle auf 6, Placebo auf 10.

Verumgruppe: Baseline = 10 →  $\Delta$  Verbesserung = 4 (10 → 6) = 40% Reduktion von 10 Anfällen

Placebogruppe: Baseline = 11 →  $\Delta$  Verbesserung = 1 (11 → 10) = 9% Reduktion von 11 Anfällen

→ Clinical Benefit = 40% - 9% = 31%

#### ③ Scorepunkte Messung

Die Funktionalität eines Leidens wird mit dem Score einer Messreihe erfasst. Die Messreihe erstreckt sich von 1 – 60 Punkten. Das Verum-Medikament verbessert die Funktion um 12 Scorepunkte, Placebo um 6.

Bei Messparameter mit 60 Punkten (= 100%) entsprechen 12 Pkte einer Verbesserung von 20%, 6 Pkte von 10%.

Verbesserung in der Verumgruppe = 20 %; Verbesserung in der Placebogruppe = 10 %

→ Clinical Benefit = 20% - 10% = 10 %

#### ④ Surrogatwerte mit Studienrange (x - y)

In einer grossen Studie bei Diabetikern schwankt der Wert von HbA1c zwischen 6,5% und 11%. In der Verumgruppe wird eine Senkung des HbA1c von 1,5% erreicht, in der Placebogruppe von 0,1%.

Der «range» laut Studie von HbA1c =  $\Delta(x - y) = 11 - 6,5 = 4,5$  (HbA1c-%) = 100%. Der «range» der Studie entspricht der klinischen Realität und ist akzeptiert\*.

$\Delta$  der Verumgruppe versus Placebogruppe =  $1,5\% - 0,1\% = 1,4\%$  HbA1c Verbesserung  
 → Clinical Benefit → Berechnung:  $1,4$  vom «range»  $4,5$  (= 100%) →  $1,4$  von  $4,5 = 31\%$

\* Der «Range» zwischen Minimum/Maximum ( $x - y$ ) soll vom VA beurteilt werden und nahe an der Realität sein. Meist kann dieser Wert aus der Studie entnommen und akzeptiert werden. Falls nicht, gelten die anerkannten Normwerte.

### ⑤ Ansprechrate in Prozent

33% weniger Bluttransfusionen (= 1°) gilt als klinisch relevant bei Patienten mit transfusionsabhängiger Thalassämie. Eine Studie zeigt, dass mit der Medikation A bei 21% der Probanden dieses Ziel erreicht wurde (= 1° Endpunkt), jedoch nur bei 4 % mit der Medikation B (Placebo).  $\Delta$  Ansprechrate von Verum versus Placebo =  $21\% - 4\% = 17\%$

→ Clinical Benefit = 17% (ARR) → NNT =  $1 : 17\% = 5,8$  → jeder 6. Patient profitiert vom Benefit der Transfusionsreduktion  
 → Zusätzlich 1x Bonus, weil ARR < 20% und die Risikominderung = 425%

#### Bonusberechnung

Am einfachsten ist die Berechnung des relativen Benefits, wenn der Benefit ( $\Delta$ ) dividiert wird durch die Ansprechrate der Kontrollgruppe.

→ 17% dividiert durch 4% = 4,25 → Bedeutet, dass 17% ist 4,25 Mal grösser als 4% = + 425% Verbesserung

### ⑥ Mortalität

Dank einer Medikation A bei fortgeschrittener Herzinsuffizienz sterben innert 24 Monaten 29% versus 42% der Probanden in der Placebogruppe. Diese Reduktion der Mortalität entspricht einer Minderung mit HR 0,70.

→ Clinical Benefit = HR 0,70

In einer andern Studie mit einer Beobachtungszeit von 2,4 Jahren ergibt sich mit Medikament A eine Mortalität von 6,8%, mit Medikament B von 4,7% →  $\Delta = 2,1\%$  = Verminderung der Mortalität innert 2,4 Jahren.

→ Clinical Benefit = 0,87% pro 1 Jahr

## 4.2 NNT als Bewertungskriterium

Noch bis Ende 20. Jahrhundert galten alleinige Änderungen des relativen Risikos (Hazard Ratio) als gutes Beurteilungsmass für Studienergebnisse. Eine realistischere Einschätzung ergibt jedoch die Änderung des absoluten Risikos. Diese Betrachtung von relativer und absoluter Änderung wurde im ESMO Modell aufgenommen.

In OLUtool Onko haben wir diese Betrachtung weitgehend übernommen. Im NonOnko Bereich haben wir in der grossen Mehrheit Studiendesigns bei denen das Ansprechen hinsichtlich Krankheitsverbesserung gemessen wird. Im Gegensatz zur Onkologie steht nicht das Kriterium «Leben oder Tod» im Vordergrund. Hier bietet sich die NNT als klinisches Mass zur Bewertung des Response an. Die NNT selbst basiert auf der Absoluten Risiko Reduktion (ARR), die wir als Clinical Benefit bestimmen (9).

OLUtool Onko (analog ESMO)	OLUtool NonOnko Typ ⑤
Relativer Benefit als Hazard Ratio (HR) u n d Absoluter Benefit als Veränderung in Monaten	Absoluter Benefit als NNT u n d Relativer Benefit als Hazard Ratio (HR) mit Bonus

### Risiko und Risikoreduktion

In einer randomisierten klinischen Studie werden 200 Patienten mit Verum (100) oder Placebo (100) über fünf Jahre behandelt. In der Placebogruppe erleiden 30 von 100 und in der Verum-Gruppe 10 von 100 Personen ein definiertes Ereignis. Daraus ergeben sich folgende Berechnungen zur relativen und absoluten Risikobetrachtung:

<b>Absolutes Risiko</b> → Patienten mit Ereignis (Verumgruppe)	10 von 100 = 10%
→ Patienten mit Ereignis (Placebogruppe)	30 von 100 = 30%
<b>Absolute Risikoreduktion (ARR)</b> → Differenz absolutes Risiko von Verum- versus Placebogruppe = $30\% - 10\% = 20\%$	
→ Das Ereignisrisiko über 5 Jahre ist in der Verumgruppe um 20 Prozente geringer als in der Placebogruppe	
<b>Number needed to treat (NNT)</b> → $1 : \text{Absolute Risikoreduktion} = \text{NNT} \rightarrow 1 : 20\% = 5$	
<b>→ Fünf Personen müssen über fünf Jahre das Verummittel einnehmen um 1 Ereignis zu verhindern</b>	

**Relatives Risiko** → Absolutes Risiko (Verum) : Absolutes Risiko (Plazebo) = 10% : 30% =  $\frac{1}{3}$  = 0.33

**Relative Risikoreduktion** → Differenz der Risiken = 1 - 0.33 = 0.67 = 67%

Relativ → Das Ereignisrisiko über 5 Jahre ist in der Verumgruppe um 67% geringer als in der Plazebogruppe

Absolut → Das Ereignisrisiko über 5 Jahre ist in der Verumgruppe um 20 Prozent geringer als in der Plazebogruppe

#### Praktisches Beispiel

Patienten, die mit einem  $\beta$ -Blocker behandelt wurden, erlitten in 12,8% einen Herzinfarkt. Bei Patienten, die mit einem AT-II-Antagonisten therapiert wurden, jedoch nur in 11% der Fälle.  $\Delta = 1,8\%$ . Die relative Risikoreduktion RRR berechnet sich mit  $1,8 : 12,8 = 0,14 = 14\%$ . Die absolute Risikoreduktion ARR liegt jedoch nur bei  $12,8\% - 11\% = 1,8\%$ .

