



CME-Beitrag

Evaluation von Psychotherapien aus der Perspektive von Qualitätssicherung und Qualitätsmanagement

Wolfgang Hiller, Gaby Bleichhardt und Amrei Schindler

Psychologisches Institut, Johannes-Gutenberg-Universität Mainz

Zusammenfassung. Dieser Beitrag befasst sich mit methodischen Fragen, die sich aus den Prämissen des Qualitätsmanagements (QM) für die Evaluation von Psychotherapien unter Routinebedingungen ergeben. Am Beispiel der Poliklinischen Institutsambulanz der Universität Mainz wird aufgezeigt, welche Anforderungen an den Datenerhebungsprozess zu stellen sind und nach welchen Kennwerten die Therapieergebnisse bewertet werden können. Es werden Qualitätskriterien für Datenqualität und Erfolgsmaße vorgeschlagen. Die Ergebnisqualität sollte sowohl mit Hilfe von Mittelwertsveränderungen (und dazugehörigen Effektstärken) als auch durch Response- und Remissionsanalysen (Beurteilung des individuellen Ansprechens auf die Therapie) gemessen werden. Entsprechend des Vorgehens in kontrollierten Therapiestudien wird vorgeschlagen, zwischen Intention-to-Treat- und Completer-Analysen zu unterscheiden. Hierfür sind Kriterien erforderlich, nach denen Therapieabschlüsse als Abbruch oder reguläre Beendigung klassifiziert werden können. Anhand eines Datensatzes von 493 abgeschlossenen Routinetherapien wird demonstriert, dass reine Completer-Analysen tendenziell zu einer Überschätzung der Therapieergebnisse führen. Eine QM-bezogene Evaluation eröffnet die Chance, über langjährige vergleichende Bewertungsprozesse ein realistisches Bild der Therapiequalität zu erhalten und daraus kontinuierliche Verbesserung abzuleiten.

Schlüsselwörter: Psychotherapie, Qualitätssicherung, Qualitätsmanagement, Evaluation, naturalistische Psychotherapieforschung

Evaluation of Psychotherapies from the Perspective of Quality Assurance and Quality Management

Abstract. This article discusses methodological questions that may arise when routine psychotherapies are evaluated according to the principles of quality management. Examples from the Psychotherapeutic Outpatient Clinic of the University of Mainz are given to demonstrate how data should be assessed and how criteria can be defined to rate treatment outcomes. Criteria for data quality and treatment parameters are proposed. Mean differences (together with effect sizes) as well as response and remission rates should be analyzed. It is also suggested to systematically differentiate between intention-to-treat- and completer analyses, which requires that clear criteria have to be set up to classify end of treatment as either regular or dropout. Data of 493 completed treatments are used to demonstrate that completer analysis tends to overestimate treatment success. Evaluations within the frame of quality management enable an institution to develop a realistic picture of treatment quality by keeping track of quality indices over longer periods of time. This again may serve as a source for continuous improvements.

Keywords: psychotherapy, quality assurance, quality management, evaluation, naturalistic psychotherapy research

Psychotherapie wird heute als Leistung der gesetzlichen oder privaten Krankenversicherungen erbracht. Gesundheitsbehörden und Kostenträger fordern vermehrt Wirksamkeitsbelege, um die entstehenden Kosten zu rechtfertigen. Auch Patienten und Verbraucherorganisationen sind an fundierten Informationen über die Effizienz der Behandlungen und den zu erwartenden Erfolgsaussichten interessiert. Psychotherapeuten und Einrichtungen der psychosozialen Versorgung reagieren auf derartige Forderungen

meist mit der Aussage, dass sich die Therapien bereits in wissenschaftlichen Studien als wirksam erwiesen hätten. Allerdings ist nicht garantiert, dass eine in der Routineversorgung angewandte Therapiemethode zu ähnlichen Effekten führt wie die gleiche Methode im Setting einer kontrollierten Therapiestudie. Deswegen ist es erforderlich, psychotherapeutische Routinebehandlungen unmittelbar zu evaluieren und die Ergebnisse mit denen aus Forschungseinrichtungen zu vergleichen (Shadish, Matt, Navarro &

Phillips, 2000). Im *Phasenmodell der Therapieevaluation*, das aus der Pharmakotherapieforschung heraus entwickelt wurde (Buchkremer & Klingberg, 2001; Linden, 1989), handelt es sich bei Studien in der Routineversorgung um *Phase-IV-Forschung*. Das Vorgehen in dieser Phase besteht im Wesentlichen aus nicht-randomisierten Effectiveness-Studien mit erweiterten Einschlusskriterien, zunehmend naturalistischen Bedingungen sowie Anwendungsbeobachtungen, die einen Bezug zur psychotherapeutischen Qualitätssicherung aufweisen.

In der Realität werden heute jedoch nur wenige Psychotherapien systematisch und fortlaufend evaluiert. Ambulante Therapeuten setzen allenfalls vereinzelt und sporadisch qualifizierte Untersuchungsmethoden ein, um Veränderungen im Verlauf ihrer Behandlungen zu erfassen. In einem von der Techniker-Krankenkasse initiierten Pilotprojekt wird gegenwärtig erstmals eine systematische therapiebegleitende Diagnostik erprobt (Fydrich, Nagel, Lutz & Richter, 2003). Im *stationären Bereich* haben einige psychosomatische und psychiatrische Kliniken Evaluationssysteme entwickelt, die fest in den Routineablauf integriert sind und zum Teil Auswertungen auf hohem Niveau erlauben (z. B. Mestel et al., 2000). Mittlerweile werden auch an einigen *psychotherapeutischen Hochschulambulanzen* Anstrengungen unternommen, die dort durchgeführten Therapien kontinuierlich wissenschaftlich zu begleiten und Aussagen über Ihre Effektivität abzuleiten (Hiller et al., 2006; Hiller & Bleichhardt, 2004). Diese Ambulanzen sind an psychologische Universitätsinstitute angegliedert, wo in der Regel umgrenzte Forschungsprojekte zu speziellen psychotherapeutischen Fragestellungen durchgeführt werden. Viele Institute sind zudem als Ausbildungsstätten zur Psychotherapeutenausbildung zugelassen. Im Rahmen dieser Ausbildung finden Routinebehandlungen in oft erheblichem Umfang statt, die zum Zwecke einer naturalistischen Therapieevaluation genutzt werden können.

Methodisch gibt es für die Evaluation von Routinetherapien noch *keine allgemein akzeptierten Standards*. Die in kontrollierten Phase-III-Studien eingesetzten Methoden können nicht eins zu eins übertragen werden, da Routinetherapien nicht auf einen bestimmten Evaluationszeitraum begrenzt sind und die behandelten Patienten meist eine eher heterogene klinische Gruppe darstellen. Bei der Entwicklung eines Evaluationssystems sind viele grundsätzliche Entscheidungen zu treffen. So muss festgelegt werden, an welchen Zeitpunkten Messungen stattfinden sollen, wie generell bzw. spezifisch diagnostische Instrumente einzusetzen sind, in welchen Zeitabständen Analysen erfolgen, ob klinische Teilgruppen ausgewertet werden und wie Therapieabbrüche zu berücksichtigen sind. Werden Psychotherapien fortlaufend evaluiert, so ist die Erhebung von Verlaufsdaten nur ein erster Schritt im Gesamtkonzept. Es muss zusätzlich spezifiziert werden, welche *Kennwerte* zu errechnen sind und welche *Konsequenzen* diese anschließend auf das Therapiekonzept der Einrichtung haben.

In diesem Artikel soll die Entwicklung eines solchen Evaluationssystems am Beispiel der *Poliklinischen Insti-*

tutsambulanz der Universität Mainz dargestellt werden. Diese Ambulanz hat als erste deutsche Hochschulambulanz ein *Qualitätsmanagement (QM)* implementiert, das 2005 nach den Normen der DIN EN ISO 9001 zertifiziert wurde. Die Zielsetzungen und Inhalte dieses QM haben wir bereits an anderer Stelle ausführlich dargelegt (Hiller et al., 2006; Hiller & Bleichhardt, 2004). Das QM umfasst neben den Betriebsabläufen auch sämtliche Aspekte der wissenschaftlichen Therapieevaluation. Bei der Entwicklung des Evaluationssystems haben wir uns im Laufe der Zeit zunehmend konsequent an den Anforderungen des QM orientiert.

Im Folgenden soll zunächst kurz das QM mit seinen Implikationen für die Planung, Durchführung und Bewertung von Therapien erläutert werden. Wir werden dann einige Betriebsanforderungen beschreiben, die für ein gut funktionierendes Evaluationssystem unabdingbar sind. Im Anschluss soll eine Auswertungsstrategie dargestellt werden. Im letzten Abschnitt gehen wir schließlich darauf ein, in welcher Form Konsequenzen aus den erhaltenen Evaluationsdaten für die Therapiequalität gezogen werden können. Zur besseren Veranschaulichung werden die einzelnen Schritte anhand von Daten der Hochschulambulanz Mainz dargestellt.

Die Idee des QM

Ein *QM-System* beschreibt und regelt alle Maßnahmen, die von einem Betrieb zur Absicherung der Qualität seiner Produkte oder Dienstleistungen getroffen werden. Psychotherapie kann grundsätzlich als Dienstleistungsprodukt verstanden werden, dessen Qualität nach bestimmten Kriterien bewertbar ist (Härter, Linster & Stieglitz, 2003). Im deutschen Gesundheitssystem besteht seit 1999 für Leistungserbringer die gesetzliche Verpflichtung zur Qualitätssicherung. Viele Arztpraxen und Krankenhäuser verfügen mittlerweile über ein zertifiziertes QM-System. Im Bereich der ambulanten Psychotherapie vollzieht sich eine ähnliche Entwicklung, obwohl die Diskussion zum Für und Wider der Qualitätssicherung noch anhält (z. B. Laireiter, 2000a; Laireiter & Vogel, 1998).

Nach Donabedian (1966) wird zwischen Struktur-, Prozess- und Ergebnisqualität unterschieden. Die *Strukturqualität* bezieht sich auf alle Aspekte der räumlichen, sachlichen und personellen Ausstattung eines Betriebs. Die *Prozessqualität* umfasst alle organisatorisch-logistischen Abläufe einschließlich der implementierten therapeutischen Maßnahmen. Angemessene Strukturen und gut funktionierende Prozesse bilden die Basis der *Ergebnisqualität*. Zu den Ergebnissen gehören alle Sachverhalte und Daten, die eine Bewertung des hergestellten Produkts bzw. der erbrachten Dienstleistung erlauben. Für das Setting der psychotherapeutischen Hochschulambulanz haben wir die einzelnen Qualitätsbereiche bereits ausführlich beschrieben (Hiller & Bleichhardt, 2004).

