

# Der prädiktive Wert der Frühbewertung von Arzneimitteln

*Fricke und Klaus haben eine Methode zur Frühbewertung des therapeutischen Stellenwerts von neuen Wirkstoffen begründet, die in Deutschland als Standard gilt. Sie bildet insbesondere die Grundlage für die primär politisch motivierte Aussage, die Hälfte aller in Deutschland zugelassenen Arzneimittel seien Scheininnovationen ohne eigenen therapeutischen Stellenwert. Der Beitrag zeigt auf, dass diese Methode weit hinter heutigen Anforderungen an eine vergleichende Nutzenbewertung zurück bleibt, die Arzneimittel im Sinne einer Gesundheitstechnik beurteilt (HTA). Sie ist unsystematisch, nicht transparent und weit gehend auf subjektive Ansichten gegründet.*

■ Ariane Höer, Bertram Häussler

## 1. Hintergrund

Im Jahr 1976 wurde das Arzneimittelgesetz geändert und bestimmte in der Neufassung, dass erstmals ab dem 1.1.1978 Wirkstoffe vor ihrer Zulassung vom Bundesgesundheitsamt bezüglich ihrer Qualität, der therapeutischen Wirksamkeit und Unbedenklichkeit zu bewerten seien. Diese Aufgabe wurde zwischenzeitlich an das Bundesamt für Arzneimittel und Medizinprodukte (BfArM) delegiert (Institut des BMG). Bewertung der Wirksamkeit heißt, dass von der Zulassungsbehörde geprüft wird, ob ein blutdrucksenkendes Mittel auch tatsächlich den Blutdruck senkt oder ob ein Analgetikum in der Lage ist, Schmerzen zu lindern. Es war und ist nicht vorgesehen, dass die Behörde im Rahmen der Zulassung auch den therapeutischen Stellenwert neuer Wirkstoffe beurteilt, also ob ein neuer Wirkstoff im Vergleich zu bereits vorhandenen Wirkstoffen besser wirkt.

Eine Bewertung des therapeutischen Stellenwerts oder noch präziser des therapeutischen Zusatznutzens eines Wirkstoffs unter den Rahmenbedingungen des jeweiligen Gesundheitssystems in medizinischer und/oder ökonomischer Hinsicht wird derzeit in zahlreichen Ländern vorgenommen. Solche Bewertungen haben in der Regel Entscheidungen zur Folge im Hinblick auf die Höhe der Kostenerstattung (Referenzpreise bzw. Festbeträge), den Verkaufspreis oder auf die Verordnungsfähigkeit eines Arzneimittels. Eine frühzeitige Bewertung wird in der Regel angestrebt, weil die jeweilige Entscheidung in der Regel vor dem Markteintritt getroffen werden soll. Im derzeit vorliegenden „Eckpunktetpapier“ (BMG 2006) ist erstmals für Deutschland vorgesehen, dass eine frühe Bewertung von Arzneimitteln der Findung „fairer Erstattungshöhen“ zu Grunde gelegt werden soll.

*Ariane Höer, Bertram Häussler, IGES – Institut für Gesundheits- und Sozialforschung, Berlin, Deutschland*

In Deutschland wird die Bewertung des therapeutischen Stellenwerts – hier überwiegend als „Nutzenbewertung“ bekannt – seit 2004 durch das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) nach Beauftragung durch den Gemeinsamen Bundesausschuss (GBA) vorgenommen. Dieses hat seit seiner Gründung insgesamt vier Aufträge zur Bewertung von Arzneimitteln abschließend bearbeitet. Darunter sind drei, die nach der im Methodenpapier beschriebenen ausführlichen Vorgehensweise durchgeführt wurden und eine, die im Verfahren einer „verkürzten Berichterstattung“ als „Rapid Report“ durchgeführt worden ist.

Eine Frühbewertung von Arzneimitteln wird in Deutschland jedoch seit langem auch von den Pharmakologen Fricke und Klaus vorgenommen, die ihre Ergebnisse seit 1982 publizieren. Sie haben 733 Arzneimittel, die zwischen dem 1.1.1978 und dem 31.12.2004 neu auf den Markt gekommen sind, einer Bewertung unterzogen und nach ihrem Klassifikationssystem einer Kategorie zwischen „A“ und „D“ zugeordnet.

Die Bewertungen von Fricke und Klaus haben eine zentrale gesundheitspolitische Bedeutung insbesondere dadurch erlangt, dass sie die Grundlage für die weit verbreitete Ansicht bilden, dass die Hälfte aller in Deutschland auf den Markt kommenden Arzneimittel – jene, die von Fricke und Klaus mit „C“ bewertet worden sind – nicht innovativ seien und keinen eigenen therapeutischen Stellenwert hätten. In jüngster Zeit werden „C“-bewertete Arzneimittel auf so genannte „Me-Too-Listen“ gesetzt, verbunden mit der Empfehlung, von der Verordnung dieser Arzneimittel Abstand zu nehmen. Es muss daher damit gerechnet werden, dass durch Verweise auf die Bewertungen von Fricke und Klaus diese in Zukunft einen deutlichen Einfluss auf das Verordnungsgeschehen haben werden. Ärzte werden möglicherweise vermehrt von der Verordnung eines Arzneimittels Abstand neh-

men das mit „C“ bewertet wurde, wenn angekündigt wird, dass die Verordnung teurer „Me-Toos“ in der Wirtschaftlichkeitsprüfung negativ sanktioniert wird.