Die NNT ist ein eindrückliches Mass, wie ein Medikament klinisch wirkt. Die Berechnung erfolgt mit  $1 : ARR = NNT$ . Im obigen «praktischen Beispiel» beträgt die NNT somit  $1 : 1,8\% = 55$  → Nur jeder 55. Patient kann vom Medikament profitieren (11).

## Welche NNT entspricht dem grossen Nutzen?

Ausgehend, dass ein Drittel der Patienten profitieren sollte, ergibt sich eine NNT \* von 3. Eine medikamentöse Behandlung gilt als guter Erfolg, falls eine NNT von 3 – 4 resultiert, was unserem postulierten Drittel entspricht. Dies im Gegensatz zu Screenings, wo eine NNT von 20 noch als Erfolg gilt (6).

In der Literatur (6) (11) gibt es einige Einschränkungen, wann die NNT als Beurteilungskriterium verwendbar ist:

- Es handelt sich um statistisch signifikante Studienergebnisse
- Die NNT gilt für den gleichen Zeitraum von Verum und Kontrollgruppe
- Gleiche Diagnostik von Verum- und Kontrollgruppe
- Das Basisrisiko (Morbidity) von Verum/Kontrolle und vom prüfaren Einzelfall sind gleich
- Keine Metaanalysen, da bei Metaanalysen das Basisrisiko in den Einzelstudien oft variiert
- Das relative Risiko wird berücksichtigt (siehe Bonus)

→ All diese Voraussetzungen sind in OLUtool erfüllt und umgesetzt

Je seltener und je schwerwiegender die Krankheitssituation ist, desto höher darf die NNT sein (6). Im Art. 71a-d KVV haben wir per definitionem immer schwerwiegende Krankheiten. Wir meinen deswegen, dass eine NNT von 3 (Drittelsregel) zu streng wäre. Wir erachten eine NNT von 5 (= 20%) als angemessen und geben dafür 3 Scorepunkte = B, sowie zusätzlich die Möglichkeit für 1x Bonus, wenn  $ARR < 20\%$  ( $NNT > 5$ ) und der relative Benefit gross ist ( $\geq 65\%$ ).

\* NNT wird immer als gerundete ganze Zahl angegeben

# 5 Vorgehen beim Rating

## Die 5 Schritte

Wie soll der VA an eine Studie herangehen, die er raten will? Diese Frage ist wichtig, weil in den Studien oft ganz viele Kriterien vorkommen, welche nicht immer direkt in OLUtool abgebildet sind. Um rasch ans Ziel zu gelangen empfiehlt sich ein Vorgehen in 5 Schritten:

	<b>OLUtool-Onko</b>	<b>OLUtool-NonOnko</b>
1	Worum geht es und was für eine Studie liegt vor?	Worum geht es und was für eine Studie liegt vor?
2	1° Endpunkt?	1° Endpunkt?
3	<b>Welche Rubrik kommt zur Anwendung?</b>	<b>Welcher Studientyp kommt vor? → Clinical Benefit</b>
4	Bonus / Malus ?	Bonus / Malus ?
5	Fazit VA	Fazit VA

## Besonderes

### Odds Ratio (OR) versus Hazard Ratio (HR)

Wir betrachten beim Rating die Studien etwas röhrenförmig und legen den Fokus auf die OLUtool Kriterien. Dies im Wissen, dass die international anerkannten und relevanten Nutzenkriterien abgebildet sind. Eine Schwierigkeit kann auftreten, wenn eine Studie ausschliesslich Odds Ratio anstelle von Hazard Ratio verwendet.

Die beiden Parameter verhalten sich etwas vereinfacht gesagt, spiegelbildlich:

- Bei der HR geht es um Risikominderung → Je kleiner HR, desto grösser die Wahrscheinlichkeit, dass der unerwünschte Effekt reduziert wird.
- Bei der OR geht es um Chancenerhöhung → Je grösser OR, desto grösser die Chance, dass der erwünschte Effekt eintritt.

Bei beiden ist 1,0 der Scheidepunkt (Spiegelpunkt). Je kleiner die HR oder je grösser die OR, desto wirksamer ist die Therapie. Bei HR ist das signifikante CI-Intervall  $< 1$ , bei OR  $> 1$ .

**Bis anhin konnten wir alle Studien im Art. 71a-d KVV mit den Kriterien von OLUtool bearbeiten, auch wenn Angaben zur OR in einer Studie vorkamen. Wir haben jedoch zurzeit keinen entsprechenden Bewertungs-Massstab für die OR. Sollten in Zukunft Studien nur mit OR vorkommen, müsste OLUtool angepasst werden.**

## 6 Rechtsprechung

Die Gerichtspraxis verlangt vom VA, dass er die Wissenschaftlichkeit einer Studie und den Einzelfall prüft. Die Wissenschaftlichkeit muss nicht im höchsten Ausmass erfüllt sein. Interimsanalysen oder kleine Kohortenstudien genügen, wenn Konsens zwischen Forschern und Praktikern besteht (Experten). Der Einzelfall muss hingegen gut zur Studie passen. Ein SL Medikament darf übersteuert werden, falls das Nutzen/Risiko Verhältnis deutlich besser ist. Ein individueller Heilversuch (Probetherapie) ist als ultima ratio zulässig.

Die Urteile vor 2011, also vor der Verordnung von Art. 71a-d KVV, behalten ihre Gültigkeit und Aktualität, da bereits damals Medikationen ausserhalb der SL angewendet wurden.

