

WIE URTEILEN VERHALTENSTHERAPIE-GUTACHTER? VON DER EINSCHÄTZUNG DES FALLBERICHTS ZUR BEURTEILUNG DES ANTRAGS

HOW DO BEHAVIOR THERAPY EXPERTS JUDGE? FROM ASSESSMENT OF THE CASE REPORT TO ASSESSMENT OF THE APPLICATION

Serge K. D. Sulz¹, Rolf Meermann², Eibe-Rudolf Rey³, Michael Sadre Chirazi-Stark⁴

Zusammenfassung

29 Gutachter beurteilten fünf unausgewählte Verhaltenstherapie-Erstanträge. Zuerst schätzten sie jeden Antrag anhand von 9 Qualitätskriterien ein. Dann entschieden sie bezüglich uneingeschränkter oder eingeschränkter Befürwortung oder Nichtbefürwortung des Antrags. Es ergab sich eine Übereinstimmung von 97 % bezüglich Befürwortung – Nichtbefürwortung. Die beiden gut geeigneten Fälle wurden mit 92 % Übereinstimmung ungekürzt befürwortet. Lediglich bei den mit Mängeln behafteten geeigneten Fällen war die Übereinstimmung der Gutachter nicht so gut. 65 % der Beurteilungen führten bei ihnen zu ungekürzter Befürwortung, 20 % zu Nachfragen und 15 % zu Kürzungen. Aufgrund der Qualität eines Antrags ergibt sich, dass bei neun erfüllten Kriterien mit einer ungekürzten Befürwortung, bei acht erfüllten Kriterien mit Nachfragen und bei sieben oder sechs erfüllten Kriterien mit Kürzungen gerechnet werden muss. Weniger als sechs erfüllte Kriterien können zur Nichtbefürwortung führen. Die inhaltlichen Ausführungen der Rückmeldungen an die Antragsteller sind differenziert und vielfältig. Das kritische Feedback der Gutachter trotz ungekürzter Befürwortung ist ein in der bisherigen Diskussion nicht berücksichtigter Bestandteil des Gutachtens. Er ging nicht in Statistiken ein, obgleich er als steuerndes Instrument ebenfalls berücksichtigt werden muss.

Schlüsselwörter

Antragsverfahren – Gutachten – Verhaltenstherapie – Psychotherapie – Qualitätssicherung – Reliabilität

Summary

29 experts judged five not chosen first applications for behavior therapy. First, they assessed each application by 9 quality criteria. Then they decided regarding limited or unlimited recommendation or non recommendation of the application. The result was an agreement of 97% regarding recommendation – non recommendation. The two well suitable cases were recommended unshortened with 92% agreement. Merely with the flawed suitable cases, the agreement of the experts wasn't as good. 65% of the assessments lead to an unshortened recommendation with them, 20% to inquiries and 15% to reductions. On account of the quality of an application it follows that with nine met criteria with an unshortened recommendation, one has to reckon with inquiries regarding eight met criteria and with reductions regarding seven or six met criteria. Less than six met criteria can lead to non recommendation. The comments in content of the reportings back to the applicant are sophisticated and varied. Despite the unshortened recommendation, the critical feedback of the experts is part of the expert's opinion which hasn't been considered in the discussion up to now. It didn't enter the statistics, although it has to be considered as a controlling instrument as well.

Keywords

application procedure – assessment – behavior therapy – psychotherapy – quality assurance – reliability

Einleitung

Heute können sich viele nicht mehr daran erinnern, dass Psychotherapie früher keine von den Krankenkassen vergütete Leistung war. Vor vierzig Jahren mussten Patienten ihre Psychotherapie selbst bezahlen. Die gesetzlichen Krankenkassen haben relativ spät die psychotherapeutische Behandlung von psychosomatischen und psychischen Erkrankungen in ihren Leistungskatalog aufgenommen – und dies auch nur nach Begutachtung jedes Einzelfalls. Der Bundesausschuss der Ärzte und Krankenkassen hat zuletzt am 11. Dezember

1998 durch Neufassung der Psychotherapie-Richtlinien erneut beschlossen, dass die Kosten für psychoanalytisch begründete Psychotherapien und Verhaltenstherapie von den Krankenkassen nur dann übernommen werden, wenn vor Beginn der Behandlung ein Gutachterverfahren durchlaufen wurde. Ein in 5. Auflage erschienener Kommentar (Faber, Dahm und Kallinke, 1999) erläutert die gegenwärtig vorgeschriebenen Vorgehensweisen. Diesem Kommentar kann auch der Text der oben genannten Psychotherapie-Richtlinien entnommen werden. Probleme mit der Begutachtung wurden immer wieder diskutiert (z. B. Sulz, 1990; Linden und Dankesreiter, 1996;

¹Centrum für Integrative Psychotherapie, München

²Professor für Psychiatrie an der Medizinischen Hochschule Hannover (MHH), Psychosomatische Fachklinik Bad Pyrmont

³Professor für Psychologie am Zentralinstitut für Seelische Gesundheit Mannheim, Leiter der Abteilung Klinische Psychologie

⁴Professor für Psychiatrie an der Universität Hamburg, Asklepios Westklinikum Hamburg, Abt. für Psychiatrie und Psychotherapie

Köhlke, 1998; Palm et al., Weigeldt, 2000; 2000; Rüger, 2001). Dies führte einerseits zu Bemühungen, das Gutachterverfahren weiterzuentwickeln, u. a. durch einen Vorschlag, das Gutachterverfahren um eine Epikrise zu ergänzen mit dem Ziel, die Ergebnisqualität in diese Qualitätssicherung aufzunehmen (DÄVT, 2000), z. B. in der von Sulz (2000) vorgeschlagenen Fassung. Andererseits arbeiten Krankenkassen und Psychotherapeuten seit Jahren gemeinsam daran, ein Instrument zu entwickeln, das die Funktion des gegenwärtigen Gutachterverfahrens ersetzen und objektive Daten zur Ergebnisqualität liefern kann. Ein dritter Ansatz kam vom Deutschen Fachverband für Verhaltenstherapie (DVT). Durch ein zertifiziertes internes Qualitätsmanagement soll die Fähigkeit zugesprochen und Befugnis zur Qualitätssicherung in größerem Umfang dem Praxisinhaber übertragen werden. Dadurch wäre eine Minimierung kostspieliger externer Qualitätssicherungsmaßnahmen möglich (Sulz, 2001). Es ist erstaunlich, dass all diese Diskussionen und Bestrebungen abliefen, ohne dass reliable empirische Daten zur Qualität des Gutachterverfahrens vorgelegen hätten. Erst 2002 wurde eine Untersuchung zur Qualität der Gutachten bei psychoanalytisch begründeten Psychotherapieverfahren veröffentlicht (Rudolf et al., 2002) mit dem Ergebnis, dass eine gute Übereinstimmung vorlag. Es wurde über zwei Studien berichtet: Den Gutachtern wurde eine Checkliste mit zehn Kriterien zur Einschätzung der Qualität eines Fallberichts/Antrags vorgegeben. Diese waren 1. Krankheitswertigkeit der Störung, 2. Biografie, 3. Diagnostische Einschätzung der Persönlichkeitsstruktur, 4. Berücksichtigung somatischer Anteile, 5. Psychodynamik, 6. Erörterung der Differenzialindikation, 7. Stimmigkeit des Behandlungsverfahrens und der Behandlungsplanung, 8. Prognostische Einschätzung mit Blick auf die Behandlungskonzeption, 9. Prognostische Einschätzung angesichts des Verlaufs und 10. Wirtschaftlichkeit der Behandlungsplanung. In der ersten Studie begutachteten 48 Gutachter je zwei ausgewählte Anträge, einen Fall ohne Mängel und einen mangelhaften Fall. Kriterium 4 (Berücksichtigung somatischer Anteile) konnte von den Gutachtern anhand der zur Verfügung stehenden Information oft nicht eingeschätzt werden und wurde deshalb ausgeklammert. Der positive Fall wurde zu 100 % befürwortet, zu 83,4 % uneingeschränkt und zu 16,6 % eingeschränkt. Der mangelhafte Fall wurde zu 54,2 % abgelehnt, zu 35,4 % kam es zu einer Einschränkung oder zu Auflagen. 10,4 % der Gutachter befürworteten diesen Fall uneingeschränkt. Damit kann von einer Fehlerquote von 10 % bei problematischen Fällen ausgegangen werden und einer Fehlerquote von 0 % bei unproblematischen Fällen. Bei uneingeschränkt befürworteten Fällen waren acht Kriterien zu 82 bis 100 % erfüllt. Nur bei dem Kriterium Persönlichkeit sahen lediglich 67 % der Gutachter das Kriterium erfüllt. Bei nicht befürworteten Fällen gaben bei 7 Kriterien zwischen 0 und 4 % der Gutachter an, dass ein Kriterium erfüllt wurde, das Kriterium Persönlichkeit sahen 25 % und das Kriterium Krankheitswertigkeit 81 % als erfüllt an. Bei eingeschränkter Befürwortung lagen diese Werte näher an den nicht befürworteten Gutachten (bei 6 Kriterien zwischen 20 % und 40 %, einmal 56 % (Wirtschaftlichkeit) und einmal bei 88 % (Krankheitswertigkeit).