Bislang war der Markterfolg von Arzneimitteln überwiegend von Qualitäts- und Preisvorteilen abhängig (Häussler et al. 2002). Die Kategorisierung von Fricke und Klaus hatte vermutlich eher geringe Bedeutung, „C“-klassifizierte Wirkstoffe hatten durchaus die Chance, sich am Markt zu entfalten, wodurch sich die Möglichkeit ergab, dass ihre Eigenschaften unter den Bedingungen der täglichen Praxis („real life“) im Vergleich zu anderen Wirkstoffen sichtbar werden konnten. Auf dieser Grundlage ergibt sich die Möglichkeit, die Ergebnisse der Frühbewertung durch Fricke und Klaus mit Bewertungen derselben Wirkstoffe zu vergleichen, die zu einem späteren Zeitpunkt durchgeführt worden sind.

Wir stellen daher im Folgenden die Frage, wie sich die Beurteilung von Wirkstoffen aktuell darstellt, die zum Zeitpunkt der Markteinführung als „C = nicht innovativ und ohne zusätzlichen Nutzen“ bezeichnet worden waren und die durch diese Frühbewertung das Etikett „die Verordnung wird nicht empfohlen“ erhielten. Vorab beschreiben wir die Vorgehensweise von Fricke und Klaus und diskutieren diese vor dem Hintergrund von Erfahrungen aus der Nutzenbewertung von Arzneimitteln.

## 1 Die Methode von Fricke und Klaus

Im Folgenden soll die Methode von Fricke und Klaus dargestellt und diskutiert werden. Es wird dabei auf die Eigenschaften der Methode, deren Basis sowie die verwendete Skala eingegangen.

### 1.1 Grundsätzliche Eigenschaften

Kennzeichnend für die Bewertung durch Fricke und Klaus ist vor allem, dass sie relativ kurz nach der Marktzulassung erfolgt. Die chemische Struktur spielt eine wesentliche Rolle bei der Bewertung. Die erste Publikation des Bewertungsergebnisses erfolgte ursprünglich in der von Fricke und Klaus gegründeten und herausgegebenen Reihe „Neue Arzneimittel“. Seit 1987 werden die Ergebnisse der Bewertung im Jahr nach der Markteinführung eines Wirkstoffs jedoch zunächst in Kurzform im Arzneiverordnungs-Report (AVR) publiziert. Die vollständige Darstellung der Bewertung erfolgt ein Jahr später in der zitierten Reihe. So erscheinen die Bewertungen der im Jahr 2004 neu eingeführten Arzneimittel zunächst als Kurzform im AVR 2005 und als ausführliche Variante in der Ausgabe 2006 der Reihe „Neue Arzneimittel“.

Die Bewertungen der in der Kurzform sowie in der später erscheinenden ausführlichen Form unterscheiden sich hinsichtlich der Ausführlichkeit der Begründung, nicht aber hinsichtlich der Einteilung in eine der Bewertungskategorien. Ein sehr wesentliches Charakteristikum

des Bewertungsverfahrens besteht darin, dass die ursprünglich getroffene Einteilung in eine Bewertungsklasse nie wieder verändert wird, auch wenn später neue Erkenntnisse zu einem Wirkstoff vorliegen. Die ursprüngliche Bewertung liegt allen Statistiken zu Grunde, die z.B. vom AVR, aber auch von zahlreichen anderen Autoren zur Beurteilung der Innovationsleistung der pharmazeutischen Industrie herangezogen wird.

### 1.2 Empirische Basis der Bewertungen

Fricke und Klaus gründen ihre Bewertungen seit 25 Jahren unverändert auf drei unterschiedliche Quellen:

1. „von Herstellern auf Anfrage zur Verfügung gestellten wissenschaftlichen Literatur und Produktinformationen, (...)“
2. sonstigen einschlägigen wissenschaftlichen Publikationen sowie (...)“
3. allgemeinen pharmakologischen und therapeutischen Erfahrungskriterien.“ (Fricke und Klaus 1982, 2005).

Die empirische Basis der Bewertungen wird nicht nach einem standardisierten Verfahren gewonnen, wie sie von den gegenwärtig praktizierten Systemen zur Nutzenbewertung angewendet werden. Es ist nicht erkennbar, dass eine systematische Literaturrecherche nach vorab definierten Suchkriterien durchgeführt wird.

Fricke und Klaus erkennen den subjektiven Charakter ihrer Bewertung selbst an und bezeichnen ihre Methode als den „Versuch einer Stellungnahme“. Sie räumen ein, dass ihr Vorgehen dadurch beeinträchtigt sei, dass für die meisten Mittel zum Zeitpunkt ihrer Markteinführung nur Informationen über die pharmakologischen Eigenschaften und die in den ersten klinischen Studien festgestellten Wirkungen am Patienten vorlägen, dagegen in der Regel noch keine ausreichenden therapeutischen Erfahrungen auf breiterer Basis gesammelt werden konnten, die den therapeutischen Stellenwert der einzelnen Präparate schon zuverlässig bestimmen ließen.

### 1.3 Die Bewertungsskala

Die Bewertung erfolgt durch einen Index, der aus drei Elementen zusammengesetzt ist:

- Reihenfolge der Zulassung innerhalb einer Wirkstoffgruppe,
- Wirksamkeit relativ zu den bisher zugelassenen Wirkstoffen,
- Beurteilbarkeit des Wirkprinzips und / oder der Wirksamkeit.