Thema	Aussage Gericht	Urteil
<b>Individueller Heilversuch</b>	Zulässig im Einzelfall als ultima ratio. Mindestens eine kleine Kohortenstudie müsste als wissenschaftliche Basis vorliegen	BGE 134 V 175 (2008)
<b>Fehlende Behandlungsalternative = Übersteuerung SL</b>	OLU-Medikation verspricht ein besseres Nutzen-Risiko-Verhältnis. Der Unterschied soll hoch sein. Zu berücksichtigen seien Wirksamkeit, Verträglichkeit und Wirtschaftlichkeit	BGE 130 V 532 E 5.41 und 6.2 (2006)
<b>Interimsanalysen</b>	Setzt voraus, dass der grosse Nutzen bereits sichtbar wird	BGE 136 V 395 E 6.3 (2011)
<b>Grosser Nutzen von OS/PFS</b>	Die Kriterien OS und PFS in RCT sind geeignete Kriterien, um den grossen Nutzen zu definieren	BGE 130 V 532 (2006)
<b>Ohne Studiendaten kein grosser Nutzen</b>	Fehlen wissenschaftliche Daten, kann kein Wirkungsnachweis begründet werden. Die Wirkung einer Therapie im Einzelfall führt nicht zur Pflichtleistung (post hoc ergo propter hoc)	BGE 136 V 395 (2011)
<b>Anerkannte Publikation</b>	Veröffentlichte Erkenntnisse ausserhalb von RCT genügen, falls Aussagen zur Wirksamkeit vorliegen und Konsens in Fachkreisen besteht, dass ein grosser Nutzen besteht	BG 9C_550/2011/12
<b>Beurteilung des grossen Nutzens</b>	Beurteilung nötig in allgemeiner Weise wie auch bezogen auf den konkreten Einzelfall. Der Nachweis muss nach wissenschaftlichen Methoden erbracht werden.	BG 9C_667/2015
<b>Wissenschaftlichkeit</b>	Ist auch gegeben, wenn eine Behandlungsart den Erkenntnissen der medizinischen Wissenschaft entspricht und von Forschern und Praktikern auf breiter Basis anerkannt ist	BG 8C_523/2016
<b>Studienpopulation/Einzelfall</b>	Der Einzelfall darf nicht wesentlich von der Studienpopulation abweichen. Eine Probetherapie wäre dann nicht zulässig	BG 9C_390/2020

**OLUtool erfüllt in allen Punkten die bisherige Gerichtspraxis. Einzig das Upgrading von D zu C aufgrund der Experten-Empfehlung von Tumorboard oder Referenzzentrum ist rechtliches Neuland.**

## 7 Literatur

- (1) R. Fries, Y. Hummel, B. Kipfer, U. Schafroth, B. Seiler, J. Zollikofer «OLUtool als neues Einheitsmodell der Vertrauensärzte» Schweizerische Ärztezeitung 2018;99(45):1571-1572
- (2) R. Fries, B. Kipfer, U. Schafroth, B. Seiler, J. Zollikofer, «Die SGV stellt die Nutzenbewertung auf eine neue Basis», Schweizerische Ärztezeitung 2018;99(23):749–50
- (3) N.I. Cherny, de Vries, et al. *Anal of Oncol* 2017, ESMO-Magnitude of Clinical Benefit Scale Version 1.1.
- (4) Schweizerische Bundesgerichts Urteil vom 23.11.2010, 9C\_334/2010
- (5) Sozialversicherungsgericht Kanton Zürich: KV.2018.00101 und BG 9C\_390/2020 vom 17.07.2020
- (6) Gysling E., «Epidemiologie für die Praxis», pharma-kritik-Jahrgang 22 , Nummer 08/2000
- (7) George A. et al, «Was ist das Risiko: Differenzierung von Risikokennzahlen, Odds Ratios und Hazard Ratios?», *Cureus*. 2020 August; 12(8): e10047, Online veröffentlicht 2020 Aug 26
- (8) Ocana and Tannock, «When are positive clinical trials in oncology truly positive?» *J Nat Cancer Inst* 2011, 103:16-20
- (9) Harris M. et al, «Mathematik und Statistik», Spektrum Akademischer Verlag 2007, ISBN 978-3-8274-1829-6
- (10) Hodi F. et al, *NEJ Med* 2010;363(8):711-23
- (11) Anlauf M., «Beurteilung von Therapien mit der number needed to treat», *Deutsches Ärzteblatt*, Jahrgang 103, Heft 48, 01. Dezember 2006
- (12) Vorhees Peter M. et al, «Daratumumab, lenalidomide, bortezomib, and dexamethasone for transplant-eligible newly diagnosed multiple myeloma: the GRIFFIN trial», *blood*® 20 Aug. 2020, Vol. 136, No 8

# 8 Anhang

## A1 Übersteuerung eines SL Präparates

### Allgemeines

Damit ein Arzneimittel im Art. 71a-d KVV leistungspflichtig werden kann, müssen folgende Voraussetzungen kumulativ erfüllt sein:

- Behördliche Zulassung durch swissmedic oder durch eine gleichwertige ausländische Zulassungsbehörde. In der Praxis sind dies EMA (Europa) und FDA (USA)
- Es handelt sich um ein Leiden, welches tödlich verlaufen oder zu schweren und chronischen gesundheitlichen Beeinträchtigungen führen kann
- Eine andere wirksame und zugelassene Behandlungsmethode fehlt. Bei Arzneimittel → SL Präparat fehlt
- Vom Arzneimittel wird ein grosser therapeutischer Nutzen erwartet

Diese Kriterien haben immer wieder zu Diskussionen geführt, weil sie mit Ausnahme der Zulassung nicht klar geregelt sind. Am meisten Schwierigkeiten verursacht die Prüfung nach der Behandlungsalternative, falls es sich um ein Arzneimittel handelt.

### Herausforderung für den VA

Von der Ärzteschaft wird ein Arzneimittel zur Kostenübernahme im Art. 71a-d KVV beantragt, obwohl ein SL Präparat für die gleiche Indikation gelistet ist.

### SL Medikamente sind WZW geprüft und lassen einen grossen Nutzen erwarten

Die Gerichtspraxis zeigt, dass trotz SL-Alternative ein Medikament im Art. 71a-d KVV akzeptiert werden kann, falls das Nutzen-Risiko-Verhältnis im Vergleich zum SL-Präparat hoch ist. Allerdings gibt es von den Gerichten keine quantifizierten Angaben, um wieviel höher das Nutzen-Risiko-Verhältnis des beantragten Art. 71 - Präparates sein muss, damit es gerechtfertigt ist, das SL - Präparat zu übersteuern, welches ja bereits einen grossen Nutzen erwarten lässt (weil es in die SL aufgenommen wurde). Die SGV versuchte, eine Quantifizierung zu formulieren. Ausgangspunkt ist wiederum die «Drittelsregel».

### Besonderes

Man könnte zum Schluss kommen, dass es genügt, ein beantragtes Medikament einfach direkt mit OLUtool auf den grossen Nutzen hin zu prüfen. Falls der grosse Nutzen ausgewiesen ist (A oder B) könnte Art. 71a-d KVV akzeptiert werden. Das wäre aber zu kurz gegriffen. Dazu folgende Überlegung: Ein neues Arzneimittel kann in OLUtool Kategorie C erreichen und dennoch 35% besser sein, weil das zugelassene Medikament mit wenig Nutzen in die SL aufgenommen wurde. Die SL-Aufnahme basiert nicht auf einem Nutzenmodell analog OLUtool, sondern folgt auch gesundheitspolitischen Überlegungen. In der Direktbestimmung mit OLUtool würden wir eine Therapieoption mit Kategorie C verweigern, da nur A oder B akzeptiert wäre.