Tabelle 1
Ziele und Kriterien der Ergebnisqualität

Ziele	Methoden	Kriterien
Hohe Patientenzufriedenheit	Regelmäßige Patientenbefragung	Rücklauf > 85 % Durchschnittsnote < 2.0 (Schulnote) Weiterempfehlungsquote > 90 %
Hohe Therapiequalität	Verlaufserhebung therapierelevanter Daten	Rücklauf aller Messungen > 85 % bei vierteljährlichen Stichprobenkontrollen von jeweils 30 Fragebögen: Gesamtfehlerzahl pro Kontrolle = 10, kein Fragebogen mit > 3 Fehlern
	Abbrecheranalyse	Abbruchquote < 20 %
	Prä-Post-Analyse	Effektstärke $d > 0.50$ Response > 50 % Remission > 40 %

Das Herzstück jedes QM bildet die so genannte *Qualitätspolitik*. Die Ambulanz (bzw. Praxis) muss in einer sehr grundsätzlichen Form ihren Stellenwert und ihre Ambitionen im Sinne einer «Unternehmensphilosophie» darlegen (z. B. Psychotherapien auf nachweisbar hohem Niveau erbringen, eine hohe Patientenzufriedenheit erzielen oder eine gute Therapeutenausbildung garantieren). Aus der Qualitätspolitik werden in einem zweiten Schritt konkrete *Qualitätsziele* abgeleitet, die wiederum durch *Qualitätskriterien* zu operationalisieren sind. Die Qualitätskriterien werden am zweckmäßigsten in Form von Kennwerten festgelegt, welche fortlaufend zu festgelegten Zeitpunkten – meistens im Jahreszyklus – erhoben werden. Anhand solcher Kennwertdaten kann die tatsächlich erreichte Qualität ermittelt und anhand der selbstgesteckten Ziele bewertet werden. Werden Qualitätskriterien nicht erreicht, so müssen über den sog. *PDCA-Zyklus* (Plan – Do – Check – Act) die Gründe analysiert, Verbesserungsmaßnahmen in die Wege geleitet und deren Auswirkungen überprüft werden.

Aus dem QM abgeleitete Evaluationsziele

Die Hochschulambulanz Mainz verfolgt die Qualitätspolitik einer *konsequenten Kundenorientierung* und der *Erreichung exzellenter Therapieergebnisse*. Die angestrebten Therapieeffekte sollen sich mit den in kontrollierten Forschungstherapien erzielten Resultaten messen können. In Tabelle 1 sind die auf die Ergebnisqualität bezogenen Ziele und Kriterien dargestellt. Als Zielgrößen werden nicht nur die Endbewertungen in Form beispielsweise von Durchschnittsnoten und Effektstärken festgelegt, sondern auch Merkmale der Datenqualität. Hohe Rücklaufquoten sollen garantieren, dass die erhaltenen Daten aussagekräftig sind und systematische Verzerrungen wie beispielsweise die Antwortbereitschaft der Patienten keine markante Rolle spielen. Mit Hilfe von Stichprobenkontrollen und einem festgelegten Toleranzbereich wird die Güte der Dateneingabe geprüft. Die in Tabelle 1 angegebene Effektstärke von $d > 0.50$ orientiert sich am Mindeststandard der Psychotherapieforschung. Wir meinen hier und

im Folgenden stets naturalistische Eingruppen-Prä-Post-Effektstärken nach Cohen (1992), bei denen die Mittelwertedifferenz an der Standardabweichung relativiert wird (vgl. auch Maier-Riehle & Zwingmann, 2000). In der Hochschulambulanz Mainz stellte der Wert von $d > 0.50$ zunächst eine Ausgangssituation dar, die im Weiteren noch näher ausdifferenziert werden soll.

Anders als in vielen Therapiestudien haben *Zufriedenheitsmessungen* aus QM-Perspektive einen hohen Stellenwert (vgl. auch Seligman, 1995). Hierfür werden meistens unkomplizierte Kurzfragebögen eingesetzt und die Bewertung erfolgt pragmatisch anhand von Schulnoten (oder äquivalenten Skalen). Die Befragungen sollten anonym erfolgen, um keinen Anlass für Antworttendenzen der sozialen Erwünschtheit oder anderer Art zu geben.

In der Hochschulambulanz Mainz erfolgt einmal jährlich eine *Patientenbefragung*, in die alle Patienten, die innerhalb eines fixen 2-Wochen-Zeitraums mindestens einen Termin haben, berücksichtigt werden. Im Kurzfragebogen wird erfasst, wie die Patienten die therapeutische Leistung, den Service des Sekretariats, den Pflegezustand der Ambulanz sowie die Leistung insgesamt beurteilen. Bei der Gesamtbeurteilung erreichten wir in den Jahren 2004 bis 2008 Werte zwischen 1.71 und 1.82 (SD 0.51–0.60) auf der Schulnotenskala 1–5. Von besonderem Interesse ist die Weiterempfehlungsquote («Würden Sie die Institutsambulanz weiterempfehlen?»), für die wir in den vergangenen Jahren Werte zwischen 92.9 % und 96.4 % erreichten. Gute Zufriedenheitswerte garantieren keine ausreichenden Therapieeffekte, jedoch muss Patientenzufriedenheit als eine wichtige Voraussetzung für die Erreichung von guten Therapieresultaten angesehen werden. Zufriedenheitsmessungen sind ein zentrales Instrument der Kundenorientierung und haben sich insbesondere in den stationären Einrichtungen der psychosozialen Gesundheit bereits breit etabliert.

Aufbau eines Evaluationssystems

Um eine systematische Evaluation betreiben zu können, müssen *Messzeitpunkte* festgelegt und geeignete *diagnosti-*

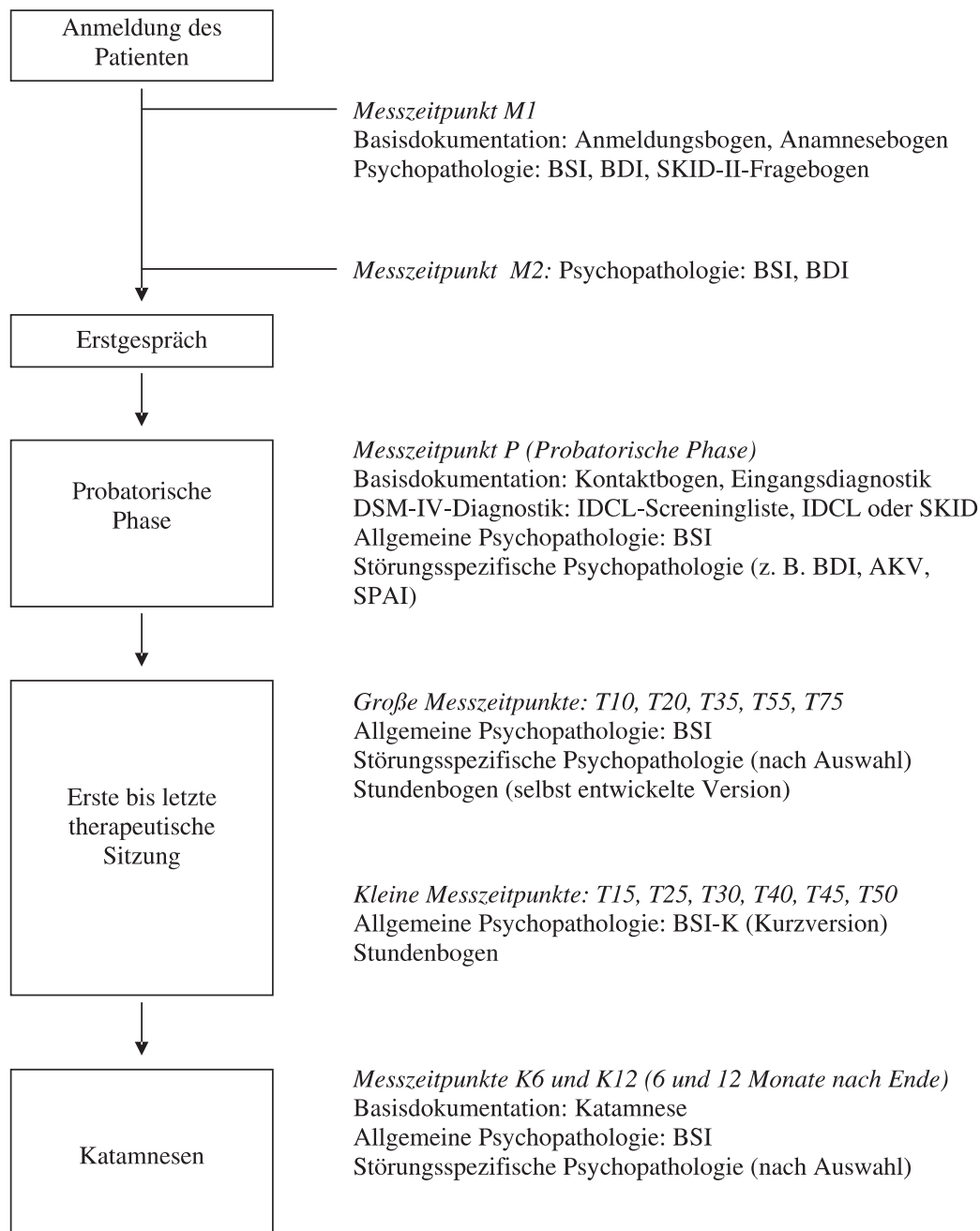


Abbildung 1. Messzeitpunkte der Therapieevaluation.

sche Instrumente ausgewählt werden. Zudem müssen Betriebsabläufe ein gewisses Zuverlässigkeitsniveau erreichen, damit die in den Qualitätskriterien beschriebene Datenqualität erreicht werden kann. In der Hochschulambulanz Mainz wurde ein für die Datenqualität verantwortliches Evaluationsteam gebildet, in dem mehrere, zumeist studentische oder in Therapieausbildung befindliche Mitarbeiterinnen mit einem Kontingent von insgesamt 40–50 Wochenarbeitsstunden tätig sind.

Eine Übersicht der Messzeitpunkte und eingesetzten diagnostischen Verfahren gibt Abbildung 1. Die von uns er-

fassten Daten lassen sich in folgende fünf Komponenten gruppieren: (1) Basisdokumentation, (2) Diagnostik der psychischen Störungen nach DSM-IV, (3) allgemeine Psychopathologie, (4) störungsspezifische Psychopathologie, (5) Stundenbewertungsbögen zur Erfassung der allgemeinen Qualität therapeutischer Sitzungen.

Die *Basisdokumentation* umfasst alle persönlichen und krankheitsbezogenen Grunddaten des Patienten. Von verschiedenen Autoren sind entsprechende Dokumentationssysteme entwickelt und publiziert worden (Laireiter, 2000b). Bestandteile der an der Mainzer Ambulanz eingesetzten Ba-

sisdokumentation sind die folgenden Komponenten: (A) Anmeldung/Kontaktaufnahme; (B) Anamnesebogen; (C) Kontaktbogen inklusive Verzeichnis wichtiger Adressen, z. B. des Hausarztes; (D) Eingangsdiagnostik; (E) Therapieabschlussbewertung des Patienten; (F) Therapieabschlussbewertung des Therapeuten; (G) Katamnesebogen (mit identischen Fassungen für Halb- und Ganzjahreskatamnese).

