



## **Qualitätsmonitoring in der ambulanten Psychotherapie**

---

# **ABSCHLUSSBERICHT**

---

**Modellvorhaben der Techniker Krankenkasse  
nach § 63 Abs. 1 SGB V**

## Impressum

---

**Wittmann, W.W., Lutz, W., Steffanowski, A., Kriz, D., Glahn, E.M., Völkle, M.C., Böhnke, J.R., Köck, K., Bittermann, A. & Ruprecht, T. (2011). Qualitätsmonitoring in der ambulanten Psychotherapie: Modellprojekt der Techniker Krankenkasse - Abschlussbericht. Hamburg: Techniker Krankenkasse.**

---

Der Abschlussbericht wurde im Auftrag der TK erstellt von den Evaluatoren des TK-Modellprojekts:

**Prof. Dr. Werner W. Wittmann, Dr. Andrés Steffanowski, Dr. David Kriz, Esther M. Glahn, Dr. Manuel C. Völkle**

Otto -Selz-Institut für Angewandte Psychologie

UNIVERSITÄT  
MANNHEIM

verantwortlich für die Kapitel (ansonsten gemeinsam)  
3.5, 3.7, 3.8.2, 3.9, 4.1, 4.2.1, 4.2.4 – 5.1.1, 5.2.-5.4.

**Prof. Dr. Wolfgang Lutz, Jan R. Böhnke, Katharina Köck, André Bittermann**

Universität Trier, Abteilung für Klinische Psychologie und Psychotherapie

 **Universität Trier**

verantwortlich für die Kapitel (ansonsten gemeinsam)  
3.6, 3.8.1, 4.2.2, 4.2.3, 5.1.2, 6, 7

---

### Kontaktadresse Mannheim

Dr. Andrés Steffanowski  
-Wissenschaftlicher Mitarbeiter-  
Otto-Selz-Institut für Angewandte Psychologie  
Arbeitsbereich 1: Methoden, Diagnostik & Evaluation  
Universität Mannheim, L 13, 15  
68131 Mannheim

[andres@steffanowski.de](mailto:andres@steffanowski.de)

### Kontaktadresse Trier

Dipl.-Psych. Katharina Köck  
-Wissenschaftliche Mitarbeiterin-  
Klinische Psychologie & Psychotherapie  
Fachbereich 1  
Universität Trier  
54286 Trier

[koeck@uni-trier.de](mailto:koeck@uni-trier.de)

---

Abkürzungen:

IG	Experimentalgruppe (Versicherte, die an der Studie teilnehmen und per Randomisierung dem modifizierten Gutachterverfahren zugeteilt wurden).
KG	Kontrollgruppe 1 (Versicherte, die an der Studie teilnehmen und per Randomisierung dem herkömmlichen Gutachterverfahren zugeteilt wurden)
KG2	Kontrollgruppe 2 (Versicherte, die nicht an der Studie teilnehmen, aber im Bereich der teilnehmenden KVen Westfalen-Lippe, Hessen und Südbaden versichert sind.
KG3	Kontrollgruppe 3 (Versicherte, die nicht an der Studie teilnehmen und auch nicht im Bereich der teilnehmenden KVen Westfalen-Lippe, Hessen und Südbaden versichert sind.
KV	Kassenärztliche Vereinigung
QS	Qualitätssicherung
RCT	Randomisierter Kontrollgruppenplan
TK	Techniker Krankenkasse

5-Datenbox-Konzeption:

PR-Box	Prädiktor-Box (vor der Intervention)
KR-Box	Kriterien-Box (nach der Intervention)
EVA-Box	Stakeholderinteressen (bestimmen die Kriterienauswahl der KR-Box)
ETR-Box	Experimentelle Treatment-Box (während der Behandlung)
NTR-Box	Nichtexperimentelle Treatment-Box (während der Behandlung)

Messzeitpunkte:

A	Messung bei Therapiebeginn „Aufnahme-Messung“
E	Messung bei Therapieende „Entlass-Messung“
K	Messung bei 1-Jahres-Katamnese „Katamnesemessung“
V	Verlaufsmessung

Stichproben:

T1:	N = 1.629 Therapeutendokumentationen zu Therapiebeginn (A-Messung)
T2:	N = 934 Therapeutendokumentationen zu Therapieende (A- und E-Messung)
P1:	N = 1.598 Patientendokumentationen zu Therapiebeginn (A-Messung)
P2:	N = 597 Patientendokumentationen zu Therapieende (A- und E-Messung)
P3:	N = 468 Patientendokumentationen zur 1-Jahres-Katamnese (A- und K-Messung)
P4:	N = 300 Patientendokumentationen (A-, E- und K-Messung liegt vor)

---

### Hinweis der Autoren:

Zur besseren Lesbarkeit wurden bei der Abfassung des Abschlussberichtes folgende Konventionen bei der Schreibweise in diesem Bericht gewählt:

- Es wird eine einheitliche Schreibweise z.B. „Therapeuten“ oder „Patienten“ verwendet, welche sowohl das männliche als auch das weibliche Geschlecht einbezieht.
- Die niedergelassenen ärztlichen und psychologischen Psychotherapeuten und Psychotherapeutinnen werden einheitlich als „Therapeuten“ bezeichnet.
- Die am Modellvorhaben teilnehmenden Patienten der zur Interventionsgruppe clusterrandomisierten Therapeuten (im Evaluationsplan als EX-Grp bezeichnet) werden als **IG** bezeichnet
- Die am Modellvorhaben teilnehmenden Patienten der zur Kontrollgruppe clusterrandomisierten Therapeuten (im Evaluationsplan als KON-Grp 1 bezeichnet) werden als **KG** bezeichnet
- Die nicht am Modellvorhaben teilnehmende Referenzgruppe (anonymisierte Leistungsdaten von TK-Versicherten) in den drei KV-Modellregionen Westfalen-Lippe, Hessen und Freiburg (im Evaluationsplan als KON-Grp 2 bezeichnet) wird als **KG2** bezeichnet.
- Die nicht am Modellvorhaben teilnehmende Referenzgruppe (anonymisierte Leistungsdaten von TK-Versicherten) außerhalb der drei KV-Modellregionen (im Evaluationsplan als KON-Grp 3 bezeichnet) wird als **KG3** bezeichnet.
- Beim Inventar für interpersonale Probleme (IIP) wird in der Studie durchgängig die deutsche Kurzversion mit 64 Items nach Horowitz, Strauss und Kordy (2000) verwendet und mit der Abkürzung **IIP-D** bezeichnet.

### Gesetzliche Grundlage

Ziele, Dauer, Art und allgemeine Vorgaben zur Ausgestaltung des TK-Modellvorhabens "Qualitätsmonitoring in der ambulanten Psychotherapie" sowie die Bedingungen für die Teilnahme von Versicherten wurden gemäß § 63 SGB V Abs. 5 vom Verwaltungsrat der TK am 24. September 2004 als 29. Nachtrag zur Satzung der Techniker Krankenkasse (TK) vom 1. Januar 2000 beschlossen und vom Bundesversicherungsamt (BVA) mit Bescheid vom 7. Januar 2005 genehmigt. Mit BVA-Bescheid vom 5. November 2007 wurde auch die vom Verwaltungsrat der TK am 28. September 2007 beschlossene Verlängerung des Modellvorhabens als 53. Nachtrag zur Satzung der Techniker Krankenkasse (TK) vom 1. Januar 2000 genehmigt (AZ II3-59015.0-4069/2003).

### Projektberatung und -betreuung

Die TK wurde bei der Konzeption des Modellvorhabens beratend unterstützt durch die Autoren dieses Berichts (Prof. Dr. Werner Wittmann und Prof. Dr. Wolfgang Lutz) sowie weiterhin durch Herrn Prof. Dr. rer. nat. Thomas Fydrich (seinerzeit Leiter des Zentrums für Psychologische Psychotherapie an der Universität Heidelberg, seit 2004 Professor für Psychotherapie und Somatopsychologie am Institut für Psychologie der Humboldt-Universität zu Berlin), Herrn Prof. Dr. phil. Rainer Richter (Leiter der Poliklinik für Psychosomatik und Psychotherapie am Zentrum für Innere Medizin des Universitätsklinikums Hamburg-Eppendorf, Präsident der Deutschen Fachgesellschaft für Tiefenpsychologisch fundierte Psychotherapie und Präsident der Bundespsychotherapeutenkammer) sowie Herrn Prof. em. Dr. Gerd Rudolf (seinerzeit Leiter des Instituts für Psychotherapieforschung und Qualitätssicherung an der Universität Heidelberg).