→ Eine Übersteuerung ist möglich, wenn die Überlegenheit  $\geq 35\%$  beträgt (= höheres Nutzen-Risiko-Verhältnis)

### Es begegnen uns zwei Varianten von Situationen:

Variante 1 : Die Studie vergleicht das neue Medikament gerade mit dem SL Präparat.

→ Übersteuerung JA, falls die Kategorie A oder B in OLUtool erreicht wird

Variante 2 : Die Studie vergleicht das neue Medikament nicht mit dem SL Präparat

→ Die Daten für das SL Präparat sind aus der Zulassungsstudie (swissmedic) zu erheben. Der Vergleich mit der neuen Medikation ergibt den Benefit ( $\Delta$ ) und damit die Nutzenkategorie nach OLUtool (A – D).

**Die direkte Nutzenbestimmung mit OLUtool klärt nicht in jedem Fall, ob eine Übersteuerung und damit Art. 71a-d KVV gerechtfertigt ist. Es braucht Regeln zur Übersteuerung.**

# Regeln zur Übersteuerung

(Gekürzte Fassung, Original siehe Homepage SGV)

## A Keine andere wirksame und zugelassene Behandlungsmethode (Medikationen)

Bereich	Grundsätzliche Kriterien zur Überlegenheit einer neuen medikamentösen Therapie → Studiendaten sind üblicherweise in peer-reviewten Fachzeitschriften publiziert → Klinisch relevante sAE oder QoL sind vergleichbar oder besser
Onkologie	<ul style="list-style-type: none"> <li>▪ HR (OS/PFS) <math>\leq 0,65 + \Delta mOS/mPFS \geq 3</math> Monate <span style="float: right;"><u>oder</u></span></li> <li>▪ Falls kein HR zum SL-Präparat: <math>mOS/mPFS \geq 35\% + \Delta mOS/mPFS \geq 3</math> Monate <span style="float: right;"><u>oder</u></span></li> <li>▪ Falls <math>mOS/mPFS</math> not reached mit HR <math>\leq 0,65</math> während <math>\geq 8</math> Monaten <span style="float: right;"><u>oder</u></span></li> <li>▪ Vergleichbare Wirkdaten, jedoch QoL mit validierter Erhebung <math>\geq 35\%</math> besser</li> <li>▪ Adjunkte Therapie               <ul style="list-style-type: none"> <li>○ Gesamt-Überleben (OS) nach Studienende <math>\geq 35\%</math> besser <span style="float: right;"><u>oder</u></span></li> <li>○ HR(DFS) <math>\leq 0,65</math> während <math>\geq 12</math> Monaten</li> </ul> </li> </ul>
Ausserhalb Onkologie	<ul style="list-style-type: none"> <li>▪ Erwartbarer klinisch relevanter Effekt <math>\geq 35\%</math></li> </ul>

mOS/mPFS = medianes OS/PFS (50%)

HR = Hazard Ratio (gemittelter Punktschätzer-Wert)

sAE = serious Adverse Events (Onko: Grad 3+4)

not reached = OS/PFS liegen über 50% der Kaplan-Meier-Kurve → mediane Marke von 50% wird nicht erreicht

ORR = Overall Response Rate (Complete + Partial Response)

QoL = Quality of Life

DFS = Disease Free Survival

### Besonderes

→ Wir verwenden den Punktschätzerwert, weil bezüglich Übersteuerung keine Aussagen vorliegen, dass der LL95%HR präziser wäre.

## B Tödlicher Verlauf möglich

Der Mangel bei diesem Passus besteht darin, dass der Zeithorizont fehlt. Unbestritten sind Krebsleiden, welche kurz- und mittelfristig zum Tod führen. Unklar wird es bei langen Krankheitsverläufen, oft im NonOnko-Bereich, wie z.B. bei Herzinsuffizienz, Diabetes, Leberzirrhose oder bei chronischen Nierenerkrankungen. Es braucht Mortalitätsdaten sowie Daten zur klinischen Patientensituation. Eine Rücksprache mit dem behandelnden Arzt kann hilfreich sein.

→ Mit der Medikation soll die durchschnittliche Lebenserwartung oder die Lebensqualität der individuellen eines Patienten relevant gesteigert werden.

## C Schwere und chronische gesundheitliche Beeinträchtigung

Laut Art. 71a-d KVV wird erwartet, dass ohne Therapie mit der beantragten Medikation die Entwicklung einer schweren und chronischen gesundheitlichen Beeinträchtigung erfolgt.

Zu beachten ist, dass eine schwere und chronische Beeinträchtigung im Fokus steht und nicht eine chronische Krankheit. Falls tatsächlich eine chronische Krankheit vermieden werden könnte, wäre Art. 71a-d KVV erfüllt. Der generelle Begriff «Beeinträchtigung» weist vor allem auf Teilaspekte von chronischen Krankheiten hin. So kann ein Funktionsdefizit als Teilaspekt eines Leidens im Art. 71a-d KVV stehen wie z.B. der Erhalt der Gehfähigkeit bei einem neurologischen Leiden.

**Folgende Kriterien sprechen für eine schwere und chronische gesundheitliche Beeinträchtigung, die eine Medikation im Art. 71a-d KVV rechtfertigen können:**

<p><b>Unter anderem droht eine schwere und chronische gesundheitliche Beeinträchtigung bei:</b></p> <ul style="list-style-type: none"> <li>➤ Organschädigung, welche mit der Medikation verhindert wird</li> <li>➤ Kontinuierlicher Steigerung eines potenziell schädlichen Arzneimittelkonsums, welche mit der Medikation verhindert wird</li> <li>➤ Pflegebedürftigkeit oder relevanter Einschränkung in den ADL (<i>Aktivitäten des täglichen Lebens</i>), welche mit der Medikation verhindert werden</li> <li>➤ Dauerhafter Verschlechterung der Lebensqualität (QoL), welche mit der Medikation verhindert wird</li> <li>➤ Minderung der Erwerbsfähigkeit (Invalidität), welche mit der Medikation abgewendet werden kann</li> </ul>
<p><b>Der Umkehrschluss kann ebenso zu Art. 71a-d KVV qualifizieren:</b></p> <ul style="list-style-type: none"> <li>○ Mit der Medikation kann erwartet werden, dass eine Organschädigung, eine Pflegebedürftigkeit, eine ADL-Qualität, die QoL oder die Erwerbsfähigkeit relevant verbessert wird</li> <li>○ Es kann erwartet werden, dass dank der Medikation ein potenziell schädlicher Arzneimittelkonsum relevant vermindert werden kann</li> </ul>

## A2 Art. 71a-d KVV, BAG 2021

### Vergütung von Arzneimitteln im Einzelfall, BAG Oktober 2021

#### Art. 71a Übernahme der Kosten eines Arzneimittels der Spezialitätenliste ausserhalb der genehmigten Fachinformation oder Limitierung

<sup>1</sup> Die obligatorische Krankenpflegeversicherung übernimmt die Kosten eines in die Spezialitätenliste aufgenommenen Arzneimittels für eine Anwendung ausserhalb der vom Institut genehmigten Fachinformation oder ausserhalb der in der Spezialitätenliste festgelegten Limitierung nach Artikel 73, wenn:

- a. der Einsatz des Arzneimittels eine unerlässliche Voraussetzung für die Durchführung einer anderen von der obligatorischen Krankenpflegeversicherung übernommenen Leistung bildet und diese eindeutig im Vordergrund steht; oder
- b. vom Einsatz des Arzneimittels ein grosser therapeutischer Nutzen gegen eine Krankheit erwartet wird, die für die versicherte Person tödlich verlaufen oder schwere und chronische gesundheitliche Beeinträchtigungen nach sich ziehen kann, und wegen fehlender therapeutischer Alternativen keine andere wirksame und zugelassene Behandlungsmethode verfügbar ist.