Die *Diagnostik der psychischen Störung* einschließlich Komorbidität erfolgt nach dem Klassifikationssystem DSM-IV, da dieses im Vergleich zur ICD-10 stärker an wissenschaftlichen Grundsätzen orientiert ist und insgesamt die bessere Qualität aufweist. Das Vorgehen erfolgt in zwei Stufen: Zunächst wird mittels der Screening-Fragen der Internationalen Diagnose-Checklisten (IDCL; Hiller, Zaudig & Mombour, 1997) und dem Screening-Fragebogen des SKID-II für die wichtigsten psychischen Störungen geprüft, ob ein diagnostischer Anfangsverdacht für eine oder mehrere Störungen vorliegt. Bei positivem Screeningbefund erfolgt eine gründliche Abklärung mit Hilfe von strukturierten bzw. halbstrukturierten Interviews anhand der IDCL und des SKID-II. Standard der Mainzer Hochschulambulanz sind die IDCL. Zu Beginn ihrer Tätigkeit wird den Ambulanztherapeuten empfohlen, die ersten Interviews anhand des SKID-I (Wittchen, Zaudig & Fydrich, 1997) durchzuführen und danach aufgrund des unkomplizierten, im Prinzip aber identischen Vorgehens zu den IDCL zu wechseln. Liegen Störungen vor, die durch die IDCL oder das SKID nicht abgedeckt sind, so werden die entsprechenden diagnostischen Kriterien anhand des DSM-IV-Manuals explizit überprüft.

Die *Veränderungsmessung* während des Therapieverlaufs erfolgt im Wesentlichen durch Verfahren der allgemeinen und störungsspezifischen Psychopathologie. Wir setzen das Brief Symptom Inventory (BSI; Franke, 2000) ein, eine 45 Items umfassende Kurzform der SCL-90-R. Die Verfahren zur Erfassung der störungsspezifischen Psychopathologie werden zum Messzeitpunkt P entsprechend der gestellten DSM-IV-Diagnosen ausgewählt. Es handelt sich beispielsweise um das Beck-Depressions-Inventar (BDI; Hautzinger, Bailer, Worall & Keller, 1994) oder die bei Panikstörung/Agoraphobie häufig verwendeten Fragebögen zu körperbezogenen Ängsten, Kognitionen und Vermeidung (AKV; Ehlers, Margraf & Chambless, 2001). Kriterien für die in der Mainzer Ambulanz ausgewählten störungsspezifischen Fragebögen stellten vor allem der Verbreitungsgrad, die internationale Vergleichbarkeit sowie ihre Veränderungssensitivität und Ökonomie dar.

Eine weitere Datenquelle sind *Stundenbewertungsbögen*, die sowohl vom Patienten als auch Therapeuten zu unterschiedlichen Zeitpunkten ausgefüllt werden. Im Wesentlichen werden darin Charakteristika des therapeutischen Prozesses und der Therapeut-Patienten-Interaktion erfasst. Wir haben die Items aufgrund einer Literaturrecherche zu ähnlichen Ansätzen systematisch zusammengestellt und bereiten derzeit eine Evaluation dieses Instruments vor.

Die in Abbildung 1 dargestellten *Evaluationsperioden*

umfassen neben der *therapeutischen Kernphase* (zwischen M2 und der letzten Messung) auch die Zeiträume vor Beginn und nach Ende der Therapie. Die erste Datenerhebung findet unmittelbar nach der (meist telefonischen) Anmeldung des Patienten statt (M1). Der Patient erhält Anamnesebogen sowie BSI, BDI und SKID-II-Fragebogen per Post zugesandt und wird gebeten, diese ausgefüllt innerhalb von sieben Tagen zurückzusenden. Es handelt sich hierbei um eine *Baseline-Messung*. Der zweite Messzeitpunkt (M2) liegt unmittelbar vor dem Erstgespräch. Das Erstgespräch stellt den ersten persönlichen Kontakt des Patienten in der Ambulanz dar. An der Hochschulambulanz Mainz beträgt die Wartezeit zwischen Anmeldung und Erstgespräch üblicherweise vier bis sechs Monate. Daher kann die Zeitspanne zwischen M1 und M2 ähnlich wie in Forschungsprojekten als *Warteperiode* (im Sinne einer Wartekontrollgruppe) angesehen werden, das heißt als *Referenzbedingung*, mit der die späteren Therapieeffekte verglichen werden können. Zum *Messzeitpunkt P* (Probatorische Phase) führt der Therapeut eine umfassende Diagnostik und Therapieplanung durch. Neben der Diagnosenstellung werden Angaben zu Medikamenten, Vorbehandlungen und allgemeinem Funktionsniveau gemäß der Achse V des DSM-IV erhoben. Nach Festlegung der für das jeweilige Störungsbild zugeordneten Fragebögen erfolgt erstmals die Erfassung der störungsspezifischen Psychopathologie. Erst nach vollständigem Vorliegen aller Daten wird der Therapiebeginn von der Ambulanzleitung freigegeben.

Die Datenerhebung in der *therapeutischen Phase* ist durch die Messzeitpunkte T5, T10 usw. gekennzeichnet (T5 kennzeichnet die fünfte therapeutische Sitzung, wobei diese Zählung im Anschluss an die probatorischen Sitzungen P1 bis P5 beginnt). In unserem Evaluationssystem wird zwischen «großen» und «kleinen» Messzeitpunkten unterschieden. An großen Messzeitpunkten wird die allgemeine und störungsspezifische Psychopathologie mit Hilfe der gleichen Fragebögen wie zum Messzeitpunkt P erhoben. Diese Messzeitpunkte liegen für alle Patienten bei T10 und T20, für Patienten mit einer Langzeittherapie nochmals bei T35. Bei Verlängerungen finden im Abstand von 20 Sitzungen weitere Erhebungen statt (T55, T75). Bei der letzten Messung im Therapieverlauf werden zusätzlich Therapieabschlussbögen gegeben. Darüber hinaus findet bei jeder fünften Sitzung ein kleiner Messzeitpunkt statt, bei dem lediglich eine 15 Item umfassende Kurzform des BSI ausgefüllt wird. Dieser sog. BSI-KF ist von uns entwickelt worden (Bleichhardt & Hiller, 2003) und weist neben einer hohen Veränderungssensitivität eine Korrelation von 0.80 mit dem GSI-Globalwert des Gesamt-BSI auf. Die kleinen Messzeitpunkte dienen als Ergänzung zu den großen Messzeitpunkten zur Erstellung eines kontinuierlichen Verlaufsprofils für die allgemeine Psychopathologie (welches auch als fortlaufendes Rückmeldesystem für Therapeuten und Supervisoren eingesetzt wird).

Nach Abschluss der Therapie werden *6- und 12-Monatskatamnesen* (K6 und K12) postalisch durchgeführt. Dabei werden die gleichen störungsspezifischen Fragebögen eingesetzt wie schon während der Therapie. Hinzu kommen

BSI sowie ein von uns entwickelter Katamnesebogen, in dem die Patienten unter anderem die Besserung in allgemeinen Symptombereichen (z. B. Stimmung, persönliche Beziehungen) global einschätzen. Das Datum des Verschickens der Katamneseunterlagen wird anhand der letzten erfolgten Therapiesitzung berechnet.

Bedingungen der Datenanalyse

Die Auswertung der erhaltenen Daten ist ein komplexer Prozess. Aus dem Grundgedanken des QM lassen sich einige Leitprinzipien für den Auswertungsprozess ableiten. Dabei werden Daten stets im Lichte der vorab aufgestellten Qualitätsziele und -kriterien interpretiert. An der Hochschulambulanz Mainz wurde ein Auswertungsprotokoll entwickelt, das die einzelnen Auswertungsstufen beschreibt. Das Auswertungsprotokoll ist selbst wiederum in das QM implementiert.

Zu den wichtigsten Besonderheiten einer QM-bezogenen Therapieevaluation gehören die drei folgenden Punkte:

1. *Keine Zeitbegrenzungen:* Es handelt sich um eine kontinuierliche, hinsichtlich der Zeitperspektive offene Evaluation. Dies unterscheidet die QM-bezogene Evaluation von Forschungsprojekten, bei denen normalerweise Daten innerhalb eines definierten Zeitfensters oder bis zum Erreichen eines Kriteriums (z. B. ausreichende Stichprobengröße) erhoben werden. Bei einem funktionierenden QM ist die Evaluation fester Bestandteil des Ambulanz- bzw. Praxisbetriebs. Sie dient nicht einer einmaligen Bewertung, sondern Qualitätsprozesse sollen fortlaufend beobachtet und gezielt gesteuert werden können.
2. *Evaluationszyklen:* Trotz der prinzipiell offenen Zeitperspektive existieren wiederkehrende Evaluationszyklen, da QM-Systeme regelmäßig – meist einmal pro Jahr – auditiert (bewertet) werden. Im DIN ISO EN 9001-System sind die Audits mit einem Qualitätsbericht und einem sog. Management-Review für das abgelaufene Jahr verbunden. Darin wird dargelegt, wie sich die Einrichtung im Berichtszeitraum entwickelt hat, inwieweit Qualitätskriterien erreicht (oder verfehlt) wurden und welche Konsequenzen daraus für das nachfolgende Jahr gezogen werden. In der Hochschulambulanz Mainz werden mittlerweile fast alle Auswertungen des Jahreszyklus routinemäßig durchgeführt. Bei etwa 200–300 Therapien pro Jahr ergeben sich für die meisten Analysen ausreichend große Datensätze. Nur für die Evaluation seltener Störungsgruppen bzw. Therapieereignisse müssen längere Evaluationszyklen festgelegt werden.
3. *Adaptive Anpassung der Qualitätskriterien:* Die Existenz von Qualitätskriterien stellt bereits einen wesentlichen Unterschied zu Forschungstherapien dar, wo meist keine Vorab-Festlegung des angestrebten Mindesttherapieeffekts stattfindet (auch bei der Hypothese, dass die untersuchte Behandlung bessere Ergebnisse als eine

Kontrollgruppe erreicht, werden üblicherweise keine Grenzwerte festgelegt). Es existieren heute nur sehr wenige Veröffentlichungen über die Effekte in ambulanten Routinepsychotherapien, so dass eine realistische Einschätzung der erreichbaren Effektstärken «guter Therapien» nahezu unmöglich ist. Auch die Orientierung an den Effekten von Forschungstherapien ist mit Unsicherheit behaftet. Wir haben bei der Entwicklung des QM-Systems eine Effektstärke von $d = 0.50$ für Mittelwertverbesserungen als zunächst vorläufigen Grenzwert festgelegt. Durch die regelmäßigen Jahresevaluationen kann sich zeigen, inwieweit dieser Grenzwert realistisch ist und welche Spielräume nach oben (und ggf. nach unten) bestehen. Zudem ist zu hoffen, dass sich zunehmend die Möglichkeit ergibt, die eigenen Therapieresultate mit denen anderer vergleichbarer Einrichtungen zu vergleichen. Beim sog. *Benchmarking* wird eine konstruktive Auseinandersetzung mit konkurrierenden Einrichtungen und Qualitätswettbewerb um die Marktführerschaft angestrebt.

Im Folgenden wird eine Strategie der Datenanalyse vorgestellt, wie wir sie in den letzten Jahren an der Hochschulambulanz Mainz entwickelt haben. Der Leitgedanke war, die Evaluationsmethodik so eng wie möglich an den Standards der streng kontrollierten Psychotherapieforschung auszurichten, so dass als Unterschied zwischen den Phasen III und IV im Wesentlichen die fehlende Randomisierung zu gelten hat.