Als externes Trustcenter und als Datenstelle fungierte das Apparatezentrum der Hogrefe Verlag GmbH & Co. KG in Göttingen. Betreut wurde das Projekt dort von Herrn Karl-Heinz Schlawis und Herrn Wolf-Rüdiger Just.

Die Gesamtleitung des TK-Modellvorhabens lag bei Hauptverwaltung der Techniker Krankenkasse in Hamburg. Verantwortlich waren dort bis zum 30.06.2006 Herr Andreas Nagel und ab dem 01.07.2006 Herr Dr. med. Thomas M. Ruprecht, Fachbereich Versorgungsmanagement - Innovationen und Angebote.

---

### Danksagung

Ein so umfassendes Projekt wie das Modellvorhaben „Qualitätsmonitoring in der ambulanten Psychotherapie“ der Techniker Krankenkasse ist stets gekennzeichnet durch die direkten wie indirekten Beiträge einer Vielzahl beteiligter Einzelpersonen und Interessensgruppen.

Besonders hervorheben möchten wir an dieser Stelle jedoch die Arbeit des Wissenschaftlichen Beirats. Seinen Mitgliedern Herrn Prof. Dr. C. E. Scheidt (Vorsitzender), Herrn Dr. J. Brockmann, Herrn Prof. Dr. F. Caspar, Herrn Prof. G. Rudolf, Herrn Prof. Dr. U. Stangier und Herrn Dr. H. Vogel sowie den Beisitzenden des Wissenschaftlichen Beirats Herrn Dr. H. Herholz, Herrn Dr. P. Hinz, Herrn A. Kintrup, Herrn J. Klauenflügel, Frau Dr. R. Simon und Herrn M. Singmann verdanken die Autoren dieses Berichts und das gesamte Projekt eine Vielzahl wertvoller Anmerkungen und Vorschläge. Es soll erst gar nicht der Versuch einer Rekonstruktion der individuellen Beiträge zu diesem Projekt unternommen werden, doch steht außer Frage, dass durch die stets konstruktive Kritik des Wissenschaftlichen Beirates das Modellvorhaben als Ganzes wiederholt profitieren konnte.

## Inhalt

1	Zusammenfassung.....	11
1.1	Methodisches Vorgehen .....	11
1.2	Ergebnisse .....	12
2	Evaluationsplan vom 04.11.2003 .....	13
2.1	Ziele der Evaluation .....	13
2.2	Evaluationsdesign.....	14
2.2.1	Vorbemerkung.....	14
2.2.2	Messinstrumente .....	15
2.2.3	Vergleichsstrategien .....	19
2.3	Statistische Datenanalyse.....	23
2.3.1	Teststärkeanalyse und Fallzahlenberechnung .....	23
2.3.2	Datenaufbereitung und Datenreduktion .....	26
2.3.3	Deskriptive statistische Analysen .....	26
2.3.4	Inferenzstatistische Analysen .....	26
2.3.5	Exploratorische Analysen .....	26
2.4	Zentrale Hypothesen .....	26
2.4.1	Mittelwertsunterschiede.....	27
2.4.2	Korrelative Zusammenhänge und Strukturgleichungsverfahren .....	27
2.4.3	Gesamtbewertung.....	28
2.5	Qualitätssicherung der Daten .....	29
2.6	Durchführung der Evaluation .....	29
2.6.1	Zeitplan.....	29
2.6.2	Aufgaben des Evaluationsteams .....	29
2.7	Anlage A: Berechnung der Break-Even-Effektgröße .....	30
2.8	Anlage B: Einstellungen zur Qualitätssicherung.....	33
2.9	Anlage C: Evaluationsfragen aus Patientensicht .....	35
2.10	Anlage D: Evaluation der Rückmeldungen u. Handlungsempfehlungen .....	36
3	Realisierung des Projekts .....	37
3.1	Evaluationsdesign und Messzeitpunkte .....	37
3.2	Auswahl der Erhebungsinstrumente.....	38
3.3	Bewertungs- und Entscheidungshilfen.....	40
3.3.1	Standardisierte Ergebnisbewertung mit Effektgrößen .....	40
3.3.2	Multiple Ergebniskriterien.....	41
3.4	Randomisierung der Studientherapeuten.....	42