Diese drei Elemente werden so kombiniert, dass insgesamt die vier Klassen „A“, „B“, „C“ und „D“ entstehen (siehe Tabelle 1).

Die vier Klassen werden in ein Kontinuum gebracht, in dem „A“ als die „wirkliche Innovation“ mit unzweifel-

hafter therapeutischer Wirksamkeit gesehen wird. „B“ hat zwar Verbesserungen gegenüber „A“, gilt aber als „nicht wirklich“ innovativ. Weder innovativ noch verbessert sind „C“-klassifizierte Wirkstoffe, sie werden häufig als „Scheininnovationen“ bezeichnet bzw. als „Me-Too“.

### 1.3.1 Reihenfolge

Aus Tabelle 1 wird erkennbar, dass die Wirkstoffe in Bezug auf die Beurteilung ihrer Wirksamkeit unterschiedlich behandelt werden. Der in einer Klasse zuerst eingeführte Wirkstoff wird mit „A“ belegt, was ihn davor bewahrt, zu irgendeinem Zeitpunkt seiner Existenz relativ zu anderen bewertet zu werden. Dies kann dazu führen, dass ein neues Arzneimittel, das zu einem späteren Zeitpunkt eingeführt wird und wesentlich besser ist als „A“ nur ein „C“ bekommt, selbst wenn es gleich gut ist wie ein mit „B“ belegter Wirkstoff.

Dem Verfahren ist also eine Bevorzugung des ersten Wirkstoffs einer Klasse inhärent, obwohl bekannt ist, dass über die Reihenfolge der Marktzulassung von Wirkstoffen oftmals zufällige Faktoren entscheiden und der Markteintrittszeitpunkt beim gleichen Wirkstoff sich sogar international unterscheiden kann (so kann ein Wirkstoff deutlich früher in den USA als in den Europa zugelassen werden und umgekehrt).

### 1.3.2 Relative Wirksamkeit

Nach den Angaben von Fricke und Klaus richten sie ihr Hauptaugenmerk bei der Bewertung auf den therapeutischen Effekt, während mögliche Nebenwirkungen in der Regel von untergeordneter Bedeutung sind. In der Nutzenbewertung ist es Standard, vorab zu erläutern, woran der Nutzen gemessen werden soll (z.B. Bastian et al. 2005).

In der Regel werden sogenannte Endpunktstudien, in denen patientenrelevante Ereignisse (bspw. Mortalität) untersucht werden, höher gewertet als Studien, die sich mit dem Effekt eines Wirkstoffs auf sogenannte „Surrogatparameter“ befassen, also etw. das Ausmaß der Blutdrucksenkung. Je nach Datenlage werden von Fricke und Klaus beide der genannten Studientypen herangezogen. Die Autoren erläutern jedoch in der Darstellung ihrer Methode nicht, auf welcher Basis der therapeutische Effekt prinzipiell bewertet wird oder welchen Stellenwert Studien unterschiedlicher Evidenzgrade in der Bewertung haben.

### 1.3.3 Beurteilbarkeit

Dieses Element der Beurteilungsskala bezeichnet im Unterschied zu den beiden anderen Elementen keine Eigenschaft des bewerteten Wirkstoffs sondern die Beleglage aus der Sicht der Bewerter. Aus der Tatsache, dass ein „D“ in nur knapp zwei Prozent aller Fälle vergeben worden ist (siehe unten, Tabelle 2), lässt sich schließen, dass die Bewerter in der großen Mehrheit der Fälle die Beleglage für ausreichend hielten. Es ist daher zu vermuten, dass Fricke und Klaus die Wirksamkeit von Wirkstoffen, die sie ausschließlich mit „D“ klassifizieren, für wissenschaftlich nicht begründbar halten. Zu den zwölf Wirkstoffen, die von Fricke und Klaus mit „D“ bewertet wurden gehören u.a. drei Phytopharmaka, Guar und Ganglioside aus Rinderhirn.

## 1.4 Die praktische Anwendung der Klassifikation

Wie bereits erwähnt, haben Fricke und Klaus bisher 733 Wirkstoffe bewertet, die seit dem 1.1.1978 neu auf den Markt gekommen sind. Die Bewertungen sind in der Reihe „Neue Arzneimittel“ erschienen, die von Fricke

Tabelle 1: Bewertungsskala von Fricke und Klaus

Klasse	Bezeichnung nach Fricke und Klaus	Reihenfolge	Wirksamkeit relativ zu bestehenden Wirkstoffen	Beurteilbarkeit
A	Innovative Struktur oder neuartiges Wirkprinzip mit therapeutischer Relevanz	erster in einer Wirkstoffklasse	entfällt	vorhanden
B	Verbesserung pharmakodynamischer oder pharmakokinetischer Eigenschaften bereits bekannter Wirkprinzipien	nicht erster	besser	vorhanden
C	Analogpräparat mit keinen oder nur marginalen Unterschieden zu bereits eingeführten Präparaten	nicht erster	nicht besser	vorhanden
D	Nicht ausreichend gesichertes Wirkprinzip oder unklarer therapeutischer Stellenwert	erster / nicht erster	entfällt (nicht definiert)	nicht vorhanden

Quelle: IGES-Darstellung nach Fricke und Klaus

und Klaus herausgegeben wird (1982, 1983, 1985, 1986, 2000, 2001) bzw. wurden in den Ausgaben des AVR als Kurzform publiziert (Schwabe und Paffrath 1987 ff.).