Flussdiagramm der Fallzahlen

Evaluationsdaten sind aufgrund von *Stichprobenfluktuationen* oft nur schwer nachvollziehbar. Im naturalistischen Setting müssen unter anderem die Zahlen der in der Einrichtung vorstellig gewordenen, der in die Therapie aufgenommenen und der die Therapie regulär beendeten Patienten unterschieden werden. Insbesondere wird jedoch von guten Therapiestudien gefordert, dass zwischen *Intention-to-Treat- (ITT)* und *Completer-Analysen* differenziert wird (Hollis & Campbell, 1999; Little & Yau, 1996). ITT erfordert den Einbezug aller ursprünglich in die Studie aufgenommenen Patienten in die Auswertung, während bei Completer-Auswertungen nur solche Patienten berücksichtigt werden, welche die Therapie regulär durchlaufen haben. International akzeptierte Richtlinien für die Darstellung der Methodik und Ergebnisse von Therapiestudien sind im sog. *CONSORT-Statement* aufgestellt worden («Consolidated Standards of Reporting Trials»; vgl. Moher, Schulz & Altman, 2004). Wir haben das CONSORT-Flussdiagramm, das einen Überblick über die Patientenzahlen zu den verschiedenen Zeitpunkten der Evaluation ermöglichen soll, auf die Bedingungen naturalistische Therapieforschung übertragen (s. Abbildung 2).

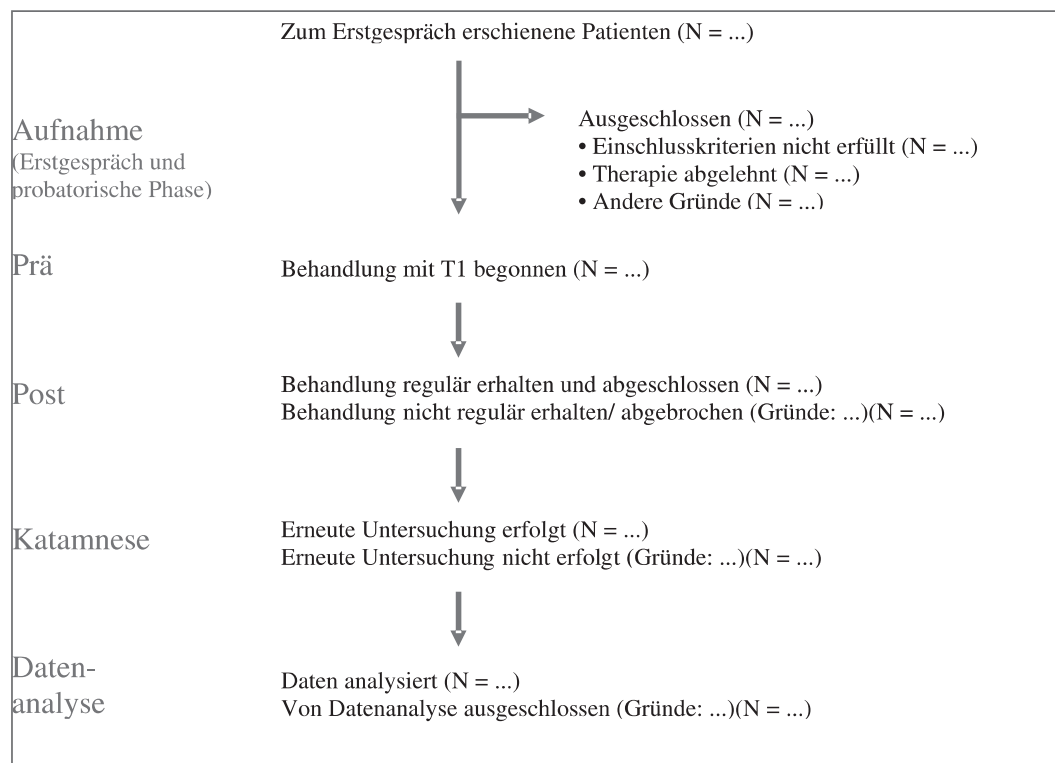


Abbildung 2. Flussdiagramm für Therapieevaluation im naturalistischen Setting (in Anlehnung an das CONSORT-Statement).

Um ITT-Analysen im naturalistischen Setting durchführen zu können, müssen zum Prä-Zeitpunkt (T1) Daten für alle in die Therapie aufgenommenen Patienten vorliegen. Die probatorische Phase zählt noch nicht als Therapiebeginn, da sie gemeinsam mit dem Erstgespräch dazu dient, die Anfangsvoraussetzungen für eine Behandlung zu prüfen (z. B. genaue Indikationsabklärung, Klärung von Behandlungsmotivation und Therapiezielen). Falls eine Therapie mit T1 begonnen hat, werden die vor der probatorischen Phase erhobenen Daten (M2) als Prä-Daten angesehen. Beim Post-Zeitpunkt muss zwischen Abbrechern (Drop-outs) und Completern unterschieden werden können. Bei Therapieabbrüchen wird der jeweils letzte erhaltene Messwert als Post-Messung angesehen (bei frühem Therapieabbruch kann dies auch der Prä-Messwert sein). Zudem kann aus dem Flussdiagramm in Abbildung 2 nachvollzogen werden, wie hoch die Ausschöpfungsrates der katamnestischen Untersuchungen war und wie oft Fälle aufgrund von fehlenden oder unvollständigen Daten von der Endauswertung ausgeschlossen werden mussten.

Definition von Therapieabbrüchen

Therapieabbrüche können Ausdruck von *Qualitätsmängeln der Therapie* sein, beispielsweise inadäquat eingesetzte Interventionen oder Schwierigkeiten des Patienten-The-

rapeuten-Verhältnisses. Deswegen ist die Erfassung und Analyse von Abbrüchen im Rahmen des QM-Systems von besonderer Bedeutung. Es empfiehlt sich, entsprechende Kennwerte zu definieren und diese fortlaufend zu erheben. Für die Hochschulambulanz Mainz haben wir im Jahr 2005 eine differenzierte Erfassung von Abbrüchen und Abbruchgründen durchgeführt (Renke, 2007).

Wie bereits dargelegt, werden Beendigungen nach dem Erstgespräch oder während der probatorischen Phase *nicht* als Abbrüche gewertet. Dies entspricht dem Vorgehen in Forschungsstudien, wo vor Beginn einer Behandlung bei jedem vorstellig gewordenen Patienten das Vorliegen der Ein- und Ausschlusskriterien geprüft wird. Die Therapien, bei denen mindestens eine T-Sitzung durchgeführt wurde und bei denen das beantragte Stundenkontingent für Kurz- oder Langzeittherapie nicht voll ausgeschöpft wurde, werden als «abgebrochen» gewertet mit Ausnahme der Therapien, deren vorzeitiges Ende aufgrund eines ausreichenden Therapieergebnisses zustande kam. In den Jahren 2003 bis 2007 wurde bei 143 von 535 Patienten die Therapie abgebrochen (26.7 %). Die möglichen Gründe hierfür können aufgrund ihrer Qualitätsrelevanz in zwei Gruppen unterteilt werden:

(a) *Potenziell oder sicher qualitätsrelevante Abbrüche*: Patient kam nicht mehr oder war nicht mehr erreichbar; Therapiemotivation des Patienten zu gering; Patient hatte eine andere Behandlung begonnen; keine Gründe angegeben;

zu geringer Therapieerfolg; keine Fortführung nach stationärer Behandlung; Patient gab an, keine Zeit mehr zu haben; Bedingungen/Therapeuten nicht akzeptiert; Therapeutenwechsel abgelehnt; Ablehnung der Krankenkasse, Patient fühlte sich gebessert ohne Nachvollziehbarkeit durch den Therapeuten; Patient hat Rechnung nicht bezahlt.

(b) *Nicht-qualitätsrelevante Therapiebeendigung*: Wechsel des Wohnorts; Therapeut empfahl andere Maßnahmen; Leidensdruck zu gering; Mutterschaft der Therapeutin.

Wie hoch sollte oder darf die Quote der qualitätsbezogenen Abbrüche maximal sein? In der Analyse der vergangenen Jahre lagen 96 potenziell oder sicher qualitätsrelevante Gründe vor, was einem Anteil von 17.9 % bezogen auf die 535 Patienten der Ausgangsstichprobe entspricht. Aus der uns bekannten Literatur können allenfalls vereinzelte Hinweise auf die Abbruchquote in regulären ambulanten Einrichtungen gefunden werden. Westbrook und Kirk (2005) ermittelten eine Dropout-Rate von 22.5 % in einer amerikanischen Mental-Health-Clinic («patients who stop attending without prior agreement») und bei Wade, Treat & Stuart (1998) waren es in einer ähnlichen Einrichtung im Kontext einer Angstbehandlung 26.4 %. Aus Deutschland sind uns keine entsprechenden Daten bekannt. Um einen vorläufigen Orientierungswert zu haben, legten wir für unser QM eine zu erreichende kritische Grenze von < 20 % fest.

Schritte der Datenanalyse

Im Folgenden soll die *Sequenz der Datenauswertung* am Beispiel unserer Ambulanz für die Jahre 2003 bis 2007 dargestellt werden. Wir beschränken uns hier auf psychopathologische Outcome-Variablen, da wir sie bei der Bewertung von Krankenbehandlungen für die entscheidenden Kriterien halten (z. B. soll die Behandlung einer Depression zuallererst dazu führen, dass sich die depressive Symptomatik bessert). Oft werden weitere Outcome-Maße wie Lebensqualität, psychosoziales Funktionsniveau, interpersonelle Probleme oder Wahrnehmungs-/kognitive Auffälligkeiten erhoben. In so gut wie allen Therapiestudien korrelieren diese Variablen jedoch mittelgradig bis stark mit den psychopathologischen Maßen, so dass die Prä-Post-Vergleiche einzelner Variablen in der Regel sehr ähnlich ausfallen.

Schritt 1: Prä-Post-Analyse der allgemeinen Psychopathologie

Wie bereits erwähnt, sind Patienten von Routinetherapien üblicherweise eine heterogene klinische Gruppe mit unter-

schiedlichen psychischen Störungen und Schweregraden. Deswegen empfiehlt sich als erster Schritt eine Analyse der *globalen Psychopathologie*. Die therapeutische Hypothese ist, dass sich die allgemeine psychische Belastung und Symptomatik unabhängig vom Vorliegen spezieller Störungsbilder durch die Behandlung bessern lässt.

Die entsprechenden Resultate unserer Hochschulambulanz sind in Tabelle 2 für die GSI-Scores des BSI dargestellt, und zwar separat für die einzelnen Evaluationsjahre (die Jahre 2003 und 2004 wurden aufgrund seinerzeit noch geringer Fallzahlen zusammengefasst). Als Prä-Wert wurde die Messung zu M2, bei deren Nichtvorliegen ersatzweise die zum Zeitpunkt P genommen. Als Post-Wert diente die letzte vorhandene Messung. Die Ausgangsstichprobe umfasst 493 Patienten, die das Kriterium erfüllten, im genannten Evaluationszeitraum ihre Therapie beendet zu haben. Rund 60 % sind weiblich, die häufigsten Diagnosenbereiche sind affektive Störungen (etwa 55 %) und Angststörungen (etwa 45 %). Weitere soziodemographische Merkmale sowie exakte DSM-IV-Diagnosenverteilungen sind in unseren Jahresberichten ab 2003 lückenlos dargestellt worden und sollen hier nicht näher vertieft werden. Die Jahresberichte können auch auf Dauer jederzeit eingesehen werden (www.klinische-psychologie-mainz.de/abteilung/download.html). Die mittlere Therapiedauer im Anschluss an die probatorische Phase betrug 33.1 Sitzungen (*SD* 17.2). Der weitaus überwiegende Teil der Therapien wurde von in Ausbildung befindlichen Therapeuten unter engmaschiger Supervision durchgeführt. Alle Therapien waren kognitiv-verhaltenstherapeutisch.