In einer zweiten Studie wurde untersucht, ob und wie die Gutachter die definierten Kriterien in ihrer Gutachtertätigkeit einsetzen. 40 Gutachter reichten 452 Routinegutachten ein. Bei der Erfüllung von neun Kriterien der Qualität des Fallberichts/Antrags lag die Übereinstimmung zwischen Einzelkriterium und Gesamteinschätzung der Begutachtung zwischen $r=0.30$ und $r=0.71$. Die Korrelation der Gesamteinschätzung der Begutachtung (Befürwortung, Kürzung, Nachfragen) mit der Einschätzung der Krankheitswertigkeit war niedrig (0.30), mit Patientenmotivation (0.46), Persönlichkeit (0.49) und bei Differenzialindikation (0.51) mäßig. Fazit ist hohe Übereinstimmung bei Befürwortung versus Nichtbefürwortung und gute Übereinstimmung in der Kriterieneinschätzung bei mängelfreien Fällen, hingegen größere Varianz bei mangelbehafteten Fällen. Diese Ergebnisse bezogen sich auf psychoanalytisch begründete Therapieverfahren und lassen sich nicht auf Verhaltenstherapien übertragen. Um auch für die Verhaltenstherapie empirische Daten zur Reliabilität des Gutachterverfahrens zu erhalten, war es nahe liegend, diese Studie für Verhaltenstherapie-Gutachten zu wiederholen. Allerdings waren die Kriterien nicht ohne Veränderung übernehmbar, da in der Verhaltenstherapie doch andere Akzente gesetzt werden.

Methoden

a) Fragestellung

In der vorliegenden Studie geht es darum, die Reliabilität der Beurteilungen von Verhaltenstherapie-Gutachtern zu untersuchen. Alle Gutachter sollen die gleichen Anträge bearbeiten, so dass ihre Beurteilungen verglichen werden können. Zuerst werden sie gebeten, die Fallberichte anhand einer Checkliste von neun Qualitätskriterien einzuschätzen. Anschließend sollen sie ihre gutachterliche Würdigung des Antrags hinsichtlich Befürwortung, Kürzung des beantragten Stundenkontingents und Nichtbefürwortung festhalten und begründen. Diese Vorgaben ermöglichen die Beantwortung folgender Fragen: Wie hoch ist die Übereinstimmung der Gutachter bei

- a) den neun einzelnen Gütekriterien
- b) Befürwortung – Nichtbefürwortung
- c) Kürzungen des Stundenkontingents
- d) den Nachfragen, die vom Antragsteller noch vor einer Befürwortung beantwortet werden müssen
- e) den Begründungen für Befürwortung – Nichtbefürwortung.

b) Datenmaterial

Die fünf begutachteten Fälle seien hier kurz skizziert, ohne die Schwachpunkte der Fallberichte explizit deutlich zu machen:

Fall A

Beschwerden und Anlass: Ein knapp vierzigjähriger Mann, dessen Frau unter einer Zwangserkrankung leidet. Beide sind seit längerer Zeit in Paartherapie. Aus dieser entstand die Erkenntnis, dass es nicht weitergehen könne, solange er keine eigene Psychotherapie mache. Er habe lange Zeit seiner Frau geholfen, dass das Haus so sauber bleibe, wie es ihr Zwang verlange. Nun aber wehre er sich vehement gegen die Tyrannei, was zu eskalierenden Auseinandersetzungen geführt habe.

Sozialdaten: verheiratet, keine Kinder, angestellt im EDV-Bereich, keine Arbeitsunfähigkeit in den letzten zwei Jahren
Befund: emotionale Beeinträchtigung, aber keine psychischen oder psychosomatischen Symptome

Diagnose: Anpassungsstörung mit überwiegend emotionalen Anteilen

Verhaltensanalyse: Es wird ein Partnerschaftskonflikt beschrieben mit zunächst dependentem und jetzt auflehndem Verhalten des Patienten.

Ziele: Selbstbehauptung des Patienten seiner Frau gegenüber

Therapieplan: Aufbau von kompetentem Sozialverhalten mit Fordern und Nein-sagen-Können

Fall B

Beschwerden und Anlass: Ein etwas über 40-jähriger Ingenieur mit sozialphobischen Symptomen, die so ausgeprägt waren, dass er nicht mehr in der Lage war, zur Arbeit zu gehen. Die Beschwerden bestehen seit einem Jahr, als seine Partnerin ihn verließ und ein Konflikt am Arbeitsplatz ihn stark belastete.

Sozialdaten: ledig, ein Kind, angestellt, vier Wochen arbeitsunfähig in den letzten zwei Jahren

Befund: Angst, sich in der Öffentlichkeit bloßzustellen, Angst vor fremden Menschen, vor allem in Gruppen. Angst, unter Menschen zu gehen. Dabei treten u. a. Atemnot, Pulsrasen, Mundtrockenheit, Schwindel, Zittern auf, außerdem Spannungskopfschmerzen

Diagnose: soziale Phobie, Spannungskopfschmerzen, ängstlich-vermeidende Persönlichkeitsstörung

Verhaltensanalyse: Symptomauslösung durch anhaltende berufliche Belastung und die Trennung. Persönliche Disposition durch die selbstunsichere Persönlichkeitsstörung mit ängstlicher Erwartungshaltung in sozialen Situationen. Symptomaufrechterhaltend wirkt die Vermeidung von Ablehnung und Ausgrenzung.

Ziele: Handlungsorientierung fördern, Problemlösestrategien entwickeln, Überlebensregel relativieren, Lernen, mit Angst umzugehen, Reduktion des Vermeidungsverhaltens

Therapieplan: Stressbewältigungs- und Problemlösestrategien, empirische Hypothesenprüfung, Angstexposition, ATP-Strategien

Fall C

Beschwerden und Anlass: Ein junger Patient mit schizophrener Minussymptomatik bei Residualsyndrom. Allmähliche Veränderung des Denkens, der Gefühle und das Sozialverhaltens seit einigen Jahren, er fühle sich beobachtet.