<sup>2</sup> Der Versicherer bestimmt nach Absprache mit der ZulassungsinhaberIn die Höhe der Vergütung. Der zu vergütende Preis muss unter dem Höchstpreis der Spezialitätenliste liegen.<sup>2</sup>

#### Art. 71b Übernahme der Kosten eines vom Institut zugelassenen nicht in die Spezialitätenliste aufgenommenen Arzneimittels

<sup>1</sup> Die obligatorische Krankenpflegeversicherung übernimmt die Kosten eines vom Institut zugelassenen verwendungsfertigen Arzneimittels, das nicht in die Spezialitätenliste aufgenommen ist, für eine Anwendung innerhalb oder ausserhalb der Fachinformation, wenn die Voraussetzungen nach Artikel 71a Absatz 1 Buchstabe a oder b erfüllt sind.

<sup>2</sup> Der Versicherer bestimmt nach Absprache mit der ZulassungsinhaberIn die Höhe der Vergütung.

#### Art. 71c Übernahme der Kosten eines vom Institut nicht zugelassenen importierten Arzneimittels

<sup>1</sup> Die obligatorische Krankenpflegeversicherung übernimmt die Kosten eines vom Institut nicht zugelassenen verwendungsfertigen Arzneimittels, das nach dem Heilmittelgesetz eingeführt werden darf, sofern die Voraussetzungen nach Artikel 71a Absatz 1 Buchstabe a oder b erfüllt sind und das Arzneimittel von einem Land mit einem vom Institut als gleichwertig anerkannten Zulassungssystem für die entsprechende Indikation zugelassen ist.

<sup>2</sup> Der Versicherer vergütet die Kosten, zu denen das Arzneimittel aus dem Ausland importiert wird. Der Leistungserbringer achtet bei der Auswahl des Landes, aus dem er das Arzneimittel importiert, auf die Kosten.

#### Art. 71d Gemeinsame Bestimmungen

<sup>1</sup> Die obligatorische Krankenpflegeversicherung übernimmt die Kosten des Arzneimittels nur auf besondere Gutsprache des Versicherers nach vorgängiger Konsultation des Vertrauensarztes oder der Vertrauensärztin.

<sup>2</sup> Der Versicherer überprüft, ob die von der obligatorischen Krankenpflegeversicherung übernommenen Kosten in einem angemessenen Verhältnis zum therapeutischen Nutzen stehen.

<sup>3</sup> Ist das Gesuch um Kostengutsprache vollständig, so entscheidet der Versicherer innert zwei Wochen darüber.

<sup>4</sup> Der Leistungserbringer stellt dem Versicherer die effektiven Kosten in Rechnung. Bei Arzneimitteln nach Artikel 71a wird der Höchstpreis der Spezialitätenliste in Rechnung gestellt, bei Arzneimitteln nach den Artikeln 71b und 71c der Preis, zu dem das Arzneimittel vom Leistungserbringer bezogen wurde, zuzüglich des Vertriebsanteils nach Artikel 67 Absatz

<sup>1</sup>quater und der Mehrwertsteuer.

# A3 OLUtool Modelle, Version 3.4 (Update 2022)

A Studienrating							
Rubrik 1	OS als 1 <sup>o</sup> Studienendpunkt	Rubrik 2	OS als 1 <sup>o</sup> Studienendpunkt	Rubrik 3	OS als 1 <sup>o</sup> Studienendpunkt	Rubrik 4	PFS als 1 <sup>o</sup> Studienendpunkt
	Medianes Überleben < 1 Jahr unter Standardtherapie oder Best Care	Medianes Überleben 1–2 Jahre unter Standardtherapie oder Best Care	Medianes Überleben > 2 Jahre unter Standardtherapie oder Best Care			Studie mit Kontrollgruppe	
4	ΔOS nach 2 Jahren ≥ 10 % LL95%HR(OS) ≤ 0,65 + ΔmOS > 3Mt. oder mOS NR ≥ 6 Mt. + HR(OS) ≤ 0,70	ΔOS nach 3 Jahren ≥ 10 % LL95%HR(OS) ≤ 0,70 + ΔmOS > 5 Mt. oder mOS NR ≥ 18 Mt. + HR(OS) ≤ 0,70	ΔOS nach 5 Jahren ≥ 10 % LL95%HR(OS) ≤ 0,70 + ΔmOS > 9 Mt. oder mOS NR ≥ 24 Mt. + HR(OS) ≤ 0,70				
3	LL95%HR(OS) ≤ 0,65 + ΔmOS > 2 Mt. oder mOS NR ≥ 6 Mt. + HR(OS) ≤ 0,70	LL95%HR(OS) ≤ 0,70 + ΔmOS > 4 Mt. oder mOS NR ≥ 12 Mt. + HR(OS) ≤ 0,75	LL95%HR(OS) ≤ 0,70 + ΔmOS > 6 Mt. oder mOS NR ≥ 18 Mt. + HR(OS) ≤ 0,75				LL95%HR(PFS) ≤ 0,65 und ΔmPFS ≥ 3 Mt. oder mPFS NR ≥ 8 Mt. + HR(PFS) ≤ 0,70
2	LL95%HR(OS) ≤ 0,70 + ΔmOS ≥ 2 Mt. oder mOS NR ≥ 6 Mt. + HR(OS) ≤ 0,75	LL95%HR(OS) ≤ 0,75 + ΔmOS ≥ 3 Mt. oder mOS NR ≥ 12 Mt. + HR(OS) ≤ 0,80	LL95%HR(OS) ≤ 0,75 + ΔmOS ≥ 4 Mt. oder mOS NR ≥ 18 Mt. + HR(OS) ≤ 0,80				LL95%HR(PFS) ≤ 0,65 und ΔmPFS ≥ 2 Mt. oder mPFS NR ≥ 6 Mt. + HR(PFS) ≤ 0,70
1	LL95%HR(OS) ≤ 0,75 + ΔmOS ≥ 2 Mt. oder mOS NR ≥ 6 Mt. + HR(OS) ≤ 0,80	LL95%HR(OS) ≤ 0,75 + ΔmOS ≥ 3 Mt. oder mOS NR ≥ 12 Mt. + HR(OS) ≤ 0,80	LL95%HR(OS) ≤ 0,75 + ΔmOS ≥ 4 Mt. oder mOS NR ≥ 18 Mt. + HR(OS) ≤ 0,80				HR(PFS) ≤ 0,75
0	LL95% > 0,75/HR(OS) > 0,80/ΔmOS < 2 Mt.	LL95% > 0,75/HR(OS) > 0,80/ΔmOS < 3 Mt.	LL95% > 0,75/HR(OS) > 0,80/ΔmOS < 4 Mt.				HR(PFS) > 0,75 oder ΔmPFS < 2 Mt.