Im ersten Block von Tabelle 2 befinden sich die Ergebnisse der unselektierten Ausgangsstichprobe, bei der Faktoren wie Ausgangsniveau der Prä-Messung und Therapieverlauf unberücksichtigt sind. Jedoch ist insbesondere die Frage pathologischer Anfangswerte von großer Bedeutung. In unseren Stichproben war bei etwa einem Drittel der Patienten der Anfangs-GSI nicht im auffälligen Bereich (definiert als Rohwert > 0.61 entsprechend der Empfehlung von Franke, 2000; dieser Grenzwert entspricht einem T-Wert von 63). In diesen Fällen kann und sollte der GSI auch nicht als Outcomemaß verwendet werden, da keine dazugehörige Behandlungshypothese zu begründen ist (im Kontext von Krankenbehandlungen ist es kein Ziel, einen nicht-pathologischen Merkmalsbereich zu verbessern). Patienten ohne pathologischen Anfangswert sind also *auszuschließen*, was in den Analysen der beiden unteren Blöcke von Tabelle 2 geschehen ist (die Werte der unselektierten Ausgangsstichprobe sind hier nur deshalb aufgeführt, um die damit verbundene Unterschätzung der tatsächlichen Therapieeffekte zu verdeutlichen). Dennoch sollten Patienten ohne pathologischen Anfangswert nicht gänzlich unberücksichtigt bleiben, worauf weiter unten noch eingegangen wird.

Tabelle 2 zeigt, dass sich die Effektstärken der ITT-Analyse im Bereich zwischen 0.60 und 0.92 und die der Completer-Analyse zwischen 0.54 und 1.07 bewegten. Bei ITT ergeben sich mit Ausnahme der zusammengefassten Jahre

Tabelle 2

Ergebnisse allgemeiner Psychopathologie (GSI des Brief Symptom Inventory BSI)

		Therapiebeginn		Therapieende				
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i> -Wert	<i>d</i>	95 % <i>CI</i>
Ausgangsstichprobe (unselektiert)								
2003/04	70	0.96	0.67	0.62	0.62	5.24**	0.53	0.37–0.67
2005	82	0.90	0.60	0.67	0.61	4.67**	0.38	0.25–0.51
2006	179	0.99	0.61	0.66	0.59	8.33**	0.55	0.46–0.64
2007	162	1.07	0.70	0.68	0.56	7.87**	0.62	0.51–0.71
Intention-to-Treat								
Nur Pat. mit pathologischem Prä-Wert								
2003/04	40	1.38	0.58	0.89	0.68	4.88**	0.78	0.60–0.99
2005	50	1.27	0.47	0.94	0.63	4.50**	0.60	0.47–0.77
2006	125	1.26	0.52	0.81	0.62	9.09**	0.79	0.70–0.90
2007	112	1.38	0.62	0.83	0.58	8.57**	0.92	0.80–1.02
Completer								
Nur Pat. mit pathologischem Prä-Wert								
2003/04	27	1.35	0.58	0.99	0.75	3.03**	0.54	0.32–0.82
2005	33	1.16	0.39	0.72	0.43	5.74**	1.07	0.94–1.22
2006	91	1.20	0.46	0.71	0.54	8.88**	0.98	0.89–1.09
2007	79	1.27	0.54	0.72	0.49	8.82**	1.07	0.95–1.18

Anmerkungen: *M* = Mittelwert; *SD* = Standardabweichung; 95 % *CI* = 95 %-Konfidenzintervall; ***p* < 0.05.

d = Prä-Post-Effektstärke nach Cohen, berechnet als Mittelwertsdifferenz in Relation zur gepoolten Standardabweichung beider Messzeitpunkte (vgl. Maier-Riehle & Zwingmann, 2000), GSI = Global Severity Index. Die Evaluationsjahre sind definiert als Therapieende zwischen KW 45 des Vorjahres bis KW 44 des Indexjahres (KW = Kalenderwoche); durch den Abschluss eines Evaluationsjahres bereits im November soll der Dateneingang bis zum Jahresende sowie die anschließende Evaluation am Jahreswechsel ermöglicht werden.

2003/04 (aus uns unklaren Gründen) geringere Effektstärken und im Jahresvergleich größere Schwankungen als bei regulär abgeschlossenen Therapien.

Bei der *Completer-Analyse* (unterster Block in Tabelle 2) wurden zunächst Patienten mit den oben definierten potenziellen oder sicheren qualitätsrelevanten Abbrüchen ausgeschlossen. Zusätzlich legten wir fest, dass zur Charakterisierung eines Patientenkontaktes im Sinne einer Psychotherapie mindestens drei Therapiesitzungen stattgefunden haben mussten (nach der probatorischen Phase). Bezogen auf das oben vorgestellte Flussdiagramm wurden nach diesem Prozedere für alle Jahreszeiträume zusammen von 475 Patienten mit Therapiebeginn 126 Non-Completer ausgeschlossen (26,6 %). Weitere Verluste aufgrund unvollständiger Datensätze bestanden nicht.

Auch wenn für die Prä-Post-Analyse nur Patienten mit pathologischem Anfangswert berücksichtigt werden, sollten Patienten mit Scores im Normalbereich dennoch nicht aus dem Blickfeld geraten. Schließlich ist nicht auszuschließen, dass sich diese Patienten im Laufe der Therapie verschlechtern. Zur Kontrolle dieses Risikos sollte eine Nebenanalyse durchgeführt werden. Im Datensatz von Tabelle 2 hatten 166 Patienten der unselektierten Ausgangsstichprobe keinen pathologischen GSI-Wert bei Therapiebeginn. Es kam in dieser hinsichtlich der allgemeinen Psychopathologie unauffälligen Stichprobe während der Therapiephase zu keiner Verschlechterung, sondern sogar zu

einer geringfügigen Verbesserung der Werte (Prä: 0.36, *SD* 0.17; Post: 0.30, *SD* 0.28; *t* = 2.86, *df* 165, *p* < 0.05, *d* = 0.27).

Der Anteil von etwa einem Drittel Patienten ohne pathologischen Anfangswert erscheint sehr hoch und weckt den Verdacht, dass möglicherweise Therapien ohne entsprechende Indikation durchgeführt wurden. Wir sind dieser Vermutung nachgegangen, konnten sie jedoch ausschließen. Es gibt in der Tat viele Fälle, in denen die allgemeine Psychopathologie kein adäquates Evaluationsmaß darstellt. Beispiele sind umschriebene Phobien oder Essstörungen ohne weitere Komorbidität. Nicht selten sind auch teilremittierte depressive Störungen, wo sich die Psychopathologie schon deutlich zurückgebildet hat, aber dennoch kein Zweifel an der Notwendigkeit einer Psychotherapie zur weiteren Stabilisierung und Rückfallprophylaxe besteht. Es ist essenziell, dass ein hochwertiges Evaluationssystem diese Konstellationen zu berücksichtigen in der Lage ist.

Schritt 2: Prä-Post-Analysen für spezielle klinische Gruppen

Das beschriebene Vorgehen kann nun auch auf spezielle Psychopathologieformen übertragen werden, das heißt auf *umschriebene Störungsgruppen*. In Tabelle 3 sind exem-

Tabelle 3

Ergebnisse störungsspezifischer Psychopathologie bei Patienten mit depressiver Störung (BDI)

		Therapiebeginn		Therapieende				
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i> -Wert	<i>d</i>	95 % <i>CI</i>
Intention-to-Treat								
Nur Pat. mit pathologischem Prä-Wert								
2003/04	27	26.3	8.0	17.0	9.9	5.13**	1.03	−1.98–4.77
2005	28	25.5	7.4	15.9	11.8	4.92**	1.00	−1.74–5.37
2006	75	23.2	6.8	13.1	8.8	9.62**	1.29	−0.24–3.29
2007	69	25.1	9.3	13.4	10.2	8.21**	1.20	−0.99–3.61
Completer								
Nur Pat. mit pathologischem Prä-Wert								
2003/04	19	25.7	8.2	16.2	11.3	3.98**	0.97	−2.71–6.06
2005	19	23.3	7.5	11.0	8.7	5.37**	1.52	−1.85–5.43
2006	56	23.4	6.9	12.0	8.8	9.40**	1.45	−0.35–3.76
2007	47	24.6	9.0	12.2	9.1	7.83**	1.37	−1.20–3.97

Anmerkungen: *N* = Stichprobengröße; *M* = Mittelwert; *SD* = Standardabweichung; 95 % *CI* = 95 %-Konfidenzintervall; ** $p < 0.05$; *d* = Prä-Post-Effektstärke nach Cohen. BDI = Beck Depressions-Inventar.

plarisches die Ergebnisse für Patienten mit depressiver Störung dargestellt (DSM-IV-Diagnosen Major Depression oder Dysthyme Störung). Jedoch sollte bedacht werden, dass selbst bei Patienten mit einer solchen Diagnose nicht zwangsläufig ein pathologischer Anfangswert vorliegen muss (z. B. bei voller oder Teilremission zwischen den depressiven Episoden). Die Analysen in Tabelle 3 beziehen sich daher nur auf Patienten mit einem auffälligen Prä-BDI-Wert von ≥ 15 Punkten. Die Therapieeffekte sind generell höher als bei den oben dargestellten Resultaten der allgemeinen Psychopathologie, was die Relevanz störungsspezifischer Analysen unterstreicht. Die Konfidenzintervalle der Effektstärken sind aufgrund der relativ geringen Jahresstichproben sehr breit und umfassen auch den Wert Null. Wird keine jahresbezogene Unterteilung vorgenommen, so liegen die Effektstärken bei 1.17 (*CI* 0.07–2.56; *N* = 198) für die ITT- und bei 1.36 (*CI* 0.06–2.88; *N* = 141) für die Completer-Analyse. Für das Stichproben-Flussdiagramm ergeben sich bei insgesamt 198 Patienten mit Therapiebeginn 57 Non-Completer (28.8 %). Es kam zu keinen Verlusten aufgrund unvollständiger Datensätze.

Ähnliche Analysen sind für andere klinische Gruppen wie Angststörungen, Zwangsstörungen, somatoforme Störungen, psychotische Störungen usw. möglich und sinnvoll.

Schritt 3: Response- und Remissionsanalysen

Mittelwertsvergleiche haben den Nachteil, dass das individuelle Ansprechen der Patienten auf die Behandlung nicht abgeschätzt werden kann. Es ist denkbar, dass trotz bedeutender Mittelwertsunterschiede ein nicht unerheblicher Anteil der behandelten Patienten nicht oder nur geringfügig auf die Therapie angesprochen hat. Deswegen sind nach

unserer Auffassung zusätzliche Responderanalysen als Teil der Routineauswertung erforderlich.