Sozialdaten: ledig, zahlreiche Krankenschreibungen in den letzten 2 Jahren, derzeit berufliche Reha-Maßnahme

Befund: Gestik, Mimik reduziert, Konzentrationsstörungen, affektflach, misstrauisch, Beziehungsideen, Antriebsmangel

Diagnose: schizophres Residuum

Verhaltensanalyse: Familienatmosphäre mit "Expressed Emotions". Keine externen krankheitsauslösenden und -aufrechterhaltenden Bedingungen eruierbar.

Ziele: Verbesserung der Stimmung, der Konzentrations- und Entspannungsfähigkeit, der sozialen Fertigkeiten, der fa-

miliären Atmosphäre, Rückfallprophylaxe

Therapieplan: Vermittlung von Problemlösetechniken, Aktivitätenplanung, Entspannungstraining, Kognitives Training, Kommunikationstraining, Früherkennungsliste von Rückfallerscheinungen

Fall D

Beschwerden und Anlass: Zwanzigjährige Patientin mit Angst in Menschenmenge und engen Räumen. Erste Panikattacke nach dem Abitur.

Sozialdaten: ledig, Musikstudentin, keine längeren Krankheitszeiten

Befund: Angst in beengten Räumen und Situationen, ausgeprägtes Vermeidungsverhalten, Panikattacken, Schlafstörungen

Diagnose: Agoraphobie mit Panikattacken

Verhaltensanalyse: Symptomauslösend ist der Auszug aus dem Elternhaus bei Ablösungsproblemen von der Mutter. Durch die Symptomatik vermeidet sie, sich den Anforderungen der Berufsausbildung zu stellen.

Ziele: psychologisches Krankheitsmodell, kognitive Angstbewältigung, Reduktion der Schlafstörungen

Therapieplan: Bedingungsanalyse, Angstbewältigungstraining, Entspannungstraining

Fall E

Beschwerden und Anlass: 50-jähriger Facharbeiter mit seit Jahren rezidivierenden depressiven Episoden, kam jetzt nach einer Trennung zur Therapie.

Sozialdaten: geschieden, Facharbeiter, in den letzten 2 Jahren 5-mal arbeitsunfähig für insgesamt 6 Monate, 5 Monate stationäre psychiatrische Behandlung.

Befund: niedergeschlagen, hoffnungslos, Grübeln, Antriebsstörungen, Schlafstörungen, verlangsamt, mnestisch und kognitive Funktionen subjektiv eingeschränkt, formales Denken, leicht eingeeengt, Schwingungsfähigkeit etwas beeinträchtigt, distanziert sich von anfänglichen Suizidgedanken. Kein psychotisches Erleben.

Diagnose: depressive Anpassungsstörung bei Trennungskonflikt

Verhaltensanalyse: Trennung von der Ehefrau wirkte als Verstärkerverlust. Wegen seiner dependenten Haltung (durch Modelllernen vom Vater übernommen) fehlt die Fähigkeit, sein Selbstwertgefühl ohne diese Beziehung aufrechtzuerhalten.

Ziele/Therapieplan: Trennung emotional und kognitiv verarbeiten, wieder Zugang zu Verstärkung erlangen durch Aufbau positiver verstärkender Aktivitäten, Aufbau von Sozialkontakten, Genusstraining

Parallel psychiatrische Behandlung mit antidepressiver Medikation.

c) Datenerhebung

Die Gutachter erhielten folgende Instruktion:

"Sie erhalten 5 Anträge mit der Beantragung von 45 Stunden Langzeit-Verhaltenstherapie (Ziffer 882, Einzelsitzungen), ohne Mitbehandlung der Bezugspersonen.

Von 100 Erstanträgen wurden per Los 5 Anträge gezogen.

Es erfolgte keine inhaltliche oder formale Selektion. Sie erhalten alle verfügbaren Informationen. Alle Angaben, die auf Formularen verfügbar waren, wurden auf einem Sozial- oder Patientendatenblatt zusammengefasst. Weitere Informationen lagen nicht vor.

Das Formblatt Stellungnahme Erstantrag Verhaltenstherapie enthält 9 Kriterien. Sie haben 4 Antwortalternativen je Kriterium:

- () nein – eindeutig nicht
- () eher nicht
- () ja – nicht ganz eindeutig
- () ja – eindeutig

Wenn Sie "nein – eindeutig nicht" ankreuzen, so ist dies gemäß Richtlinien ein Grund, den Antrag nicht zu befürworten. Inwiefern sich die Bewertung "eher nicht" mit einer Befürwortung vereinbaren lässt, ist bei einigen wenigen Kriterien abzuwägen, bei den anderen Kriterien ist dies ebenfalls ein Grund für eine Nichtbefürwortung.

Bitte senden Sie uns das ausgefüllte Formblatt zur Stellungnahme zurück.

Falls Sie einen Gutachter-Code erfinden und angeben, so können wir Ihnen Ihr Vergleichsprofil zukommen lassen bzw. Sie können dies aus den Ihnen zugesandten Auswertungsdiagrammen herauslesen, ohne dass Sie uns mitteilen müssen, welches Ihr Code ist – wie Sie es wünschen. Darüber hinaus hätten wir dann die Möglichkeit, die Intrarater-Varianz mit zu berücksichtigen.

Eine Frage ist: Soll ich so bewerten wie immer oder so wie es mir nunmehr anhand der Checkliste und der besonderen Aufmerksamkeit innerhalb der vorliegenden Studie möglich ist? Antwort: Gehen Sie vor wie immer. Entscheiden Sie so wie immer über Befürwortung oder Nichtbefürwortung."

Die neun Kriterien sind eine Operationalisierung der Psychotherapie-Richtlinien und des von Faber, Dahm und Kallinke verfassten Kommentars (1999):

1. Bei den Beschwerden handelt es sich um eine behandlungsbedürftige Krankheit im Sinne der Psychotherapie-Richtlinien.
2. Für die vorliegende Erkrankung besteht eine Indikation für Verhaltenstherapie.

3. Die Anamnese beschreibt die für das Fallverständnis notwendigen Aspekte der Lebens- und Krankengeschichte.
4. Der Befund entspricht der gestellten ICD-10-Diagnose.
5. Die Verhaltensanalyse erklärt Entstehung und Aufrechterhaltung der Störung.
6. Die formulierten Therapieziele werden der Störung gerecht, sind in Zielrichtung und Ausmaß des intendierten Änderungsprozesses angemessen und sind nicht anderweitig (z.B. psychosomat. Grundversorgung) erreichbar.
7. Der Behandlungsplan beschreibt eine Verhaltenstherapie und ist erfolgversprechend.
8. Die Prognose reicht für eine Befürwortung des Antrags aus.
9. Die geplante Behandlung ist notwendig und wirtschaftlich und umfasst das Notwendige (evtl. incl. Pharmako-, Somato- und/oder Soziotherapie).

Sie stimmen inhaltlich, aber nicht in der Formulierung mit den Kriterien für tiefenpsychologisch orientierte Gutachten überein, die in der Studie von Rudolf et al. (2002) angewandt wurden. Für die Einschätzungen des Grades der Erfüllung dieser Kriterien wurde die erwähnte vierstufige Skala verwendet, die den Gutachter zwingt, sich bei Zweifelsfällen für eine der beiden Seiten zu entscheiden. Dadurch war die Möglichkeit gegeben, die Skala nachträglich zu dichotomisieren (erfüllt versus nicht erfüllt): nein – eindeutig nicht; eher nicht erfüllt; ja – nicht ganz eindeutig erfüllt und ja – eindeutig erfüllt.