LL95%HR(OS) = Hazard Ratio von Overall Survival, ausgehend vom unteren HR-Grenzwert des 95%-Konfidenzintervalls (LL = Lower Limit)  
mOS oder mPFS = Median Overall Survival oder Median Progression Free Survival (±50%)  
ΔmOS oder ΔmPFS = Differenz (Δ) der Medianwerte

Rubrik 5	1 <sup>o</sup> Studienendpunkt ist PFS oder ORR, beide ohne Kontrollgruppe Spezialfall: ORR mit Kontrollgruppe → falls ΔORR ≥ 20% gilt ORR absolut	Rubrik 6	1 <sup>o</sup> Studienendpunkt ist OS oder DFS (iDFS) oder pCR Adjuvante und Neoadjuvante Therapie
2	Progressionsfreiheit: mPFS ≥ 6 Mt. oder mPFS NR ≥ 6 Mt. Ansprechrate: ORR ≥ 60% oder ORR 30-60% und Duration of Response (DoR) ≥ 9 Mt.	4	ΔOS nach ≥ 3 Jahren ≥ 5% oder HR(DFS oder iDFS) ≤ 0,65 ΔOS = OS gemessen am Studienende DFS = Disease-free-survival
1	Progressionsfreiheit: mPFS ≥ 3 Mt. oder mPFS NR ≥ 4 Mt. Ansprechrate: ORR 30-60% oder ORR 20-60% und DoR ≥ 4 Mt.	3	ΔOS nach < 3 Jahren ≥ 3% oder HR(DFS oder iDFS) ≤ 0,80 ΔOS nach < 3 Jahren ≥ 3%, falls Lebenserwartung ≤ 2 Jahre Neoadjuvant: Anteil pCR ≥ 50% (ohne Δ)
0	Progressionsfreiheit/Ansprechrate: mPFS < 3 Mt./mPFS NR < 4 Mt. oder ORR < 20%	0	Adjutant: ΔOS < 3% oder HR(DFS/iDFS) > 0,80 oder Neoadjuvant: pCR < 50%

Rubrik 5	1 <sup>o</sup> Studienendpunkt ist PFS oder ORR, beide ohne Kontrollgruppe Spezialfall: ORR mit Kontrollgruppe → falls ΔORR ≥ 20% gilt ORR absolut	Rubrik 6	1 <sup>o</sup> Studienendpunkt ist OS oder DFS (iDFS) oder pCR Adjuvante und Neoadjuvante Therapie
2	Progressionsfreiheit: mPFS ≥ 6 Mt. oder mPFS NR ≥ 6 Mt. Ansprechrate: ORR ≥ 60% oder ORR 30-60% und Duration of Response (DoR) ≥ 9 Mt.	4	ΔOS nach ≥ 3 Jahren ≥ 5% oder HR(DFS oder iDFS) ≤ 0,65 ΔOS = OS gemessen am Studienende DFS = Disease-free-survival
1	Progressionsfreiheit: mPFS ≥ 3 Mt. oder mPFS NR ≥ 4 Mt. Ansprechrate: ORR 30-60% oder ORR 20-60% und DoR ≥ 4 Mt.	3	ΔOS nach < 3 Jahren ≥ 3% oder HR(DFS oder iDFS) ≤ 0,80 ΔOS nach < 3 Jahren ≥ 3%, falls Lebenserwartung ≤ 2 Jahre Neoadjuvant: Anteil pCR ≥ 50% (ohne Δ)
0	Progressionsfreiheit/Ansprechrate: mPFS < 3 Mt./mPFS NR < 4 Mt. oder ORR < 20%	0	Adjutant: ΔOS < 3% oder HR(DFS/iDFS) > 0,80 oder Neoadjuvant: pCR < 50%

Bonus/Malus Punkte	Adverse Events (AE)	Grad 3 + 4 bei Studien mit Kontrollgruppe ΔAE ≥ 30% oder ohne Kontrollgruppe AE ≥ 50%	- 1
→ In der Summe maximal 1 Punkt → Bonus nur möglich, falls mindestens 1 Scorepunkt erreicht worden ist	Long term plateau in survival curve (OS)	OS als 1 <sup>o</sup> Studienendpunkt und OS Plateau im Bereich 10–49% → Plateau für ≥ 12 Mt.	+1
	ORR (= PR + CR)	ORR Prüfmedikament < 30%. Gilt nicht für Rubrik 5	-1
	Studiendefizite siehe Liste	Wird ein Studiendefizit bewertet, müssen nutzenrelevante Folgen begründet sein	-1
	QoL	Definierter Studiendefizit mit validiertem Fragebogen erhoben und mit signifikanter Aussage. Kein Bonus, falls Ergebnis «non inferior»	+1
	Expertenbeurteilung	Internationale Fachgremien wie IQWiG/G-BA, NICE und andere sowie CH-Konsensus Empfehlungen. Der VA dokumentiert die Expertenbeurteilung	- 1 / +1
	Besonderes im Bereich Hämatologie	Anteil «MRD Status negativ» oder sCR nach Therapie ≥ 50% (ohne Δ)	+ 1
	MRD = Minimal-Residual-Disease sCR = stringente Complete Remission	Kein Malus für AE bei Blutwerten, da vor Therapie ein Teil von Krankheit	-
		Single arm Studie ohne weitere Therapieoption mit Ansprechrate ORR ≥ 60%	Upgrading von C auf B möglich

Total Studienrating	5 → a	4 → a	3 → b	2 → c	1 → c	0 → d
→ Summe aus Scorepunkten und maximal 1 Bonus/Malus-Punkt						

→ Up- oder Downgrading wird vom VA begründet	B Bewertung Einzelfall		Upgrading möglich	Downgrading möglich
	→ Up- oder Downgrading um maximal eine Nutzenkategorie	Relevante Zusatz-Aufwendungen: Durch Medikation bedingte Hospitalisation/Pflegeaufwand → nötig (-) oder vermeidbar (+)		x
→ Upgrading maximal auf B	Nutzenrelevante Abweichung des Einzelfalls von der Studienpopulation: Kind/Erwachsene, Prä-/Postmenopausen, Therapielinie, Diagnose/Krankheitsstadium			x
	Einschlusskriterien: AZ, Selbstständigkeit (ADL), Arbeitsfähigkeit, Kognition, Pflegebedürftigkeit, Relevante Komorbiditäten, limitierende Lebenserwartung (nach Rücksprache mit dem behandelnden Arzt)		x	x