Therapieresponse muss von *Remission* unterschieden werden. Eine Remission liegt vor, wenn sich ein Patient unter Therapie aus dem pathologischen Bereich einer Psychopathologieskala in den unauffälligen Bereich verändert hat. Dies sollte zwar das Ziel jeder Behandlung sein, jedoch muss bei psychotherapeutischen Behandlungen auch dann von Erfolg gesprochen werden, wenn sich beispielsweise ein Patient aus einer anfänglich sehr schwer ausgeprägten Symptomatik in den mittleren oder leichtgradigen Störungsbereich verbessert hat. Response kann also auch bei fehlender Remission vorliegen.

Wie sollte Response nun definiert werden? Hierzu werden in der Literatur unter dem Stichwort der «klinischen Signifikanz» verschiedene Vorschläge gemacht. Diese lassen sich grob in die beiden folgenden Positionen unterteilen:

1. *Verteilungsabhängige Definitionen*: Am bekanntesten ist der von Jacobson und Truax (1991) entwickelte «Reliable Change Index» (RCI). Damit lässt sich errechnen, wie groß eine individuell erreichte Veränderung auf einer bestimmten Skala mindestens sein muss, um als «klinisch bedeutsam» bewertet zu werden (ausführliche Darstellung und Diskussion z. B. bei Wise, 2004). Der kritische Wert ist abhängig von Verteilungsparametern der jeweils zugrunde liegenden Stichprobe und der Reliabilität des eingesetzten Messverfahrens. Ein verwandter Ansatz besteht darin, die zu erreichende Mindestveränderung anhand der Stichprobenstreuung festzulegen, meist 1 bis 2 Standardabweichungen. Bei Remissionsanalysen wird zusätzlich auf die Verteilungen nicht-klinischer Populationen zurückgegriffen. Verteilungsabhängige Definitionen haben jedoch erhebliche Nachteile (Existenz verschiedener Normierungsdaten, Settingab-

Tabelle 4
Ergebnisse der Response- und Remissionsanalysen

	GSI		BDI	
	Response	Remission	Response	Remission
Intention-to-Treat				
2003/04	55.0	40.0	55.9	41.2
2005	50.0	36.0	51.1	48.9
2006	64.8	46.4	68.0	59.2
2007	61.6	42.0	69.0	54.0
Completer				
2003/04	48.2	37.0	54.2	45.8
2005	63.6	42.4	70.0	66.7
2006	71.4	52.8	75.0	65.3
2007	65.8	49.4	74.6	55.9

Anmerkungen. Alle Angaben in %. *Response*: Rückgang mind. 50 % im pathologischen Wertebereich und mind. 25 % im gesamten Wertebereich. *Remission*: Score bei Behandlungsbeginn pathologisch (GSI > 0.61, BDI > 14), bei Behandlungsende unauffällig; außerdem Responsekriterium erreicht.

hängigkeit usw.). Wie die Literatur zeigt, ist es äußerst schwer, Konsens zwischen verschiedenen Forschergruppen und klinischen Einrichtungen über die Größe der Mindestveränderung zu erreichen. Wenn neue klinische Studien mit den betreffenden Instrumenten publiziert werden, entsteht Unsicherheit darüber, welche der existierenden Verteilungen heranzuziehen ist und ob die Charakteristika der Stichprobe, aus der die Verteilung stammt, tatsächlich mit der vorliegenden Gruppe behandelter Patienten vergleichbar ist.

2. *Methoden der prozentualen Wertereduktion*: Ein pragmatischer Ansatz besteht darin, von einer klinisch bedeutsamen Verbesserung dann auszugehen, wenn ein Patient unter Therapie eine bestimmte prozentuale Reduktion seines Psychopathologiescores erreicht hat. Diese Methode hat sich in vielen Feldern der Psychopharmaforschung bereits als Standard durchgesetzt (Bandelow et al., 2006). In der medikamentösen Depressionsbehandlung werden für die ResponderEinstufung üblicherweise mindestens 50 % Reduktion auf Fremdbeurteilungsverfahren (z. B. HAMD) gefordert. Bei der Behandlung von Psychosen aus dem schizophrenen Formenkreis werden niedrigere Werte angesetzt, meist 25 bis 30 %. Es bietet viele Vorteile, diese Methode auch bei der Psychotherapieevaluation anzuwenden, insbesondere die dadurch entstehende Vergleichbarkeit mit der Psychopharmaforschung und die Unabhängigkeit von Populations- und Stichprobenverteilungen.

In unserer Evaluationsstrategie haben wir uns daher auf die Methode der prozentualen Reduktion festgelegt. Durch die Orientierung am 50 %-Kriterium der Pharmaforschung wird ein *gleichwertiger Qualitätsanspruch* erhoben. Im Verlauf unserer entsprechenden Datenanalysen fiel allerdings auf, dass bei Anwendung des 50 %-Kriteriums auf die diversen Outcome-Instrumente keineswegs Einheitlichkeit erreicht wird. Dies liegt darin begründet, dass der

als nicht-auffällig definierte «Normalbereich» einzelner Instrumente in Relation zum Gesamtwertebereich unterschiedlich breit ist. In einem Verfahren mit relativ engem Normalbereich ist ein bestimmter prozentualer Rückgang «weniger wert» als in einem anderen Verfahren mit breitem Normalbereich.

Ein fiktives Beispiel soll dies verdeutlichen: Zwei Verfahren haben einen gleichen Gesamtwertebereich zwischen 0 und 50 Punkten. In Verfahren 1 sei der nicht-pathologische Bereich zwischen 0 und 10, in Verfahren 2 zwischen 0 und 20. Ein Patient bessert sich nun von 40 auf 30, was eine absolute Besserung um $10/40 = 25\%$ darstellt. In Verfahren 1 mit dem pathologischen Bereich 10–50 wäre die relative Besserung $10/30 = 33\%$ (die Symptomatik hat sich innerhalb des pathologischen Wertebereichs um ein Drittel gebessert), in Verfahren 2 mit dem pathologischen Bereich 20–50 jedoch $10/20 = 50\%$ (innerhalb des pathologischen Bereichs um die Hälfte gebessert). Es wird also deutlich, dass der Rückgang von 40 auf 30 Punkten in Verfahren 2 (mit größerem Normalbereich) eine stärkere Besserung darstellt als in Verfahren 1 (mit kleinerem Normalbereich). Dies hat zur Konsequenz, dass Responsequoten auf der Basis absoluter Prozentreduktionen zwischen verschiedenen Verfahren in aller Regel nicht vergleichbar sind.

Wir haben uns daher entschlossen, zwar eine Reduktion von mindestens 50 % für unsere Responsedefinition zugrunde zu legen, dies jedoch bei allen Verfahren nur auf den pathologischen Wertebereich zu beziehen. Um auszuschließen, dass im Grenzbereich zwischen «pathologisch» und «unauffällig» bereits minimale Besserungen den Responderstatus erreichen (z. B. wäre auf einer Skala mit Normalbereich 0–20 ein Rückgang von 22 auf 21 bereits eine 50-prozentige Reduktion, was der klinischen Interpretation widerspricht), wird neben der 50 %-Reduktion im pathologischen Bereich zusätzlich eine Mindestreduktion von 25 % auf der Gesamtskala gefordert (im o. g. Beispiel einer Skala mit Normalbereich 0–20 müsste ein Patient also

bei einem Ausgangswert von 22 mindestens um 6 Punkte gebessert sein, um als Responder eingestuft zu werden).

Die Ergebnisse der mit diesen Kriterien erfolgten Auswertungen der Hochschulambulanz Mainz sind in Tabelle 4 zusammengefasst. Angegeben sind sowohl Response- als auch Remissionsraten für GSI und BDI der einzelnen Evaluationsjahre entsprechend der ITT- und Completer-Analysen. Die Responderanteile bewegen sich zwischen knapp 50 und 75 % und sind bei den Completern gegenüber der ITT-Analyse deutlich höher (mit Ausnahme der zusammengefassten Jahre 2003/04). Die Remissionsraten (definiert als Score bei Therapieende im unauffälligen Bereich *plus* Erreichen des oben beschriebenen Responsekriteriums) liegen im Bereich zwischen 36 und 67 % und sind sowohl beim BDI als auch bei den Completer-Analysen etwas höher.

Im Rahmen des QM müssen wir der Frage nachgehen, *warum ein nicht unerheblicher Teil der Patienten nicht auf die Behandlung anspricht*. Mit der Suche nach Prädiktoren für das Eintreten oder Ausbleiben des Therapieerfolgs eröffnet sich ein großes Weiterentwicklungspotenzial (Steketee & Chambless, 1992). Für die QM-Ziele (vgl. Tabelle 1) sollte auch ein Kriterium für die Responseanalyse festgelegt werden. Auch hier scheint man Neuland zu betreten, denn unseres Wissens sind bislang aus ambulanten psychotherapeutischen Einrichtungen keine entsprechenden Ergebnisse publiziert worden. Basierend auf den hier dargestellten Ergebnissen haben wir spezifiziert, künftig einen Responderanteil von mindestens 50 % und Remissionsraten von mindestens 40 % erreichen zu wollen. Wie bei den anderen Kennwerten müssen auch diese Kriterien in Zukunft auf ihre Angemessenheit überprüft werden.

Schritt 4: Analyse der Warteperiode vor Therapiebeginn

Im Evaluationssystem der Mainzer Hochschulambulanz wird routinemäßig (s. o.) der *Wartezeitraum* bis zum Therapiebeginn evaluiert (s. o.). Dadurch haben wir die Möglichkeit, unsere Therapieverläufe mit etwaigen Veränderungen der Warteperiode zu vergleichen. Diese stellt also gewissermaßen eine Art *Kontrollbedingung* dar.

Für eine unselektierte Stichprobe der Jahre 2003 bis 2007 (unabhängig vom späteren Zustandekommen einer Therapie) von 289 Patienten lagen GSI-Werte für die Zeitpunkte M1 (bei Anmeldung) und M2 (vor Erstgespräch) vor. Von diesen hatten 206 einen pathologischen GSI bei M2. Die Wartezeit dieser Gruppe betrug im Mittel 130.5 Tage (*SD* 109.6), also knapp viereinhalb Monate. Der mittlere GSI-Wert blieb in der Warteperiode im Prinzip unverändert mit minimaler Tendenz zu Verschlechterung (*M1*: 1.34, *SD* 0.67; *M2*: 1.36, *SD* 0.62; Effektstärke $d = -0.03$ mit 95 %-CI zwischen -0.12 und 0.05). Interessanterweise gab es einen leichten Rückgang der Werte, wenn die M2-Erhebung nicht vor sondern erst nach dem Erstgespräch zu

Beginn der probatorischen Phase erfolgen konnte (*M1*: 1.45, *SD* 0.61; *M2* nach Erstgespräch: 1.34, *SD* 0.59 bei $N = 138$; $d = 0.18$ mit 95 %-CI zwischen 0.08 und 0.28). Dies weist darauf hin, dass schon durch den persönlichen Kontakt des Erstgesprächs ein gewisser therapeutischer Effekt erzielt wird. Es sollten daher im Rahmen der Betriebsabläufe alle Anstrengungen unternommen werden, die Endmessung der Warteperiode *vor* dem ersten persönlichen Kontakt mit dem Patienten vorzunehmen. Es empfiehlt sich, dieses Ziel als QM-Kriterium zu definieren oder die Betriebsabläufe so zu gestalten, dass Erstgespräche ohne Vorliegen der M2-Messung nicht stattfinden können.