Ihre abschließende Stellungnahme konnten die Gutachter auf folgende Weise geben:

- () Befürwortung ohne Stundenkürzung
 - () ohne Bedenken
 - () mit Bedenken
- () Befürwortung mit Kürzung des Stundenkontingents um von ... Sitzungen
- () Keine Befürwortung
- () Ich kann erst über eine Befürwortung entscheiden, wenn folgende offene Fragen vom Antragsteller beantwortet wurden: ...

Kürzung und Nichtbefürwortung mussten in einem Text beliebiger Länge begründet und die eventuellen Fragen formuliert werden.

Kriterium zur Einschätzung eines Fallberichts	Fall A	Fall B	Fall C	Fall D	Fall E	Korrelation mit Stellungnahme
1. Beschwerden behandlungsbedürftig?	1,4	4,0	3,9	4,0	4,0	0,70***
2. Indikation für Verhaltenstherapie?	1,6	4,0	3,8	3,9	4,0	0,72***
3. Anamnese ergibt Fallverständnis?	2,1	3,7	3,5	3,2	3,6	0,66***
4. Befund entspricht ICD-10-Diagnose?	1,3	3,7	3,2	3,3	3,7	0,68***
5. Verhaltensanalyse erklärt Störung?	1,3	3,6	3,5	3,0	3,6	0,73***
6. Therapieziele werden Störung gerecht?	1,7	3,6	3,5	3,5	3,5	0,75***
7. Therapieplan beschreibt erfolgversprechende VT?	1,5	3,6	3,6	3,5	3,6	0,78***
8. Prognose ausreichend?	1,7	3,6	3,5	3,5	3,7	0,74***
9. Therapie notwendig, wirtschaftlich, umfasst Notwendiges?	1,2	3,7	3,5	3,6	3,7	0,77***

Tabelle 1: Einschätzung der Kriterienerfüllung auf einer vierstufigen Skala

Letzte Spalte: Spearman-Korrelation mit der Stellungnahme des Gutachters (1= Nichtbefürwortung, 2 = Kürzung, 3 = Fragen, 4 = Befürwortung mit Bedenken, 5 = Befürwortung ohne Bedenken). Zweiseitige Signifikanzprüfung bei N = 149, *** = sign. auf dem 1-Promille-Niveau)

d) Stichprobe

Alle 42 psychologischen und ärztlichen Gutachter wurden eingeladen, an der Studie teilzunehmen. Dreißig (71 %) Gutachter sandten die Gutachten zurück. Einer dieser Gutachter hatte Fall A nicht bearbeitet. Dadurch ergaben sich für Fall A 29 und für die übrigen Fälle je 30 Gutachten, insgesamt lagen also 149 bzw. 150 Gutachten vor. Die Gutachter bearbeiteten die Anträge zwischen September und Dezember 2002. Die Datenerhebung war anonym. Es wurde ein frankiertes Rückcouvert beigelegt, so dass auch der Poststempel die Herkunft der Gutachten nicht zeigen konnte. Gutachter, die ein Feedback für ihre Antworten haben wollten, konnten einen Code angeben. Dadurch können sie ihre Antworten mit dem Durchschnitt der anderen Gutachter vergleichen.

e) Auswertung

Die Auswertung wurde mit SPSS durchgeführt. Kontingenzanalysen wurden mit Chi-Quadrat auf Signifikanz geprüft, es wurden Rangkorrelationen nach Spearman berechnet. Mittelwertvergleiche wurden dem nonparametrischen Mann-Whitney-U-Test für nicht abhängige Stichproben unterzogen. Alle Signifikanzprüfungen wurden als zweiseitige Tests angelegt, und es wurde eine Alpha-Adjustierung vorgenommen.

Ergebnisse

Die Kriterien wurden, wie in Tabelle 1 gezeigt, von den Gutachtern bei den einzelnen Fällen geratet.

Insgesamt lag mit Ausnahme des ungeeigneten Falles (A) eine Kriterienerfüllung vor (Werte ab 3). Die Kriterien Behandlungsbedürftigkeit und Indikation für Verhaltenstherapie waren bei den Fällen B bis E durchschnittlich eindeutig erfüllt (Werte mindestens 3,8). Diese beiden Kriterien differenzieren demnach am besten zwischen dem abgelehnten und den befürworteten Fällen. Bei allen anderen Kriterien ergab sich ein Mittelwert, der 24-mal etwa zwischen 3 (nicht eindeutig erfüllt) und 4 (eindeutig erfüllt) und 4-mal nahe bei 3 lag. Diese Kriterien waren durchschnittlich bei den Fällen B bis E also erfüllt, aber nicht immer eindeutig. So wurde moniert, dass bei Fall C und D die Übereinstimmung von Befund und ICD-10-Diagnose und bei Fall D außerdem der Beitrag der Anamnese zum Fallverständnis und die Erklärung der Störung durch die Verhaltensanalyse als zwar ausreichend, aber nicht ganz eindeutig erklärt oder dargelegt waren. Die letzte Spalte von Tabelle 1 zeigt die Korrelationen zwischen der durchschnittlichen Erfüllung der Kriterien und der Stellungnahme der Gutachter. Der Zusammenhang ist mit Korrelationen zwischen 0,66 und 0,78 gut.

Zahl erfüllter Kriterien	Fall A	Fall B	Fall C	Fall D	Fall E
0	13				
1	8				
2	4				
3	3				
4	1		1	1	
5					
6				2	
7			1	4	3
8		5	4	6	2
9		25	24	17	25
gesamt	29	30	30	30	30
Mittelwert	1,0	8,8	8,6	8,1	8,7
Standardabweich.	1,16	0,38	1,0	1,23	0,64

Tabelle 2: Die begutachteten Fälle und Zahl der bei jedem Fall erfüllten Kriterien (zweistufige Skala: Die Werte 1 und 2 aus Tabelle 1 werden hier zu "nicht erfüllt" und die Werte 3 und 4 zu "erfüllt")

Wir können die vierstufige Skalierung der Kriterien auf eine zweistufige Skalierung reduzieren: "eindeutig nicht" und "eher nicht" wird zu "nicht erfüllt" und "eindeutig ja" und "nicht ganz eindeutig ja" werden zu erfüllt zusammengefasst. Dadurch kann die Frage beantwortet werden, wie viele Gutachter bei einem einzelnen Fall ein bestimmtes Kriterium als erfüllt bzw. nicht erfüllt betrachteten. Bei Fall A sahen 13 Gutachter kein Kriterium als erfüllt, acht Gutachter nur ein Kriterium und insgesamt sahen 29 Gutachter 4 oder weniger, d. h. weniger als die Hälfte der Kriterien als erfüllt an. Nur bei Fall C und D sah jeweils ein Gutachter weniger als 50 % der Kriterien erfüllt. Oder: Bei den Fällen B bis E sahen 26 von 30 Gutachter mindestens 7 Kriterien als erfüllt an. Die Fälle B, C und E unterscheiden sich am wenigsten. Bei ihnen haben 25, 24 und 25 Gutachter alle neun Kriterien als erfüllt gesehen. Damit bleibt neben Fall A als weiterer problematischer Fall der Fall D, bei dem nur 17 von 30 Gutachtern alle neun Kriterien erfüllt sahen. Dies ist zugleich derjenige Fallbericht, bei dem die Ratings der Gutachter am höchsten streuten.

Die Stellungnahmen der Gutachter

Es wurden 149 Stellungnahmen von 30 Gutachtern abgegeben (Tab. 3). Davon waren 73 Befürwortungen ohne Bedenken, 22 mit Bedenken, 11 mit Kürzungen der Stundenzahl um 15 bis 20 Sitzungen, und 15-mal wurde der Antragsteller um die Beantwortung offener Fragen gebeten. Schließlich wurden 28 Nichtbefürwortungen ausgesprochen.