Das Ergebnis der Bewertungen lautet für ca. je ein Viertel der Wirkstoffe „A“ oder „B“ und in etwa der Hälfte der Fälle „C“ (siehe Tabelle 2). Eine detaillierte Betrachtung zeigt jedoch, dass in der Praxis in 78 Fällen (10,6%) mehr als nur eine Bewertung vergeben wurde. Vor allem in Kombination mit einer „A“-Bewertung sind kombinierte Klassifikationen wie z.B. „A/C“ nicht außergewöhnlich. Die Kombination von „B“ und „C“ wird sogar in beiden Reihenfolgen („B/C“ und „C/B“) vergeben. In der Beschreibung der Methode sind solche Klassifikationen nicht vorgesehen, so dass es unklar bleibt, was damit gemeint ist. Im AVR werden die kombinierten Klassifikationen unkommentiert wiedergegeben. Eine Definition der Mischklassifikationen wird fallweise in der ausführlichen Darstellung für einzelne Wirkstoffe in der Reihe „Neue Arzneimittel“ (Fricke und Klaus 1982 ff.) geliefert.

### 1.5 Zusammenfassende Bemerkungen zur Methode

Die Bewertungsmethode von Fricke und Klaus muss vor dem Hintergrund ihrer Entstehungsgeschichte gesehen werden. Sie ist zu Beginn der 80er Jahre entstanden, zu einer Zeit, als das methodische Inventar der evidenzbasierten Medizin noch weitgehend unbekannt war. Es war damals innovativ, über die Zulassungsunterlagen von Arzneimitteln hinaus eine Bewertung des therapeutischen Stellenwerts vorzunehmen. Es war jedoch unbekannt, eine systematische Methode zu entwickeln, die die Bewertungsergebnisse von der subjektiven Ebene wegführen. Ein Blick auf das umfangreiche methodische Inventar von Institutionen wie dem britischen NICE oder dem deutschen IQWiG zeigt, wie weit Fricke und Klaus hinter modernen Standards zurück bleiben. Ihre Ergebnisse sind daher in starkem Maße vom nicht-explizierten methodischen Glauben der Autoren abhängig.

Mit einer heute gebräuchlichen Formulierung haben Fricke und Klaus eine „eminenzbasierte“ Methode begründet und haben diese bis heute beibehalten. Dies scheint sich weitgehend unbeobachtet vollzogen zu haben, da bisher keine methodische Kritik an einer so wenig nachprüfaren Vorgehensweise geäußert worden ist.

## 2 Evaluation der Ergebnisse der Frühevaluation von Fricke und Klaus

Es ist bekannt, dass sich die vergleichende Bewertung von Wirkstoffen ändert, wenn im Lauf ihrer Anwendung in der Praxis neue Erkenntnisse anfallen. Dies kann auf verschiedene Gründe zurück zu führen sein, zu denen unter anderem folgende gehören:

- Der Wirkstoff, der eine Gruppe neu begründete, wird vom Markt genommen und der zunächst als „Me-To“ eingestufte Wirkstoff verbleibt im Markt (z.B. der Wirkstoff Entacapon, der nach Marktrücknahme von Tolcapon als einziger COMT-Hemmer für die Behandlung des Morbus Parkinson zur Verfügung steht).
- Ein neuer Wirkstoff erweist sich trotz ähnlichen Wirkprinzips als besser verträglich (bspw. zeichnen sich die selektiven Serotoninwiederaufnahme-Hemmer (SSRI) Citalopram, Paroxetin und Sertralin durch weniger Wechselwirkungen im Vergleich zu den Vorläufern Fluoxetin und Fluvoxamin aus).
- Ein Wirkstoff wird für zusätzliche Indikationen zugelassen und dann primär eingesetzt als der mit „A“ bewertete Analogwirkstoff (so wird Nafarelin – wie sein Vorläufer Leuprorelin – für die hormonsuppressive Therapie bei Endometriose eingesetzt, doch ist die Hauptindikation von Nafarelin vermutlich die Vorbereitung der assistierten Fertilisation).

Der Zuordnung von Wirkstoffen zu der Kategorie „C“ kommt die gesundheitspolitisch größte Bedeutung zu,

Tabelle 2: Übersicht zur Bewertung von 733 seit 1978 eingeführten neuen Arzneimittel durch Fricke und Klaus

Erste Nennung	ggf. weitere Nennung					gesamt	in %
	A	B	C	D			
A	149	4	20	23		196	26,7
B		163	18	1		182	24,8
C		4	331	8		343	48,8
D				12		12	1,6
	149	171	369	44		733	100,0

Quelle: IGES nach Schwabe und Paffrath (1987 ff.) sowie Fricke und Klaus (1982, 1983, 1985, 1986)

weil sie in zunehmendem Maße als Begründung für einen Verzicht zur Verwendung dieser Arzneimittel Verwendung findet und daher einen zunehmend starken Einfluss auf das Verordnungsgeschehen nimmt. Es soll daher untersucht werden, inwiefern die Bewertungen zu Beginn des Markteintritts zu späteren Zeitpunkten noch Bestand haben.