Bewertung der Therapieresultate im Rahmen des QM

Abschließend möchten wir auf die Frage eingehen, wie die erhaltenen Therapieresultate zu bewerten sind. Beim gegenwärtigen Stand der naturalistischen Psychotherapieforschung ist es kaum möglich, gut begründete Einschätzungen darüber zu entwickeln, ob die eigenen Therapieergebnisse nun gut oder schlecht sind. Es fehlen schlicht Vergleichsdaten. Beim Verfassen dieses Artikels lagen uns bis auf eine Ausnahme (Hahlweg, Fiegenbaum, Schröder, Frank & von Witzleben, 2004) keine publizierten Daten aus Hochschulambulanzen oder psychotherapeutischen Praxen deutschsprachiger Länder vor. In der internationalen Literatur finden sich aus den letzten Jahren nur ganz vereinzelte naturalistische Studien. Zwei Studien verwendeten den BDI. Persons, Bostrom und Bertagnolli (1999) erzielten bei der Behandlung überwiegend depressiver Patienten mit kognitiver Verhaltenstherapie Prä-Post-Effektstärken von etwa 1.40–1.50 (nachträgliche Berechnung aufgrund der publizierten Daten durch uns). Westbrook und Kirk (2005) erreichten in einer großen gemischten klinischen Stichprobe eine Effektstärke von 1.15, wenn nur Patienten mit einem Anfangs-BDI von 10 oder mehr Punkten einbezogen wurden. In der erwähnten Studien von Hahlweg et al. (2004) wurden 416 Patienten mit Panikstörung und Agoraphobie durch 52 Therapeuten von Institutsambulanzen mit Konfrontationsmethoden behandelt. Im GSI der SCL-90-R erzielten die Autoren eine Effektstärke von 0.99. In den drei genannten Studien wurden jedoch keine ITT-Analysen durchgeführt.

Es ist zu hoffen, dass in den nächsten Jahren vermehrt Daten aus der Routineversorgung erhoben und publiziert werden, was ein Benchmarking im Sinne des QM ermöglichen würde. Eine gewisse Bewertung kann aber auch innerhalb einer Einrichtung vorgenommen werden, wenn im Jahreszyklus regelmäßige Evaluationen stattfinden und sich über die Entwicklung dieser Zahlen im Laufe der Zeit eine Art «interner Standard» für gute und weniger gute Jahresresultate entwickeln kann. Eine andere Möglichkeit ist die Orientierung an den in Forschungstherapien erreichten

Tabelle 5
Implikationen des QM für die Therapieevaluation

QM-Merkmale	Maßnahmen
Datenqualität	
– Vollständigkeit	– Hohe Rücklaufquote definieren und sichern (Kennwert, Betriebsabläufe).
– Eingabegenauigkeit	– Korrektheit der Dateneingabe garantieren (z. B. Stichprobenkontrollen, Kennwert).
Qualität des Untersuchungsplans	
– Therapiebeginn	– Den Zeitpunkt des Therapiebeginns als «Intention-to-Treat» definieren (z. B. T1).
– Therapieabbrüche	– Qualitätsrelevante Abbrüche definieren und deren zuverlässige Erfassung sichern (Betriebsabläufe).
– Patientenfluss	– Zu- und Abgänge im Gesamtablauf dokumentieren (z. B. CONSORT-Flussdiagramm).
Qualität der Datenanalyse	
– Behandlungshypothesen	– Gezielt klinische Fragestellungen analysieren (z. B. Besserung bei pathologischem Anfangswert).
– Outcome-Maße	– Das Therapieergebnis mittels Effektstärken, Response und Remission bestimmen.
– Behandlungsabsicht	– Sowohl Intention-to-Treat- als auch Completer-Analysen berechnen.
Therapiequalität	
– Bewertungszyklen	– Die Therapiequalität in regelmäßigen Zyklen prüfen, z. B. jährlich (Betriebsabläufe).
– Kundenzufriedenheit	– Regelmäßig die Patientenzufriedenheit zu Service und Leistung erfassen (Betriebsabläufe, Kennwert).
Bewertung/Benchmarking	
– Interne Längsschnittbewertung	– Aus den Ergebnissen heraus Maßnahmen zur Weiterentwicklung der Therapiequalität entwickeln.
– Externes Benchmarking	– Mit anderen Einrichtungen bzw. Forschungsergebnissen vergleichen.

Effekten, wobei aber unklar ist, ob dies realistisch ist (vgl. Shadish et al., 2000). Sicherlich kann es nicht falsch sein, durch weiterentwickelte Therapiekonzepte zu versuchen, die Wirksamkeit der Routinetherapien möglichst nahe an die der Forschungstherapien heranzuführen. Eine Grob-orientierung könnten die *allgemein akzeptierten Konventionen* für mittelwertsbezogene Effektstärken sein, das heißt unzureichend bei $d < 0.50$, akzeptabel bei d bis 0.80 und gut bis sehr gut bei $d > 0.80$ (vgl. Maier-Riehle & Zwingmann, 2000). Für Responseraten gibt es nach unserer Kenntnis keinerlei Konventionen.

Diskussion

In diesem Artikel wurde demonstriert, wie sich psychotherapeutische Routinebehandlungen unter den Prämissen eines QM-Systems evaluieren lassen. Die wichtigsten Komponenten sind nochmals in Tabelle 5 zusammengefasst. Wir haben bei unseren Arbeiten der letzten Jahre mit einigem Erstaunen festgestellt, dass für die Analyse von Routinetherapien kaum methodische Standards existieren, weder in der Fachliteratur noch im tatsächlichen Handeln von Ambulanzbetreibern und Praxisinhabern. Erste Ansätze liegen in den Arbeiten der Stuttgarter Forschungsstelle für Psychotherapie (z. B. Kordy, Hannover & Bauer, 2003) sowie von Grawe und Mitarbeitern im Rahmen der Berner Psychotherapieforschung vor (z. B. Grawe & Baltensperger, 1998). Diese Ansätze sind aber äußerst komplex und haben bislang in den Einrichtungen der ambulanten Routineversorgung keine breite Basis finden können. Zudem

sind die speziellen Rahmenbedingungen des QM darin nicht berücksichtigt.

Das in der Evaluationsforschung entwickelte Methodenrepertoire kann auf die Analyse von Routinetherapien angewandt werden. Jedoch stehen Routinetherapien unter einigen besonderen Voraussetzungen. So steht in der Regel keine Kontrollbedingung zur Verfügung. Ein- und Ausschlusskriterien für die Aufnahme von Patienten in die Therapie sind heterogener und oft nicht explizit definiert. Die behandelten Störungsbilder sind vielfältiger als bei umschriebenen Forschungsfragestellungen, was die Auswahl von Instrumenten und die Festlegung von Erfolgskriterien zunächst komplizierter macht. Nicht leicht zu beantworten ist auch die Frage, wann in Routinetherapien von Abbrüchen gesprochen werden sollte, wobei man davon ausgehen muss, dass die Festlegung in dieser Frage erhebliche Auswirkungen auf die Ergebnisse hat. Schließlich sollten auch noch die erforderlichen Betriebsroutinen bedacht werden, die erforderlich sind, um aussagekräftige Daten zu gewinnen. Hier sind besonders die Rücklaufquoten von Bedeutung. In unserer Ambulanz hat es sich herausgestellt, dass durch stringent organisierte Betriebsabläufe ein Rücklauf von über 90 % erreichbar ist. Es ist kaum sinnvoll, viel Aufwand in die Etablierung eines Evaluationssystems zu stecken und nach einigen Jahren feststellen zu müssen, dass die gesammelten Daten aufgrund des geringen Rücklaufs keine wirklich zuverlässige Beurteilung der Therapietugenden erlauben.

Wir möchten abschließend betonen, dass es aus unserer Sicht alles andere als einen trivialen Unterschied darstellt, ob die Evaluation von Routinetherapien aus QM-Perspektive betrieben wird oder nicht. Die QM-bezogene Evaluation ist wesentlich verbindlicher und in ihrer Grundidee

selbstkritisch, da nicht von Qualität ausgegangen wird, sondern die kontinuierliche Verbesserung das Ziel darstellt. Die Bereitschaft zu kontinuierlicher Verbesserung setzt Interesse an Mängeln und Schwächen der eigenen Arbeit voraus. Die in diesem Artikel dargestellten Evaluationsdaten belegen, dass unsere Therapien noch weit von dem entfernt sind, was viele Forschungsberichte an Wirksamkeit suggerieren. Das gilt sowohl für die Effektstärken als auch für den Anteil von Patienten, die nicht oder nur gering auf die Therapie ansprechen.

Leider drängt sich bei naturalistischen Studien nicht selten der Eindruck auf, dass sie den Charakter von «Bestätigungsforschung» haben, das heißt eine Einrichtung oder Berufsgruppe möchte möglichst plakativ belegen, dass das, was sie tut, von hoher Qualität ist. Solche Interessen werden durch ein QM natürlich nicht ausgeschaltet. Wird das QM aber richtig verstanden und ernst genommen, so entwickelt sich eine eher offene Haltung gegenüber den eigenen Schwächen («positive Fehlerkultur»). Evaluation wird zum Werkzeug der Therapiesteuerung und Weiterentwicklung. Während Forschungstherapien üblicherweise durch das Auslaufen des Projekts und die Abschlusspublikation beendet werden, stimuliert die QM-bezogene Evaluation Neu- und Weiterentwicklungen über viele Jahre hinweg. Verbesserungen oder Verschlechterungen sind relativ zeitnah zu erkennen. Dies ist eine völlig andere Situation als bei Forschungstherapien, bei denen vor allem von Praktikern (oft nicht zu unrecht) vermutet wird, dass die guten Therapieeffekte mit einmaligen günstigen Bedingungen zusammenhängen (z. B. außergewöhnlich gut motivierte und eingearbeitete Therapeuten, hohe Manualdisziplin, ausgewählte Patienten). Die QM-bezogene Evaluation eröffnet die Chance, langfristig ein realistischeres und stabileres Bild über die Wirksamkeit von Psychotherapie zu gewinnen als durch temporäre Forschungsbehandlungen (die andererseits als Impulsgeber von großer Bedeutung sind).