	Häufigkeit	Gültige Prozente	Fall A	Fall B	Fall C	Fall D	Fall E
Nichtbefürwortung (1)	28	18,8	28				
Kürzung (2)	11	7,4			2	7	2
Offene Fragen (3)	15	10,1		2	8	4	1
Befürwortung mit Bedenken (4)	22	14,8	1	10	2	3	6
Befürwortung ohne Bedenken (5)	73	49,0		18	18	16	21
Gesamt	149	100,0	29	30	30	30	30

Tabelle 3: Die Beurteilungen der Gutachter

Wie verteilen sich diese Stellungnahmen auf die Fälle? Fall A wird von 28 Gutachtern abgelehnt. Fall B wird von ebenso vielen Gutachtern befürwortet, 18-mal ohne und 10-mal mit Bedenken. Fall C wird 18-mal ohne Bedenken befürwortet und 8-mal wird vom Gutachter nachgefragt. Fall D wird 16-mal ohne Bedenken befürwortet, und 7-mal erfolgen Kürzungen des beantragten Stundenkontingents. Fall E wird 21-mal ohne Bedenken und 6-mal mit Bedenken befürwortet. Drücken wir Tabelle 3 in Prozenten aus, so ergibt sich Folgendes: Der ungeeignete Fall (A) wurde zu 96,6 % nicht befürwortet, die gut geeigneten Fälle (B und E) wurden zu 91,7 % befürwortet und die geeigneten Fälle mit Mängeln (C und D) wurden zu 65 % befürwortet, zu 15 % gekürzt und zu 20 % resultierten Nachfragen.

Der Zusammenhang zwischen Kriterienerfüllung und Stellungnahme des Gutachters

Lässt sich aus dem Rating der Kriterienerfüllung die Stellungnahme der Gutachter vorhersagen? Wir können die Summe über alle neun Kriterien je Fall bilden und diese mit der gestuften Stellungnahme korrelieren. Eine Kontingenzanalyse ergibt ein Chi-Quadrat, das auf dem Ein-Promille-Niveau signifikant ist. Der Kontingenzkoeffizient ist 0.81. Geht man nicht vom Nominalskalens- sondern vom Ordinalskalenniveau aus, so ergibt sich eine Spearman-Korrelation von 0.84 zwischen Kriterienerfüllung (vierstufige Skala) und Stellungnahme.

Weniger als vier erfüllte Kriterien führten mit einer Ausnahme zur Nichtbefürwortung. Neun erfüllte Kriterien führten in 90 % der Fälle zur ungekürzten Befürwortung. Bereits das Nicht-erfüllt-Sein von einem Kriterium reduzierte die relative Häufigkeit ungekürzter Befürwortung auf 56 %. Bei sieben erfüllten Kriterien wurden nur 50 % ungekürzt befürwortet, und bei sechs erfüllten Kriterien gab es keine ungekürzten Befürwortungen mehr. Dass auch neun erfüllte Kriterien zu Kürzung und Nachfragen führten, lag daran, dass zwei Gutachter die ursprüngliche Differenzierung in eindeutig (4) und nicht ganz eindeutig (3) erfülltem Kriterium als Anlass zur Kürzung nahmen. Sie mochten eine ungekürzte Befürwortung ebenso wie sechs weitere Gutachter, die die Beantwortung

offener Fragen verlangten, nur dann geben, wenn ein Kriterium wirklich eindeutig erfüllt war. Für sie war nicht die Zahl erfüllter Kriterien entscheidend, sondern die Art des Mangels und die fachliche Abwägung seiner Auswirkungen auf die beantragte Therapie.

Welche Begründungen der Nichtbefürwortung gaben die Gutachter an?

Da die Zahl der nicht erfüllten Kriterien kein explizites Argument der Gutachter ist, müssen deren inhaltliche Begründungen betrachtet werden. Zweiundzwanzig Gutachter gaben an, dass es sich um keine behandlungsbedürftige Erkrankung im Sinne der Richtlinien handelt. Bei vieren fand sich eine Aufzählung anderer Mängel, und zwei gaben keine Begründung (siehe Tab. 4).

Für Fall A liegen 29 Gutachten vor. Wenn wir davon ausgehen, dass der wesentlichste Mangel von Fall A war, dass es sich um keine behandlungsbedürftige Erkrankung im Sinne der Richtlinien handelte, so haben 96,6 % der Gutachter übereinstimmend eine nicht befürwortende Stellungnahme abgegeben. 75,9 % der Gutachter haben ihre Nichtbefürwortung übereinstimmend begründet.

Wie begründeten die Gutachter die Kürzungen?

Fall C:

- Bericht über Vorbehandlung nachreichen. Medikamentöse Compliance nicht geklärt.
- Psychiatr. Befundbericht bzw. Entlassungsbericht der stationären psychiatr. Behandlung fehlt. Zwischen Residuum und anderen Verhaltensproblemen wird nicht differenziert. Ziele nicht eindeutig. Neuropsycholog. Interventionen unklar.

Fall D:

- Ich kann die Indikation für eine Langzeittherapie nicht erkennen.
- Ich habe Zweifel bezüglich "schizotyp".
- Die Verhaltensanalyse unzureichend, das "Schizotype" ist nicht nachvollziehbar, wohl eher soziophobisch gemeint.
- Krankheits- und Lebensgeschichte dürftig. Verhaltens-

Häufigkeit	Begründungstext
22	Keine behandlungsbedürftige Erkrankung im Sinne der Richtlinien
1	Beschreibung der aktuellen Symptomatik ist unklar, die Anamnese ist unvollständig, die Lerngeschichte und Bedingungsanalyse der Störung ist unzureichend, die Diagnose ist falsch, Behandlungskonzept ist oberflächlich und nichts sagend
1	Verhaltensanalyse entspricht nicht den geforderten Minimalkriterien. Diagnose auf der Grundlage der Darstellung sehr vage
1	Keine Übereinstimmung von Befund und Diagnose, Therapieziele nicht zu unterstützen, z. B. Therapie bei anderer Therapeutin, unzureichender Konsiliarbericht
1	Verhaltensanalyse ist zu allgemein, Behandlungsplan unkonkret, Prognose angesichts der ungeklärten Partnersituation fraglich
2	Keine Begründung angegeben

Tabelle 4: Begründungen der Nichtbefürwortung (28 Nichtbefürwortungen von 29 Gutachtern)

analyse greift die relevanten biograf. Faktoren nicht auf. Dadurch keine Makroanalyse, die zu einem spezifischen Therapieplan führen würde.

- Indikation wegen Angststörungen gegeben. "Schizotyp" und "psychot." bleiben unklar. Deshalb Klärung und psychiatr. Befund nötig.
- Krankheitsanamnese ungenügend, Therapieziele zu global. Fehlende Prognose.
- Diagn. Klärung erforderlich, Befund beschreibt eher anamnest. Komponenten. Warum erfolgt keine Konfrontationsbehandlung? Evtl. SSRI-Hemmer anwenden?

Fall E:

- Für die beschriebene Anpassungsstörung müsste eine KZT ausreichend sein. Auf die seit Jahren bestehenden depressiven Episoden (siehe Konsiliarbericht) wird nicht eingegangen.
- Keine Begründung

Kürzungsvorschläge der Gutachter an die Krankenkasse erfolgen im Vergleich zu Nachfragen oft, weil die Notwendigkeit eines baldigen Therapiebeginns gesehen wird und ein Briefverkehr durch Nachfragen zu einer zu großen Verzögerung führen würde. Es ist deshalb nicht unbedingt als strengere Stellungnahme zu sehen. Ein weiterer Grund für Kürzungen sind Unklarheiten bezüglich der Prognose oder Differenzialindikation (oft im Zusammenhang mit der Fallkonzeption und dem Therapieplan), die nach 15 bis 25 Sitzungen besser beurteilt werden kann. In diesem Falle ist die Befürwortung des gekürzten Stundenkontingents eine Probetherapie. Tatsächlich beziehen sich nur zwei Begründungen ausschließlich auf diagnostische Aspekte, deren Klärung keiner Probetherapie bedarf, sondern durch Nachfragen aufgeklärt werden könnten. Damit ist die Schlüssigkeit der Kürzungen nachvollziehbar.