## 2.1 Daten und Vorgehensweise

Da – wie bereits beschrieben – Fricke und Klaus nur diese Frühbewertung vornehmen, müssen als Quellen für spätere Bewertungen andere Publikationen herangezogen werden. Hierzu bieten sich vor allem die Bewertungen im AVR an, weil diese Publikation eine enge konzeptionelle Nähe zu Fricke und Klaus hat: Im AVR werden die Ergebnisse der Klassifikation jeweils von den Autoren Fricke und Schwabe im Kapitel „Neue Arzneimittel“ zuerst publiziert. In keiner anderen Publikation außer der Reihe „Neue Arzneimittel“ findet sich eine so detaillierte Darstellung entsprechend der Vorgehensweise von Fricke und Klaus wie im AVR. Schließlich gibt der AVR auch in jedem Jahr eine Übersicht der Anzahl von Wirkstoffen insgesamt sowie der Anzahl innovativer Wirkstoffe, die seit 1986 auf den Markt kamen. Im AVR finden sich außerdem jährlich aktualisierte Themenkapitel zu fast allen Indikationsgebieten. In diesen wird die Qualität

einzelner Wirkstoffe kommentiert und diese Kommentare im AVR bilden daher die zentrale Quelle für „Spätbewertungen“, die wir zum Vergleich mit den Frühbewertungen heranziehen.

Insgesamt haben Fricke und Klaus 343 Wirkstoffe mit „C“ bewertet. Darunter befanden sich zwölf Kontrastmittel und acht Zytostatika. Diese Wirkstoffe haben wir von der Evaluation ausgeschlossen, weil einerseits Kontrastmittel als Wirkstoffgruppe im AVR überhaupt nicht besprochen werden. Von den Zytostatika werden im AVR nur solche Wirkstoffe erwähnt, die zu den 3000 häufigsten ambulant verordneten Arzneimitteln gehören, was im AVR 2005 nur einer Anzahl von sieben klassischen Zytostatika entsprach. Zudem ist eine pauschale Bewertung von Zytostatika kaum möglich und selbst in Standardwerken der Onkologie (z.B. Schmoll et al. 2006) findet sich keine wertende Kommentierung zu den einzelnen Wirkstoffen. Ferner haben wir jene 68 Wirkstoffe von der Evaluation ausgeschlossen, die heute nicht mehr auf dem Markt sind.

Von den verbleibenden 255 „C“-Wirkstoffen sind 42 Antinfektiva, für deren spätere Bewertung wir die Publikation „Antibiotika-Therapie“ (Stille et al. 2005) konsultieren, die als Standardwerk auf ihrem Gebiet gilt. Für die verbliebenen 213 Wirkstoffe wird der AVR für das Jahr 2005 herangezogen. Da in den verwendeten Quellen insgesamt 62 Wirkstoffe nicht erwähnt wurden und

weitere 62 Wirkstoffe zwar erwähnt, aber nicht kommentiert wurden, verbleiben insgesamt 30 Antinfektiva und 101 übrige Wirkstoffe, die unserer Evaluation unterzogen werden.

Im Zuge dieser Evaluation wurde untersucht, welche Botschaft in Bezug auf die Verwendung der in der Frühbewertung mit „C“ klassifizierten Wirkstoffe aus heutiger Sicht abgegeben wird. Die heutigen Bewertungen wurden vor allem daraufhin untersucht, ob ihnen noch immer die Botschaft zu entnehmen ist, dass bei der Verordnung der Wirkstoffe besondere Zurückhaltung geübt bzw. ganz darauf verzichtet werden soll.

Dazu wurden die Kommentierungen im AVR 2005 bzw. bei Stille et al. (2005) wie folgt klassifiziert (Beispiele siehe Tabelle 4):

**Tabelle 3: Übersicht der 343 Wirkstoffen, die seit 1978 auf dem Markt eingeführt wurden und von Fricke und Klaus mit C, C/B oder C/D klassifiziert wurden**

	Antinfektiva (Quelle: Stille 2005)		Alle übrigen Wirkstoffe (Quelle: AVR 2005)		Gesamt abs.
	abs.	in %	abs.	in %	
<b>In die Untersuchung einbezogene Wirkstoffe</b>	42	100,0	213	100,0	255
Nicht erwähnt in Quelle	9	21,4	53	24,9	62
Erwähnt aber nicht kommentiert in Quelle	3	7,1	59	27,7	62
Kommentiert in Quelle = in die Evaluation einbezogen	30	71,4	101	47,4	131
<b>In die Untersuchung nicht einbezogene Wirkstoffe</b>					88
2005 nicht mehr im Handel					68
Zytostatika					8
Kontrastmittel					12
Alle mit „C“ klassifizierten Wirkstoffe					343

Quelle: IGES-Analyse auf der Basis von Stille (2005) und AVR (2005)

- **Eindeutig positiv bzw. sicher nicht negativ:** Der Kommentar weist auf einen belegten oder möglichen Zusatznutzen hin (incl. bessere Verträglichkeit im Vergleich zu anderen Wirkstoffen) oder der Kommentar nennt für den Wirkstoff eine zusätzliche Therapieoption oder der Wirkstoff wird als Reservetherapeutikum für bestimmte Indikationen bezeichnet. In diese Kategorie fallen auch Kommentare, die für mehrere Wirkstoffe einer Gruppe Vorteile nennen, ohne diese untereinander zu vergleichen sowie Kommentare, die einen oder mehrere Wirkstoffe als Behandlungsoption für bestimmte Indikationen nennen.
- **Unentschieden:** Der Kommentar stellt die positiven, aber auch negativen Eigenschaften des Wirkstoffs differenziert dar. Es wird jedoch darauf hingewiesen, dass eine abschließende Beurteilung noch immer nicht möglich ist oder dem Sinn nach dem Text nicht zu entnehmen ist.
- **Negativ:** Der Kommentar attestiert dem Wirkstoff Nachteile, fehlende Wirksamkeit, lehnt die Verordnung aus Kostengründen ab bzw. weist auf die höheren Kosten im Vergleich zu anderen Wirkstoffen hin.