Literatur

- Bandelow, B., Baldwin, D. S., Dolberg, O. T., Andersen, H. F. & Stein, D. J. (2006). What is the threshold for symptomatic response and remission for major depressive disorder, panic disorder, social anxiety disorder, and generalized anxiety disorder? *Journal of Clinical Psychiatry*, 67, 1428–1434.
- Bleichhardt, G. & Hiller, W. (2003). *BSI-15 – ein änderungssensitiver Messwert für die allgemeine psychopathologische Belastung*. Unveröffentlichtes Manuskript der Abteilung Klinische Psychologie und Psychotherapie, Johannes-Gutenberg Universität Mainz.
- Buchkremer, G. & Klingberg, S. (2001). Was ist wissenschaftlich fundierte Psychotherapie? Zur Diskussion um Leitlinien für die Psychotherapieforschung. *Nervenarzt*, 72, 20–30.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Donabedian, A. (1966). Evaluating the quality of medical care. *Milbank Memorial Funds Quarterly*, 44, 166–203.
- Ehlers, A., Margraf, J. & Chambless, D. (2001). *Fragebogen zu körperbezogenen Ängsten, Kognitionen und Vermeidung (AKV)*. Göttingen: Hogrefe.
- Franke, G. H. (2000). *Brief Symptom Inventory von Derogatis*. Göttingen: Hogrefe.
- Fydrich, T., Nagel, A., Lutz, W. & Richter, R. (2003). Qualitätsmonitoring in der ambulanten Psychotherapie: Modellprojekt der Techniker Krankenkasse. *Verhaltenstherapie*, 13, 291–295.
- Grawe, K. & Baltensperger, C. (1998). Figurationsanalyse. Ein Konzept und Computerprogramm für die Prozeß- und Ergebnisevaluation in der Therapiepraxis. In A.-R. Laireiter & H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung. Ein Werkstattbuch* (S. 179–207). Tübingen: dgvt.
- Härter, M., Linster, H. W. & Stieglitz, R.-D. (Hrsg.). (2003). *Qualitätsmanagement in der Psychotherapie. Grundlagen, Methoden und Anwendung*. Göttingen: Hogrefe.
- Hahlweg, K., Fiegenbaum, W., Schröder, B., Frank, M. & von Witzleben, I. (2004). Klinische Brauchbarkeit der Konfrontationstherapie für agoraphobische Patienten: Einfluss von syndromaler Komorbidität und therapeutischer Erfahrung. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 33, 130–138.
- Hautzinger, M., Bailer, M., Worall, H. & Keller, F. (1994). *Beck-Depressions-Inventar (BDI)*. Bern: Huber.
- Hiller, W. & Bleichhardt, G. (2004). Qualitätssicherung in einer psychotherapeutischen Hochschulambulanz. *Verhaltenstherapie und Verhaltensmedizin*, 25, 315–344.
- Hiller, W., Bleichhardt, G., Haaf, B., Legenbauer, T., Mauer-Matzen, K. & Rübler, D. (2006). Zertifizierung einer Psychotherapeutischen Hochschulambulanz nach DIN EN ISO 9001. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 35, 225–233.
- Hiller, W., Zaudig, M. & Mombour, W. (1997). *IDCL – Internationale Diagnosen Checklisten für DSM-IV und ICD-10 (Manual, 31 Checklisten nach DSM-IV und Screening-Blatt)*. Göttingen: Hogrefe.
- Hollis, S. & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *British Medical Journal*, 319, 670–674.
- Kordy, H., Hannöver, W. & Bauer, S. (2003). Das Stuttgart-Heidelberger Modell zur Qualitätssicherung in der stationären Psychotherapie. In M. Härter, H. W. Linster & R.-D. Stieglitz (Hrsg.), *Qualitätsmanagement in der Psychotherapie. Grundlagen, Methoden und Anwendung* (S. 289–304). Göttingen: Hogrefe.
- Laireiter, A.-R. & Vogel, H. (Hrsg.). (1998). *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung. Ein Werkstattbuch*. Tübingen: dgvt.
- Laireiter, A.-R. (2000a). Diagnostik, Dokumentation und Qualitätssicherung von Psychotherapie. In A.-R. Laireiter (Hrsg.), *Diagnostik in der Psychotherapie* (S. 441–458). Wien: Springer-Verlag.
- Laireiter, A.-R. (Hrsg.). (2000b). *Diagnostik in der Psychotherapie*. Wien: Springer-Verlag.
- Linden, M. (1989). Die Phase IV der Therapie-Evaluation. *Nervenarzt*, 60, 453–461.
- Little, R. & Yau, L. (1996). Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics*, 52, 1324–1333.
- Maier-Riehle, B. & Zwingmann, C. (2000). Effektstärkenvarian-

- ten beim Eingruppen-Prä-Post-Design: Eine kritische Betrachtung. *Rehabilitation*, 39, 189–199.
- Mestel, R., Erdmann, A., Schmid, M., Klingelhöfer, J., Stauss, K. & Hautzinger, M. (2000). 1–3-Jahres-Katamnesen bei 800 depressiven Patienten nach stationärer Psychosomatischer Rehabilitation. In M. Bassler (Hrsg.), *Leitlinien zur stationären Psychotherapie – Pro und Contra* (S. 243–273). Gießen: Psycho-Sozial-Verlag.
- Moher, D., Schulz, K.F. & Altman, D.G. für die CONSORT Gruppe. (2004). Das CONSORT Statement: Empfehlungen zur Qualitätsverbesserung von Reports randomisierter Studien im Parallel-Design. *Deutsche Medizinische Wochenschrift*, 129, T16–T20.
- Persons, J., Bostrom, A. & Bertagnolli, A. (1999). Results of randomized controlled trials of cognitive therapy for depression generalize to private practice. *Cognitive Therapy and Research*, 23, 535–548.
- Renke, J. (2007). *Wirkfaktoren in der Psychotherapie: Identifizierung leicht zugänglicher Prädiktoren für Abbruchverhalten im ambulanten Setting*. Unveröffentlichte Diplomarbeit, Johannes-Gutenberg-Universität Mainz.
- Seligman, M. E. P. (1995). The effectiveness of psychotherapy. *American Psychologist*, 50, 965–974.
- Shadish, W. R., Matt, G., Navarro, A. & Phillips, G. (2000). The effects of psychological therapies and clinically representative conditions: A meta-analysis. *Psychological Bulletin*, 126, 512–529.
- Steketee, G. & Chambless, D. L. (1992). Methodological issues in prediction of treatment outcome. *Clinical Psychology Review*, 12, 387–400.
- Wade, W. A., Treat, T. A. & Stuart, G. L. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology*, 66, 231–239.
- Westbrook, D. & Kirk, J. (2005). The clinical effectiveness of cognitive behaviour therapy: Outcome for a large sample of adults treated in routine practice. *Behaviour Research and Therapy*, 43, 1243–1261.
- Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment*, 82, 50–59.
- Wittchen, H.-U., Zaudig, M. & Fydrich, T. (1997). *Strukturiertes Klinisches Interview für DSM-IV. Achse I*. Göttingen: Hogrefe.

Prof. Dr. Wolfgang Hiller

Johannes-Gutenberg-Universität Mainz
Psychologisches Institut
Abteilung Klinische Psychologie und Psychotherapie
Staudingerweg 9
D-55099 Mainz
E-Mail: hiller@mail.uni-mainz.de

CME-Fragen

1. Zu den Zielen der Phase-IV-Forschung in der Psychotherapie gehören
 1. Effectiveness-Studien.
 2. die Randomisierungskontrolle.
 3. der Ausschluss von Komorbidität.
 4. die Evaluation natürlicher Therapieverläufe.
 5. die Manualentwicklung.

Welche Aussagenkombination ist richtig?

- A. Antworten 1 und 5 sind richtig.
- B. Antworten 1 und 4 sind richtig.
- C. Antworten 2, 3 und 4 sind richtig.
- D. Antwort 2 ist richtig.
- E. Alle Antworten außer 3 sind richtig.

2. Die Ergebnisqualität des QM-Systems einer psychotherapeutischen Einrichtung

1. ist in Deutschland gesetzlich definiert.
2. ist nur für größere Einrichtungen, nicht für Privatpraxen von Bedeutung.
3. setzt eine gute Struktur- und Prozessqualität voraus.
4. wird auch als «konsequente Kundenorientierung» bezeichnet.
5. kann über Daten zum Therapieverlauf überprüft werden.

Welche Aussagenkombination ist richtig?

- A. Antworten 3 und 5 sind richtig.
- B. Antworten 4 und 5 sind richtig.
- C. Antworten 1, 2 und 3 sind richtig.
- D. Alle Antworten außer 2 sind richtig.
- E. Nur Antwort 5 ist richtig.

3. Welches sind Besonderheiten einer QM-bezogenen Therapieevaluation?

1. Daten werden in wiederkehrenden Zyklen ausgewertet.
2. Die methodischen Ansprüche an die Testverfahren sind nicht so hoch wie bei Forschungstherapien.
3. Es werden nur Patienten mit bestimmten Einschlusskriterien berücksichtigt.
4. Die Evaluation ist hinsichtlich der Zeitperspektive offen.
5. Qualitätskriterien können adaptiv angepasst werden.

Welche Aussagenkombination ist richtig?

- A. Antworten 1 und 5 sind richtig.
- B. Antworten 1, 4 und 5 sind richtig.
- C. Antworten 2, 3 und 5 sind richtig.
- D. Alle Antworten sind richtig.
- E. Antworten 4 und 5 sind richtig.

4. Unter «intention-to-treat» versteht man

1. die Fallzahlentwicklung entsprechend des CON-SORT-Statements.
2. eine Auswertungsmethode unter Berücksichtigung aller in die Therapie aufgenommenen Patienten.
3. die Behandlung unabhängig vom zu erwartenden Therapieerfolg.
4. die Bemühungen des Therapeuten um bestmögliche Behandlung.
5. die probatorische Phase der Therapie.

Welche Aussagekombination ist richtig?

- A. Antworten 3 und 5 sind richtig.
- B. Antwort 1 ist richtig.
- C. Antwort 2 ist richtig.
- D. Antworten 1 und 2 sind richtig.
- E. Antworten 2 und 4 sind richtig.

5. Therapieabbrüche

1. gefährden die Prozessqualität einer psychotherapeutischen Einrichtung.
2. dürfen in nicht mehr als 5 % aller Behandlungsfälle auftreten.
3. werden bei Completer-Analysen berücksichtigt.
4. sind Ausdruck eines schlechten Therapeut-Patienten-Verhältnisses.
5. können auf Qualitätsmängel einer Einrichtung hinweisen.

Welche Aussagekombination ist richtig?

- A. Antworten 3 und 5 sind richtig.
- B. Antworten 1, 3 und 5 sind richtig.
- C. Antworten 4 und 5 sind richtig.
- D. Antwort 3 ist richtig.
- E. Alle Antworten außer 1 sind richtig.

Um Ihr CME-Zertifikat zu erhalten, schicken Sie bitte den ausgefüllten Fragebogen mit einem frankierten Rückumschlag bis zum 31.03.2009 an die nebenstehende Adresse. Später eintreffende Antworten können nicht mehr berücksichtigt werden.

Dipl.-Psych. Johanna Helmsen
Universität Bremen, ZKPR
Grazer Str. 6
D-28359 Bremen

FORTBILDUNGSZERTIFIKAT

Die Ärztekammer Niedersachsen erkennt hiermit 1 Fortbildungspunkt an.

«Evaluation von Psychotherapien aus der Perspektive von Qualitätssicherung und Qualitätsmanagement»

Die Antworten bitte deutlich ankreuzen!

1 2 3 4 5

- | | | | | | |
|----|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| A. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| B. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| C. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| D. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| E. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Stempel

Zeitschrift für Psychiatrie,
Psychologie und
Psychotherapie (ZPPP)

Verlag Hans Huber

Datum

Unterschrift

Ich versichere, alle Fragen ohne fremde Hilfe beantwortet zu haben.

Name

Berufsbezeichnung, Titel

Straße, Nr.

PLZ, Ort