Welche Fragen stellten die Gutachter an die Antragsteller?

Die Fragen betreffen überwiegend die Fälle C und D. Bei einer psychotischen Erkrankung wollen Gutachter den Bericht eines Psychiaters, andere den Entlassungsbericht der Psychiatrischen Klinik. Sie wünschen einen vollständigeren psychischen Befundbericht, der nicht offen lässt, ob bestimmte Symptombereiche überhaupt untersucht wurden. Und sie wünschen eine bessere Übereinstimmung von ICD-10-Diagnose und psychischem Befund. Vereinzelt wurden Spezifizierungen der Verhaltensanalyse, der Zielanalyse und des Therapieplans (besonders Individualisierung) oder der Indikationsbegründung für eine Langzeittherapie angefordert:

- Psychiatr. Konsiliarbericht einreichen, wenn dieser überzeugt, Befürwortung (4x C, 3x D)
- Kopie Entlassungsbericht der psychiatr. Klinik (2x C)
- Welche Medikation besteht? (2x C)
- Psych. Befund ist unvollständig, wichtige Symptombereiche bleiben unerwähnt (3x C, 3x D)
- Verhaltensanalyse nicht schlüssig (2x C, 3x D)
- Diagnose unklar, stimmt z. B. nicht mit Anamnese und Befund und Bedingungsmodell überein (4x C, 2x D)
- Spezifizierung der Ziele: Was ist erreichbar? (B)
- Therapieplan unvollständig bzw. unangemessen (C, D).

- Notwendigkeit Langzeittherapie nicht klar (2x C)
- Individualisierter Behandlungsplan erbeten (2x B, C, D).

Welche Bedenken äußerten die Gutachter trotz ungekürzter Befürwortung?

Befürwortende Gutachter bringen ganz ähnliche Kritik vor wie kürzende und fragende Gutachter. Sie geben sich jedoch mit der Kritikäußerung und der Bitte um Beachtung des Feedbacks beim weiteren Vorgehen zufrieden. Dieses recht kritische Feedback ist ein wichtiger Bestandteil des Dialogs zwischen Gutachter und Antragsteller. Da dieser Aspekt in der bisherigen Diskussion keine Berücksichtigung fand, sollen die Bedenken nur wenig gekürzt wiedergegeben werden:

Fall A:

- Unklare Behandlungsmotivation, fehlende aktive Veränderungsbereitschaft, diffuse Behandlungsziele.

Fall B:

- Die Entwicklung d. selbstunsicheren Persönlichkeitsstörung ist aus der Biografie und der Darstellung der lerngeschichtlichen Entwicklung nicht eindeutig nachzuvollziehen. Eine Verhaltensanalyse des Spannungskopfschmerzes fehlt, ebenso Therapieziele und Behandlungsplan.
- Der psych. Befund von Therapeut und Psychiater ist unvollständig.
- Befund unzureichend, Verhaltensanalyse unpräzise, wechselt Verhaltensexzesse und -defizite, Diagnose nur teilweise nachvollziehbar, Therapieplan wenig präzise.
- Diagnose F60.6 kann nicht nachvollzogen werden. Für die Störung 44.2 werden spezifische Interventionen vermisst. Insgesamt unklarer Behandlungsplan.
- Wie werden die aufgelisteten Interventionsmethoden in einen Gesamtbehandlungsplan eingebettet?
- Funktionalität Spannungskopfschmerz nicht dargestellt. Behandlungsplan zu schematisch und undifferenziert: Individualisieren! Spannungskopfschmerz nicht im Therapieplan enthalten.
- Behandlungsplan zu schematisch, entwickelt nicht hinreichend eine individualisierte Behandlungsstrategie. Da aber die genannten Methoden zielführend sind und der Antrag ansonsten kompetent ist, erfolgt Befürwortung.
- Behandlungsplan zu wenig individualisiert.
- In Verhaltensanalyse fehlt der Zusammenhang mit der Lebensgeschichte. Aufrechterhaltende Bedingungen sind nicht sauber aufgeführt. Therapieplan zu schematisch.
- Fehlende Prognose.

Fall C:

- Biograf. fehlt schulisch-beruflicher Werdegang, Verhaltensanalyse berücksichtigt nicht Drogen als auslösenden Faktor, Status des Drogenkonsums fehlt. Engmaschige psychiatr. Begleitbehandlung wichtig. Für Fortführungsbericht bitte psychiatr. Stellungnahme.

Fall D:

- Ich kann die Indikation für eine Langzeittherapie nicht erkennen.
- Psychiatrische Abklärung unbedingt empfohlen.
- Behandlungsplan ist zu schematisch (Fehlen der übergeordneten und individualisierten Strategie). Gegebenenfalls

wären mehrstündige Expositionen zu planen. Da Gesamteindruck aber gut, erfolgt Befürwortung.

- Behandlungsplan zu wenig individualisiert. Nervenärztliche Mitbehandlung/Abklärung DD.

Fall E:

- Die dependenten Persönlichkeitszüge werden im psychischen Befund, Therapiezielen und Behandlungsplan nicht ausreichend berücksichtigt.
- Diagnostisch ggf. die Kriterien einer depr. Episode erfüllt. Behandlungsplan zu schematisch, dennoch zielführend. Da Gesamtinhalt des Antrags überzeugt, erfolgt Befürwortung.

Wir können uns die notwendigen Implikationen der von den Gutachtern geäußerten Bedenken am Beispiel der Prognose deutlich machen. Der Gutachter kann aufgrund seiner Erfahrung die Prognose aus einem Fallbericht recht zuverlässig herauslesen, auch wenn der Antragsteller diese nicht explizit formuliert. Deshalb ist die fehlende Angabe des Antragstellers zur Prognose zwar ein zu kritisierender Mangel. Aber der Gutachter kann trotz dieses Mangels entscheiden, ob die Prognose ausreichend ist. Dann ist weder ein Nachfragen noch eine Kürzung zwingend. Nur wenn der Bericht so unklar bleibt, dass die Prognose offen bleibt, muss nachgefragt werden oder, wenn die Prognose erst durch eine Probetherapie zuverlässig genug erkennbar sein wird, ein gekürztes Stundenkontingent befürwortet werden. Man kann hingegen auch den edukativen Standpunkt einnehmen, dass der Antragsteller sich selbst Gedanken zur Prognose machen muss und diese auch explizit dem Gutachter mitzuteilen hat.

Was eine ausreichende Individualisierung des Therapieplans ist, scheint unter den Gutachtern ebenfalls nicht sehr gut übereinzustimmen. Dies ist die insgesamt am häufigsten geäußerte Kritik (9-mal von 149 befürwortenden Stellungnahmen), die nie zu einer Nichtbefürwortung oder Kürzung führte, viermal zu Nachfragen und fünfmal zu kritischem Feedback bei Befürwortung.

Gibt es persönliche Stile der Gutachter?

Angesichts der Streuung der gutachterlichen Stellungnahmen können wir weitere Untersuchungen anstellen. Worin bestehen die Unterschiede der Beurteilungen? Wir müssen beden-

Stellungnahme	Strenge Gutachter	Nicht strenge Gutachter	Signifikanz
Nichtbefürwortung der Therapie	15	13	n.s.
Kürzung der Stunden	10	1	***
Noch Fragen vor Entscheidung	15	0	***
Bedenken bei der Befürwortung	9	13	n.s.
Befürwortung ohne Bedenken	30	43	n.s.