NUTZENKATEGORIE	A	B	C / C*	D
Grosser Nutzen = Pflichtleistung aus OKP	JA	JA	NEIN → Probetherapie möglich	Nein

## C Besonderheiten und Erläuterungen

- Regeln zur Modellbewertung** → Ein Studien-Rating erfolgt erst dann, wenn die Kriterien zum Eintritt in den Art. 71 a-d KVV erfüllt sind
- Ein Rating basiert auf publizierten Studiendaten in anerkannten, peer-reviewed Fachzeitschriften oder Publikationen mit Peer-review im Rahmen von internationalen Fachärzte-Kongressen
  - Ein Rating basiert auf der Studie mit der höchsten Evidenz, welche vom Antragsteller benannt oder zur Verfügung gestellt wird
  - OLUtool verlangt signifikante Studienergebnisse. Keine Signifikanz liegt vor, falls  $p > 0,05$  oder Konfidenzintervall CI  $\geq 1$
  - Auswertung von Subgruppen sind dann akzeptiert, wenn diese im Studiendesign eingeplant waren
  - Conditional Approval mit Forderung auf Datennachlieferung
  - Biologische Daten als 1<sup>o</sup> Studienendpunkt → Maximal Studienrating B
  - Studien ohne Kontrollgruppe (Sonderregelung Hämatologie) → Maximal Studienrating C
  - Studien ohne Indikationen, lediglich Zulassung für genetischen Marker (Tumor-Wachstumstreiber) → Maximal Studienrating C
  - Case Reports → Studienrating D, Upgrading auf C möglich, falls Antrag von Tumorboard
  - Interdisziplinäre Tumorboard-Empfehlung → Upgrading von D zu C im Schlussrating (Einzelfall) möglich
  - Interdisziplinäre Tumorboard-Empfehlung für Medikation ohne Studie, basierend auf molekulargenetischer Analyse → Kategorie C im Schlussrating (Einzelfall) möglich
- Biologische Daten** = im Klinikalltag etablierte Messparameter wie Laborwerte, Nervenleitgeschwindigkeit, bildgebende Befunde usw. Sie sind klinischen Messparametern gleichgestellt, falls eine enge Verknüpfung mit dem therapeutischen Nutzen durch den VA begründet wird
- Kategorie C / C\*** (C\*: sprich C – Stern)
- Im Regelfall beinhaltet Kategorie C, dass die Kosten der Probetherapie von der ~~Quassa~~ übernommen werden (meist 1–3 Monate).
  - C\* wird vom VA empfohlen, wenn die übliche C-Lösung vermutlich nicht umsetzbar ist, jedoch im Patienteninteresse eine Finanzierung der Probetherapie angestrebt werden sollte. Die finanzielle Abgeltung sowie die Kriterien für die Probetherapie sind auszuhandeln. Der VA unterstützt auf Wunsch den Versicherer bei Limitationen oder der Messung von Therapiezielen.
  - Ohne Lösung für die Probetherapie entfällt die Leistungspflicht, da bei C/C\* der ~~grosse~~ Nutzen zwar im Einzelfall möglich, aber nicht generell zu erwarten ist. Der Therapievorschlag soll erst starten, wenn die Finanzierung der Probetherapie und die allfällige Weiterbehandlung geregelt ist.

Studiendefizite	Generell Hinweise auf Bias
Homogenität der Studienpopulation ist ungenügend	Stupido/Volltext fehlt. Defizit ist darin begründet, dass die Studie hinsichtlich Nutzen nicht vollständig beurteilbar ist
Kontrollgruppe nicht adäquat → Vergleich nicht mit Standardtherapie oder Vergleich mit Unterdosierung	Andere beurteilte, nutzenrelevante Mängel
Wirksamkeitsdaten sind nicht eindeutig als Primär-Endpunkte in der Studie definiert	

## A Studienrating

Typ	1° Endpunkt Studie	Mess-Parameter von 1° Endpunkt	Bestimmung von Clinical Benefit
①	Funktionsänderung	%-Verbesserung ausgehend von Baseline	$\Delta\%$ Verum und Kontrolle → berechnen
②	Anzahl Ereignisse	%-Verbesserung ausgehend von Baseline	$\Delta\%$ Verum und Kontrolle → berechnen
③	Scorepunkte Messung	%-Verbesserung ausgehend von Baseline Total Scorepunkte = 100%	$\Delta\%$ Verum und Kontrolle von total Scorepunkte → berechnen
④	Surrogatwerte mit Studienrange (x - y)	%-Verbesserung ausgehend von Baseline Masszahl Studienrange (x - y) = 100%	$\Delta\%$ Verum und Kontrolle von Studienrange (x-y) → berechnen
⑤	Ansprechrate in %	%-Ansprechrate von 1° Endpunkt	$\Delta\%$ = Prozentpunkte (PP) = ARR → direkt aus Studie zu entnehmen
⑥	Mortalität	Mortalitätssenkung in Jahren oder HR	$\Delta$ direkt aus Studie zu entnehmen oder berechnen

Score	Typ ① - ④	Typ ⑤	Typ ⑥	Score
Clinical Benefit als $\Delta$ → Massgebend ist der 1° Endpunkt der Studientypen ① - ⑥ → $\Delta\%$ -Werte kommen vor HR	$\Delta \geq 50\%$	$\Delta \geq 40\%$ (PP)	$\Delta \geq 2\%$ pro Jahr	4
	$\Delta \geq 30\%$ oder $HR \leq 0,65$	$\Delta \geq 20\%$ (PP)	$\Delta \geq 1\%$ pro Jahr oder $HR \leq 0,80$	3
	$\Delta \geq 10\%$ oder $HR \leq 0,70$	$\Delta \geq 10\%$ (PP)	$\Delta \geq 0,5\%$ pro Jahr oder $HR \leq 0,85$	2
	$\Delta \geq 5\%$ oder $HR \leq 0,75$	$\Delta \geq 5\%$ (PP)		1
	$\Delta < 5\%$ oder $HR > 0,75$	$\Delta < 5\%$ (PP)	$\Delta < 0,5\%$ pro Jahr oder $HR > 0,85$	0

Bonus/Malus-Punkte	Serious Adverse Events (sAE)	$\Delta$ sAE bei Studien mit Kontrolle $\geq 30\%$ , ohne Kontrollgruppe $\geq 50\%$	Score
→ In der Summe maximal 1 Punkt → Bonus nur möglich, falls mindestens 1 Scorepunkt erreicht	• Ansprechen (RR) ohne $\Delta$	→ $\geq 30\%$ Symptombefreiheit für $\geq 1$ Jahr zu erwarten	+1
	• $\Delta$ Ansprechen (Typ ⑤) < 20%	→ Falls $HR \leq 0,35$ (Relative Risikominderung $\geq 65\%$ )	+1
	Studiendefizite	→ Falls nutzenrelevante Folgen begründet sind (siehe Liste)	-1
	QoL	Definierter Studienendpunkt mit validiertem Fragebogen erhoben und mit signifikanter Aussage. Kein Bonus, falls Ergebnis «non inferior»	+1
	Expertenbeurteilung	Internationale Fachgremien wie IQWiG/G-BA, NICE und andere sowie CH-Konsensus Empfehlungen. Der VA dokumentiert die Expertenbeurteilung	+1 / -1