Bei einem eindeutig positiv bzw. sicher nicht negativ eingestuftem Kommentar wurde davon ausgegangen, dass ein Signal zu zurückhaltender Verordnung heute nicht mehr gegeben wird. Ob sich die Wirkstoffe derzeit noch am Markt befinden, wurde auf Grundlage der Angaben des ABDA-Artikelstamms überprüft (ABDATA 2005).

## 2.2 Ergebnisse

Eine Übersicht der Ergebnisse unserer Evaluation von 131 Wirkstoffen, die in der Frühbewertung durch Fricke und Klaus mit „C“ klassifiziert wurden, zeigt Tabelle 5.

Für insgesamt 86 Wirkstoffe (65,5%) fand sich eine positive oder eindeutig nicht negative Kommentierung. Für 24 der Wirkstoffe war die Wertung uneindeutig, d.h. es ließ sich weder eine eindeutig positive noch eine eindeutig negative Beurteilung aus dem Kommentar herauslesen. Nur für insgesamt 21 Wirkstoffe fand sich ein negativer Kommentar. Hierbei ist hervorzuheben, dass bei sechs dieser Wirkstoffe der Kommentar deshalb als negativ eingestuft wurde, weil von der Verordnung nicht wegen der therapeutischen Qualitäten des Wirkstoffs ab-

**Tabelle 4: Beispiele für Einstufung der Kommentare der Spätwertung von ursprünglich mit „C“ klassifizierten Wirkstoffen als „eindeutig positiv bzw. sicher nicht negativ“, „uneindeutig“ bzw. „negativ“.**

Einstufung	Wirkstoff ATC-Kode	Einführung	Quelle	Kommentar
<i>eindeutig positiv bzw. sicher nicht negativ</i>	Ramipril C09AA05	1990	AVR 2005; S. 258	Die in der AIRE-Studie belegte Erhöhung der Überlebenschancen durch Ramipril bei herzinsuffizienten Patienten nach akutem Myokardinfarkt war auch fünf Jahre nach Therapiebeginn noch nachweisbar
	Ciprofloxacin J01MA02	1991	Stille et al. 2005; S. 127	Klassische Standardsubstanz der Fluorchinolone
	Ropinirol N04BC04	1997	AVR 2005; S. 813	Dopaminagonist; erster Vertreter der Nichtergolinderivate; deutlich seltener Dyskinesien als mit Levodopa
<i>uneindeutig</i>	Glimepirid A10BB12	1996	AVR 2005; S. 377	Die Vorzüge von Glimepirid werden daher nicht einheitlich beurteilt. Während viele Diabetologen Glimepirid gegenüber anderen beta-zytotropen Substanzen bevorzugen, fasst eine niederländische Autorengruppe ihre Beurteilung mit einem kritischen Kommentar zusammen
	Dutasterid G04CB02	2003	AVR 2005; S. 925	Hemmt zusätzlich TYP I der 5alpha-Reduktase, ob dadurch ein klinisch relevanter Vorteil besteht, wurde bisher nicht überprüft
	Escitalopram N06AB10	2003	AVR 2005; S. 833	Wurde in zahlreichen Studien untersucht und erwies sich zum Beispiel dem Venlafaxin in einer kontrollierten Studie als ebenbürtig. Manche Autoren nehmen einen schnelleren Wirkungseintritt als bei Citalopram an, trotzdem ist Citalopram Standard in kürzlich angelaufener STAR*D-Studie.
<i>negativ</i>	Flecainid C01BC04	1982	AVR 2005; S. 320	In der CAST-Studie gab es häufiger Todesfälle oder Herzstillstände unter Flecainid oder Encainid; „mehr Schaden als Nutzen“
	Norfloxacin J01MA06	1984	Stille et al. 2005; S. 141	Norfloxacin ist kein Ersatz für systemisch wirksame Fluorchinolone, sondern Ersatz für Harnwegsdesinfizienzien.
	Insulin glulisin A10AB06	2004	AVR 2004; S. 76	Anwendung kann nicht empfohlen werden

Quelle: IGES, AVR 2005

geraten wurde, sondern weil andere, gleich wirksame Stoffe derzeit kostengünstiger sind. Als Beispiel seien hier die Statine Pravastatin und Fluvastatin genannt. Im AVR 2005 finden sich keinerlei Hinweise darauf, dass sich diese beiden Wirkstoffe qualitativ von Simvastatin unterscheiden (S. 702 ff.). Es wird jedoch im selben Kapitel (S. 711 f.) das Einsparpotenzial berechnet, dass sich durch die Substitution dieser Wirkstoffe durch Simvastatin erreichen ließe. Daher wurde von uns die Kommentierung für Fluvastatin und Pravastatin als „negativ“ eingestuft.