Tabelle 6: Vergleiche der Stellungnahmen von strengen und nicht strengen Gutachtern

Kontingenzanalyse. *** = Chi-Quadrat ist signifikant auf dem 1 Promille-Niveau, n.s. = Wahrscheinlichkeit von Chi-Quadrat ist größer als 0.05 (zweiseitige Prüfung)

ken, dass 13 von 30 Gutachtern vier und einer fünf Befürwortungen von fünf Anträgen aussprachen. Es muss geprüft werden, ob diese Gutachter auch die Kriterienerfüllung positiver einschätzten als die strengeren Gutachter.

Dem ist nicht so (Tab. 5). Bei keinem Kriterium ergaben sich signifikante Mittelwertunterschiede zwischen diesen beiden Gutachtergruppen. Es könnte allerdings sein, dass bei allen Fällen außer Fall D sehr ähnliche Kriterieneinschätzungen vorlagen. Um dies zu prüfen, wurde die gleiche Signifikanzprüfung nur für Fall D durchgeführt. Aber auch bei diesem Fall gab es keine signifikanten Unterschiede zwischen den beiden Gutachtergruppen. D. h., obwohl sie die Fälle ähnlich einschätzten, kam es bei den nicht strengen Gutachtern kaum zu Nachfragen und Kürzungen. Bei dem abzulehnenden Fall waren sie sich jedoch mit der anderen Gutachtergruppe einig. Es bleibt zu fragen, ob sie bei den Befürwortungen auch öfter auf die Äußerung von Bedenken verzichteten.

Auch dies ist nicht der Fall (Tab. 6). Dass Kürzungen und Fragen bei den beiden Gutachtergruppen signifikant verschieden sind, liegt an der Konstruktion der beiden Gruppen und

Kriterien zur Einschätzung der Qualität der Fallberichte	Mittelwerte		Signifikanz (Mann-Whitney U-Test)
	Strenge Gutachter	Nicht strenge Gutachter	
Beschwerden behandlungsbedürftig?	3,5	3,4	n.s.
Indikation für Verhaltenstherapie?	3,5	3,5	n.s.
Anamnese ergibt Fallverständnis?	3,3	3,2	n.s.
Befund entspricht ICD-10-Diagnose?	3,0	3,1	n.s.
Verhaltensanalyse erklärt Störung?	3,0	3,0	n.s.
Therapieziele werden Störung gerecht?	3,2	3,2	n.s.
Therapieplan beschreibt erfolgversprechende VT?	3,2	3,1	n.s.
Prognose ausreichend?	3,2	3,2	n.s.
Therapie notwendig, wirtschaftlich, umfasst Notwendiges?	3,1	3,3	n.s.

Tabelle 5: Vergleich der strengen und nicht strengen Gutachter hinsichtlich ihrer Einschätzung der Kriterienerfüllung der Fallberichte Mann-Whitney-U-Test, n.s. = Wahrscheinlichkeit des U-Wertes größer 0.05 (zweiseitige Prüfung)

Kriterien	Strenge Gutachter	Nicht strenge Gutachter	Alle Gutachter
1. Beschwerden behandlungsbedürftig?	0	0	0
2. Indikation für Verhaltenstherapie?	0	0	0
3. Anamnese ergibt Fallverständnis?	4	3	7
4. Befund entspricht ICD-10-Diagnose?	5	0	5
5. Verhaltensanalyse erklärt Störung?	5	2	7
6. Therapieziele werden Störung gerecht?	1	0	1
7. Therapieplan beschreibt erfolgversprechende VT?	1	0	1
8. Prognose ausreichend?	3	0	3
9. Therapie notwendig, wirtschaftlich, umfasst Notwendiges?	1	0	1

Tabelle 7: Welche Kriterien waren bei Fall D nicht erfüllt?

Signifikanzprüfung der Unterschiede zwischen strengen und nicht strengen Gutachtern mit Qui-Quadrat. Alle Prüfungen waren nicht signifikant mit Wahrscheinlichkeiten weit über 0.05.

kann deshalb nicht interpretiert werden. Am wichtigsten ist, dass die Einteilung in strenge und nicht strenge Gutachter sich nicht in der Häufigkeit der Nichtbefürwortungen niederschlägt. D. h., die nicht strengen Gutachter filtern ebenso klar Fälle heraus, die nicht befürwortet werden können. Ebenso bedeutsam ist der Vergleich der Zahl der Äußerungen von Bedenken bei Befürwortung ohne Kürzung. Dass hier keine signifikanten Unterschiede zwischen den beiden Gutachtergruppen vorliegen, zeigt, dass die nicht strengen Gutachter bei befürworteten Anträgen ebenso kritisches Feedback geben. Ihre Kritik führt aber seltener zu Kürzungen und sie halten eine Therapie auch dann für gerechtfertigt, wenn einzelne Punkte im Bericht unklar bleiben.

Trotzdem bleibt die Frage bestehen, welche Mängel des Fallberichts D die schlechtere Beurteilung erklären können.

Tabelle 7 zeigt, dass es leichte, nicht signifikante Unterschiede zwischen strengen und nicht strengen Gutachtern in der Beurteilung von Fall D gibt (Einschätzung des Befundes, der Verhaltensanalyse und der Prognose). Insgesamt ist Fall D dadurch gekennzeichnet, dass Anamnese, Befund und Verhaltensanalyse von etwa einem Viertel der Gutachter als unzureichend eingeschätzt wurden.

Bei Fall C gab es nur bei der Einschätzung des Befundes leichte, ebenfalls nicht signifikante Unterschiede bei den beiden Gutachtergruppen. Insgesamt wurde bei Fall C viermal der Befund als unzureichend eingeschätzt.

Diskussion

Dreißig Verhaltenstherapie-Gutachtern wurden fünf unausgewählte Anträge zur Begutachtung vorgelegt. Neunundzwanzig gaben zu jedem Fall und einer zu vier Fällen eine Stellungnahme ab. Sie füllten eine Checkliste zur Einschätzung der Qualität der Fallberichte mit neun Kriterien aus, die eindeutig fehlende, eher fehlende, nicht ganz eindeutige und eindeutige Erfüllung des jeweiligen Kriteriums erfasste. Zudem gaben sie an, ob sie den Antrag ungekürzt mit oder ohne Bedenken befürworten, gekürzt befürworten oder noch Fragen stellen, bevor sie endgültig entscheiden. Bei vier von fünf Fällen waren alle Kriterien erfüllt, wenngleich nicht immer eindeutig. Aufgrund der Kriterien konnten die Fälle in unge-

eignet (ein Fall), geeignet mit Mängeln (zwei Fälle) und gut geeignet (zwei Fälle) eingeteilt werden. Die Stellungnahmen wiesen eine gute Übereinstimmung auf. Der ungeeignete Fall wurde mit 97 % Übereinstimmung abgelehnt, die beiden gut geeigneten Fälle wurden mit 92 % Übereinstimmung ungekürzt befürwortet. Lediglich bei den mit Mängeln behafteten geeigneten Fällen war die Übereinstimmung der Gutachter nicht so gut. 65 % der Beurteilungen führten bei ihnen zu ungekürzter Befürwortung, 20 % zu Nachfragen und 15 % zu Kürzungen. Es kann gesagt werden, dass bei neun erfüllten Kriterien mit einer ungekürzten Befürwortung, bei acht erfüllten Kriterien mit Nachfragen und bei sieben oder sechs erfüllten Kriterien mit Kürzungen gerechnet werden muss. Weniger als sechs erfüllte Kriterien können zur Nichtbefürwortung führen.