Studienrating	5 → a	4 → a	3 → b	2 → c	1 → c	0 → d
→ Summe aus den Scorepunkten und Bonus/Malus-Punkt						

## B Bewertung Einzelfall

→ Up- oder Downgrading wird vom VA begründet  → Up- oder Downgrading um maximal eine Nutzenkategorie  → Upgrading auf maximal B	Empfehlung zur Beurteilung durch den Vertrauensarzt (VA)	Upgrading möglich	Downgrading möglich
	Relevante Zusatz-Aufwendungen: Durch Medikation bedingte Hospitalisation/Pflegeaufwand → nötig (↓) oder vermeidbar (↑)	X	X
	Nutzenrelevante Abweichung des Einzelfalls von der Studienpopulation: - Kind/Erwachsene, Prä-/Postmenopause, Therapielinie, Krankheitsstadium - Typ ① → Baseline des Einzelfalls liegt in relevanter Grösse ausserhalb vom Studienrange (x-y)		X
	Einzelsituation: - AZ, Selbständigkeit (ADL), Arbeitsfähigkeit, Kognition, Pflegebedürftigkeit - Relevante Komorbiditäten, limitierte Lebenserwartung → VA nimmt Rücksprache mit dem behandelnden Arzt	X	X

NUTZENKATEGORIE	A	B	C / C*	D
Grosser Nutzen = Pflichtleistung aus OKP	JA	JA	NEIN → Probetherapie möglich	Nein

**C Besonderheiten und Erläuterungen**

**Regeln zur Modellianwendung**

Ein Studien-Rating erfolgt erst dann, wenn die Kriterien zum Eintritt in den Art. 71 a-d KVV erfüllt sind

- Ein Rating basiert immer auf publizierten Studiendaten in anerkannten, peer-reviewten Fachzeitschriften oder Publikationen von internationalen Fachärzte-Kongressen mit durchgeführtem peer-review
- Ein Rating basiert auf der Studie mit der höchsten Evidenz, welche normalerweise vom Antragsteller benannt oder zur Verfügung gestellt wird. Der VA kann aber auch selbständig bestimmen, welche Studie er für das Rating verwenden will. Ob bei komplementär aufgebauten Studien das Beste aus mehreren Studien (Studienmix) akzeptiert wird, entscheidet der VA
- Bei mehreren Endpunkten und unklarem 1<sup>st</sup> Studienendpunkt, wird der Parameter mit der höchsten klinischen Relevanz bewertet
- OLUtool verlangt signifikante Studienergebnisse. Keine Signifikanz liegt vor, falls  $p > 0,05$  oder HR-Konfidenzintervall  $CI \geq 1$
- Auswertung von Subgruppen sind dann akzeptiert, wenn diese im Studiendesign eingeplant waren
- Conditional Approval mit Forderung auf Datennachlieferung → Maximal Studienrating B
- Studienvolltext liegt dem VA nicht vor (nur Abstract, Poster usw.) → Maximal Studienrating B
- Surrogatendpunkte/Biologische Daten (Siehe Erläuterungen) → Maximal 3 Score Punkte und maximal Studienrating B
- Studien ohne Kontrollgruppe → Maximal Studienrating C, falls GRF  $\geq 90\%$
- Case Reports → Studienrating D, Upgrading auf C möglich, falls Antrag von RZ
- Referenzzentrum (RZ) → Upgrading von D zu C als Schlussrating im Einzelfall möglich

**Kategorie C / C\*** (C\*: sprich C – Stern)

- Im Regelfall beinhaltet Kategorie C, dass die Kosten der Probetherapie von der Pharma übernommen werden (meist 1 – 3 Monate).
- C\* wird vom VA empfohlen, wenn die übliche C-Lösung voraussichtlich nicht umsetzbar ist, jedoch im Patienteninteresse eine Finanzierung der Probetherapie angestrebt werden sollte. Die finanzielle Abgeltung sowie die Kriterien für die Probetherapie sind auszuhandeln. Der VA unterstützt auf Wunsch den Versicherer bei Limitationen oder der Messung von Therapiezielen.
- Ohne Lösung für die Probetherapie entfällt die Leistungspflicht, da bei C/C\* der grosse Nutzen zwar im Einzelfall möglich, aber nicht generell zu erwarten ist. Der Therapieversuch soll erst starten, wenn die Finanzierung der Probetherapie und die allfällige Weiterbehandlung geregelt ist.

**Erläuterungen**

Clinical Benefit: Als Oberbegriff für den Benefit, wie er in den NonOnko Studientypen vorkommt → Weitere Ausführungen im Scriptum

Hazard Ratio (HR): Die höchste Scorepunktzahl wird nicht vergeben, da HR nur einen relativen Benefit darstellt.

Surrogatendpunkte/Biologische Daten: Im Klinikalltag etablierte Messparameter wie Laborwerte, Nervenleitgeschwindigkeit, bildgebende Befunde usw. Sie sind klinischen Messparametern gleichgestellt, falls eine enge Verknüpfung mit dem therapeutischen Nutzen durch den VA begründet wird.

Referenzzentrum (RZ): Im «Nationalen Konzept Seltene Krankheiten» des Bundes sollen spezielle Zentren bezeichnet werden, die befähigt sind, seltene Krankheiten abzuklären und zu therapieren. Die Bezeichnung dieser Zentren durch das BAG ist noch ausstehend. Wir betrachten universitäre Zentren oder Fachzentren mit universitärer Anbindung als gleichwertig.

Beizug von Experten: Wir empfehlen den Beizug von Experten, falls Unklarheiten zur klinischen Relevanz bestehen.

Durchlässigkeit der Modelle: Es kann vorkommen, dass eine Studie aus der Onkologie anhand der Messparameter besser zum NonOnko Modell passt und umgekehrt. In diesen eher selten Fällen (meist im Bereich Hämatologie) kann jeweils das besser passende Modell angewendet werden.

**Studiendefizite - Auswahl einiger Prüfkriterien**

Homogenität der Studienpopulation ist ungenügend	Generell Hinweise auf Bias
Kontrollgruppe nicht adäquat → Vergleich nicht mit Standardtherapie oder Vergleich mit unterdosierter Medikation	Studien-Volltext fehlt. Defizit liegt darin begründet, dass die Studie hinsichtlich Nutzen, nicht vollständig beurteilbar ist
Wirksamkeitsdaten nicht eindeutig als Endpunkte in der Studie definiert	Andere begründete, nutzerrelevante Mängel