### 3 Methodische Diskussion

Bei nur 21 von 131 Wirkstoffen, die wir unserer Evaluation unterziehen konnten, wurde die von Fricke und Klaus durch die „C“-Kategorisierung initial ausgesprochene negative Empfehlung zu einem späteren Zeitpunkt aufrecht erhalten. Dies sind weniger als 20 Prozent der Wirkstoffe – nur zehn Prozent waren es, wenn man lediglich die Antiinfektiva betrachtet. In fast zwei Drittel der Fälle entsprach die spätere Bewertung entweder einer eindeutigen Empfehlung oder es wurde explizit keine negative Bewertung abgegeben. Die Erfahrungen, die mit den Wirkstoffen im Laufe der Zeit gemacht werden konnten, haben offensichtlich zu einer veränderten Bewertung geführt. Der prädiktive Werte einer Frühbewertung nach dem Verfahren von Fricke und Klaus muss daher stark in Frage gestellt werden.

Eine Reihe von Einwänden gegen dieses Ergebnis soll im Folgenden diskutiert werden.

- Dass die spätere Bewertung von anderen Autoren vorgenommen wurde und dass dies zu anderen Urteilen führen konnte, kann durchaus möglich sein. Dem ist allerdings entgegen zu halten, dass die Mehrzahl der Bewertungen aus dem AVR stammt, dessen Heraus-

geber der Methode von Fricke und Klaus sehr nahe stehen. Diese stehen im Übrigen auch nicht im Verdacht, systematisch zu Gunsten innovativer Produkte zu urteilen. Auch die Autoren der Antibiotika-Bewertungen (Stille et al.) sind für ihre kritische Position bekannte Experten.

- Dass die evaluierten 131 Wirkstoffe eine gezielte Auswahl aus den insgesamt 343 mit „C“ bewerteten Wirkstoffen darstellen könnten, könnte das Ergebnis zu Ungunsten einer Konstanz der Urteile über die Zeit verzerrern. Die Selektivität der Auswahl ist durchaus möglich, aber sie wurde nicht von uns getroffen, sondern von den Autoren des AVR bzw. der „Antibiotika-Therapie“. Sie haben sich bei ihrer Auswahl vermutlich durch die therapeutische Bedeutung der Arzneimittel leiten lassen, so dass sich die Evaluation auf therapeutisch bedeutsame Wirkstoffe bezieht. Unsere Ergebnisse sind daher sicherlich relevant, wenn auch nicht vollständig.

### 4 Praktische Bedeutung der Ergebnisse

Aus unserer Sicht haben die Ergebnisse unserer Evaluation folgende Bedeutung:

1. Die Bewertungen von Fricke und Klaus sollten *im Sinne einer Frühbewertung* mit großer Vorsicht rezipiert werden. Sie sind methodisch nicht auf dem neuesten Stand, intransparent und stark von den subjektiven Ansichten der Bewerter geprägt. Ihr prädiktiver Wert ist – zumindest für Wirkstoffe, die eine gewisse therapeutische Bedeutung erlangt haben – in Bezug auf ablehnende Urteile – gering.
2. Die Verwendung der Klassifikation von Fricke und Klaus zur Beschreibung des innovativen bzw. nicht-innovativen Charakters des Wirkstoffangebots auf dem deutschen Arzneimittelmarkt schließt sich erst

Tabelle 5: Ergebnis der 131 in die Untersuchung einbezogenen Wirkstoffe hinsichtlich ihrer Kommentierung im aktuellsten AVR (2005) bzw. in Standardwerken (Stille 2005)

Botschaft in Bezug auf den zusätzlichen Nutzen zum gegenwärtigen Zeitpunkt (2005) ist ...	Antiinfektiva (Quelle: Stille 2005)		Alle übrigen Wirkstoffe (Quelle: AVR 2005)		Gesamt	
	abs.	in %	abs.	in %	abs.	in %
... eindeutig negativ	3	10,0	18	17,8	21	16,0
... uneindeutig	0	0,0	24	23,8	24	18,3
... eindeutig positiv bzw. sicher nicht negativ	27	90,0	59	58,4	86	65,6
	30	100,0	101	100,0	131	100,0

Quelle: IGES-Analyse auf der Basis von Stille (2005) und AVR (2005)

recht aus, weil es sich nur um Frühbewertungen handelt, die als Spätbewertungen nicht haltbar sind. Es stellt sich die Frage, warum Fricke und Klaus einer solchen Verwendung ihrer Methode nicht selbst entgegen treten, die eine äußerst große Verbreitung gefunden hat (AVR, Arzneimittelbrief, Scholz und Schwabe, Kaesbach und Schleert u.a.).