Die inhaltlichen Ausführungen der Rückmeldungen an die Antragsteller sind differenziert und vielfältig. Das kritische Feedback der Gutachter trotz ungekürzter Befürwortung ist ein in der bisherigen Diskussion nicht berücksichtigter Bestandteil des Gutachtens. Er ging nicht in Statistiken ein, obgleich er als steuerndes Instrument ebenfalls berücksichtigt werden muss. Es geht ja nicht um Kosteneinsparung, sondern um Qualitätssicherung. Wenn ein gefundener Mangel die Erfolgsaussichten einer Therapie bzw. deren Wirtschaftlichkeit und Notwendigkeit in Frage stellt, so muss keine Nichtbefürwortung erfolgen. Er ist dann ein Hinweis für den Antragsteller, dass sich die Qualität seines Berichts in einem oder mehreren spezifischen Aspekten der tolerierten Qualitätsgrenze nähert. Zugleich wurde deutlich, dass die Gutachter bezüglich dieser Toleranzgrenze mit 96,6 % eine sehr hohe Übereinstimmung haben. Bezüglich des Umgangs mit leichten Mängeln war die Übereinstimmung mit 70 % nicht so hoch. Die beiden Schwellen zwischen Befürwortung trotz Bedenken, Nachfragen und Kürzung waren von Gutachter zu Gutachter bei im Grund sehr ähnlicher Einschätzung der Qualität des Berichts verschieden. Welches dieser drei Vorgehensweisen die wirksamere Qualitätssicherung darstellt, konnte bislang nicht untersucht werden.

Es gibt eine strenge Gruppe von 15 Gutachtern, die nachfragen und kürzen, und eine nicht strenge Gruppe von 13 Gutachtern, die zwar inhaltlich die gleichen Bedenken äußern,

aber trotzdem ungekürzt befürworten. Die strengen und nicht strengen Gutachter unterscheiden sich nicht in der Nichtbefürwortung ungeeigneter Fälle. Und sie unterscheiden sich nicht signifikant in der Äußerung von Bedenken bei leichten Mängeln. Der Unterschied liegt nur darin, dass eine Gutachtergruppe bei durchschnittlich gleicher Einschätzung der Kriterienerfüllung schneller zum Instrument der Kürzung greift und dass diese Gruppe auch mit der Befürwortung zögert, bis weitere Fragen geklärt sind. Da nicht bekannt ist, wie oft sie nach Eingang der Antworten nicht befürworten und wie oft sie befürworten, sind die hier vorliegenden Querschnittsdaten nicht ausreichend, um die endgültige Nichtbefürwortungsquote dieser Gutachtergruppe zu berechnen. Insgesamt zeigte sich, dass die neun Kriterien zur Einschätzung der Qualität der Fallberichte sehr bedeutsam für die abschließende Beurteilung im Gutachten sind und dass bei den hier vorgelegten Fällen die Kriterien Krankheitswertigkeit/Behandlungsbedürftigkeit und Indikation besonders gut zwischen geeigneten und ungeeigneten Fällen differenzierten. Durch eine stärkere Orientierung an den Kriterien zur Einschätzung der Qualität eines Fallberichts könnte jedoch die Gutachterübereinstimmung gesteigert werden, verbunden mit einer Operationalisierung der Konsequenzen, die aus mangelhafter Erfüllung der Kriterien resultieren. Da es sich um eine Pilotstudie handelt, beruhen die Berechnungen auf einer sehr geringen Anzahl von Fällen und sind in ihrer Generalisierbarkeit eingeschränkt. Es können jedoch Tendenzen aufgezeigt werden, die in umfangreicheren Studien zu differenzierteren Fragestellungen führen können. Im Vergleich mit der Studie von Rudolf et al. (2002) zeigt sich, dass eine ähnliche Trennschärfe zwischen geeigneten und ungeeigneten Fällen besteht, dass die Fälle in beiden Studien mit guter Übereinstimmung gesehen wurden, dass aber die Übereinstimmung nicht gut ist, sowohl bezüglich der Einschätzung von Mängeln anhand der Kriterienliste als auch bezüglich der Konsequenzen, die aus der Mängelfeststellung bei leichten Mängeln resultieren. Dies führt zu dem Ergebnis, dass das Gutachterverfahren sinnvoll ist, ungeeignete Fälle ausfindig zu machen, dass es aber als Instrument der externen Qualitätssicherung mit dem Ziel der steten Qualitätssteigerung zur Minimierung von Mängeln, die nicht zu einer Nichtbefürwortung führen, entwicklungsfähig ist. Die hier untersuchten Kriterien bilden die Varianz der Gutachterbeurteilungen gut ab, und ihre Verwendung kann vielleicht ein Beitrag zur Weiterentwicklung des Gutachterverfahrens sein. Inwiefern andere - objektivere und überwiegend quantifizierte - Instrumente diese Aufgabe besser erfüllen können, bleibt dahingestellt.

Literatur

- Deutsche Ärztliche Gesellschaft für Verhaltenstherapie: Vorschlag zur Weiterentwicklung des Gutachterverfahrens unter Berücksichtigung der Ergebnisbewertung. *Psychotherapie* 2001; Bd. 6, 149
- Faber FR, Dahm A, Kallinke D: Kommentar Psychotherapie-Richtlinien. 5. Aufl. München: Urban und Fischer, 1999
- Köhlke HU: Qualitätssicherung durch Gutachterverfahren – Aber wie qualitätsgesichert ist das Verfahren selbst? In: Laireiter AR, Vogel H (Hrsg.): *Qualitätssicherung in der Psychotherapie und psychosozialer Versorgung*. Tübingen: DGVT, 1998, 139-149
- Linden M, Dankesreiter R: Die Durchführung von Verhaltenstherapie im Rahmen der kassenärztlichen Versorgung: Fallberichte, Antragsverfahren und Kassenabrechnung. In Margraf J (Hrsg.): *Lehrbuch der Verhaltenstherapie*. Band 1. Berlin: Springer, 1996, 537-642
- Palm W, Bertram H, Schneider-Reinsch E, Pruemmer C, Feldmann M: Qualitätssicherung in der Diskussion (mit 8 Einzelbeiträgen). *Verhaltenstherapie und psychosoziale Praxis* 2000;32:292-307
- Rudolf G, Jakobsen T, Hohage R, Schlösser A: Wie urteilen Gutachter? Entscheidungsmuster von Psychotherapiegutachtern auf der Grundlage der Gutachterkriterienliste. *Psychotherapeut* 2002;47: 249-253
- Rüger U: Zum Gutachterverfahren bei psychoanalytisch begründeten Behandlungsverfahren. *Psychotherapeutische Praxis* 2001;3: 139-149
- Sulz SKD: Probleme der Antragsbegutachtung. *Prax Klin Verhaltensmed Reha* 1990;12:258-261
- Sulz SKD: Qualitätskriterien der Psychotherapie-Ausbildung. *Psychotherapie* 1998;3:8-15
- Sulz SKD: Der Therapieabschlussbericht. *Psychotherapie* 2001; Bd. 6, 148
- Sulz SKD: Qualitätssicherung in der ambulanten Psychotherapie. *Psychotherapie* 2001;6:30-56
- Weigelt I: Kassenantrag: tiefenpsychologisch fundierte und analytische Therapie. In: Senf W, Broda M (Hrsg.): *Praxis der Psychotherapie*. Stuttgart: Thieme, 2000, 507-513

Dr. Dr. Serge K.-D. Sulz
 Nymphenburger Str. 185
 80634 München
 Fax • 089-132 133
 cipmedien@aol.com