3. Wenn man davon ausgeht, dass auch eine methodisch einwandfreie Frühbewertung nicht auf Erfahrungen einer breiten Anwendung zurück greifen kann, stellt sich vor dem Hintergrund unserer Ergebnisse die Frage, wie valide eine Frühbewertung überhaupt sein kann, wie gut sie die therapeutische Leistungsfähigkeit eines Wirkstoffs überhaupt vorhersehen kann. Wenn absehbar ist dass auf der Basis dieser Frühbewertungen Entscheidungen getroffen werden, die indirekt zum Ausschluss von Arzneimitteln führen, weil der Marktzugang im Bereich der GKV praktisch eingeschränkt wird, sind Einbußen bei der Verfügbarkeit effektiver Arzneimittel die Folge.
4. Eine sinnvolle Regulierung des Arzneimittelmarktes sollte auf jeden Fall zulassen, dass neue Wirkstoffe ihre Leistungsfähigkeit in der praktischen Anwendung unter Beweis stellen können – unter der Voraussetzung natürlich, dass sie die bisher üblichen Prüfungen auf Wirksamkeit und Sicherheit bestanden haben.
5. Eine Frühbewertung sollte daher zum einen höchsten methodischen Standards genügen und zum anderen vermeiden, durch unangemessene Prüfungen potenziell brauchbare Arzneimittel auszugrenzen. Dies wird umso weniger der Fall sein, als Bewertungen nicht nur von einer einzigen Stelle getroffen werden.
6. Die Bewertung sollte zu einem festgesetzten Zeitpunkt wiederholt werden und die Frühbewertung mit einem Verweis darauf versehen werden, bis zu welchem Zeitpunkt sie als gültig anzusehen ist, wie es so wie es für evidenzbasierte Leitlinien Standard ist.
7. Als Signal an die pharmazeutische Industrie – sowohl des In- als auch des Auslands – sollten auch in Zukunft die Ergebnisse einer Innovationstätigkeit belohnt wer-

den. Eine Forcierung des Wettbewerbs unter den Krankenkassen könnte einen maßgeblichen Beitrag dazu leisten, die kollektive Suche nach effektiven Arzneimitteln und damit den Innovationswettbewerb zu verbessern.

### Literatur

- ABDATA (2005) ABDA-Artikelstamm. ABDATA-Pharmadaten-Service, Eschborn / Taunus
- Bastian H, Bender R, Kaiser T, Kirchner H, Lange S, Müller de Cornejo G, Noelle G, Sawicki PT (2005) Methoden. <http://www.iqwig.de/index.download.a356e36e87a02176d2305a0c513dc52a.pdf> (30.05.2006)
- BMG (2006) Die Eckpunkte der Gesundheitsreform. (04.06.2007) URL: [http://www.die-gesundheitsreform.de/gesundheitspolitik/pdf/eckpunkte\\_gesundheitsreform\\_2006.pdf](http://www.die-gesundheitsreform.de/gesundheitspolitik/pdf/eckpunkte_gesundheitsreform_2006.pdf)
- Fricke U, Günther J, Zawinell A (2005) Anatomisch-therapeutisch-chemische Klassifikation mit Tagesdosen für den deutschen Arzneimittelmarkt. Hrsg. Wissenschaftliches Institut AOK (WIdO), Bonn
- Fricke U, Klaus W (1982) Kritische Wertung der neuen Arzneistoffe. *Offizinpharmazie* 4:6-47
- Fricke U, Klaus W (1983) Die neuen Arzneimittel. Wirkungsweise und therapeutischer Stellenwert. Eine Übersicht von April 1981 – Dezember 1982. *Offizinpharmazie* 7:6-62
- Fricke U, Klaus W (1985) Die neuen Arzneimittel. Wirkungsweise und therapeutischer Stellenwert. Eine Übersicht von Januar 1983 – Juni 1984. *Offizinpharmazie* 10:1-71
- Fricke U, Klaus W (1986) Die neuen Arzneimittel. Wirkungsweise und therapeutischer Stellenwert. Eine Übersicht von Juli 1984 – März 1985. *Die Offizin* 1:1-35
- Fricke U, Klaus W (2000 ff.) Neue Arzneimittel. Fakten und Bewertungen von 2000 bis 2003 zugelassenen Arzneimitteln. Band 11 ff. Stuttgart: Wissenschaftliche Verlagsgesellschaft
- Häussler B, Gothe H, Reschke P, Höer A, Hagenmeyer EG, Ryll A, Hempel E (2002) Analog-Wirkstoffe im Arzneimittelmarkt: Therapeutischer Nutzen und Bedeutung für die Ausgaben der Krankenversicherungen. *Schriftenreihe Strukturforchung im Gesundheitswesen*, Bd. 30. Berlin
- Kaesbach W, Schleert W (2000) Finanzierbarkeit von Innovationen im GKV-Markt. In: Klauber J, Schröder H, Selke, GW (Hrsg.) *Innovationen im Arzneimittelmarkt*. Springer, Berlin: 193-202
- Netzwerk evidenzbasierte Medizin (2006) Was ist Evidenzbasierte Medizin und was nicht? URL: <http://www.ebm-netzwerk.de/grundlagen/grundlagen/definitionen> (01.06.2006)
- NN (2006) Neue Arzneimittel 2004. *Arzneimittelbrief* 40:2
- Schmoll HJ, Höffken K, Possinger K (Hrsg.) (2006) *Kompendium Internistische Onkologie. Standards in Diagnostik und Therapie. Teil 1: Epidemiologie, Tumorbiologie, Zytostatika, Prinzipien der Tumorthherapie, Supportive Maßnahmen*. Springer Medizin Verlag, Heidelberg
- Schwabe U, Paffrath D (Hrsg.) (1987 ff.) *Arzneiverordnungs-Report 1987 ff.* Springer, Stuttgart, New York
- Stille W, Brodt Hr, Groll AH, Just-Nübling G (2005) *Antibiotika-Therapie*. Schattauer, Stuttgart