---

3.5	Schulung der Therapeuten .....	45
3.6	Hotline für Therapeuten .....	46
3.7	Datenerhebung und Fragebogenrücklauf .....	47
3.7.1	Fragebogenrücklauf der Therapeuten.....	48
3.7.2	Fragebogenrücklauf der Patienten .....	50
3.8	Gründe für fehlende Dokumentationen.....	52
3.8.1	Nichtteilnahme von Patienten am Modellvorhaben.....	52
3.8.2	Fehlende Abschluss- und Katamnesedokumentationen .....	54
3.9	Abschließende Bewertung.....	56
4	Stichprobenbeschreibung und Überprüfung der Randomisierung .....	58
4.1	Repräsentativität der Daten: Stichprobe TK.....	58
4.2	Therapeutenangaben bei Behandlungsbeginn: Stichprobe T1 .....	60
4.2.1	Basisdokumentation .....	60
4.2.2	Therapieeingangsdagnostik.....	67
4.2.3	Erfassung der Therapieziele mit dem Berner Inventar für Therapieziele (BIT-CT).....	78
4.2.4	Psychopharmakologische Behandlung .....	80
4.2.5	Therapeuten-Ratings bei Therapiebeginn .....	82
4.3	Patientenangaben bei Therapiebeginn (Stichprobe P1) .....	82
4.4	Zusammenfassende Bewertung .....	83
5	Wirksamkeit (Makroebene).....	84
5.1	Abschlussdokumentation der Therapeuten (Stichprobe T2) .....	84
5.1.1	Behandlungsverlauf und Situation bei Therapieende .....	84
5.1.2	Therapieziele zum Ende der Therapie .....	86
5.2	Angaben der Therapeuten zur OPD (optional, nur IG).....	87
5.2.1	Zu Therapiebeginn ausgewählte Konflikt- und Strukturbereiche .....	87
5.2.2	Konstruktion der Fokusskala .....	89
5.2.3	Prä-Post-Vergleich auf der Fokusskala .....	90
5.2.4	Validierung der Fokusskala anhand der Patientenmessungen .....	91
5.2.5	Zusammenfassende Bewertung .....	91
5.3	Therapeutenratings .....	92
5.3.1	Drop-Out-Analyse für die Therapeutenangaben.....	92
5.3.2	Status- und Ergebnisindikatoren der Therapeutenangaben .....	93
5.3.3	Vergleich zwischen Interventions- und Kontrollgruppe .....	94
5.3.4	Vergleich zwischen Therapiebeginn und Therapieende .....	100
5.3.5	Statistische Kontrolle möglicherweise vorhandener Confounder .....	101
5.4	Patientenratings .....	103
5.4.1	Drop-Out-Analyse für die Patientenangaben .....	103

5.4.2	Status- und Ergebnisindikatoren der Patientenangaben .....	105
5.4.3	Vergleich zwischen Interventions- und Kontrollgruppe (Stichprobe P2) .....	106
5.4.4	Statistische Kontrolle möglicherweise vorhandener Confounder (Stichprobe P2) .....	110
5.4.5	Vergleich zwischen Therapiebeginn und Therapieende (Stichprobe P2).....	111
5.4.6	Bias-Korrektur der Prä-Post-Effekte (Stichprobe P2) .....	112
5.4.7	Vergleich zwischen Interventions- und Kontrollgruppe (Stichprobe P3) .....	113
5.4.8	Statistische Kontrolle möglicherweise vorhandener Confounder (Stichprobe P3) .....	117
5.4.9	Vergleich zwischen Therapiebeginn und 1-Jahres-Katamnese (Stichprobe P3) .....	118
5.4.10	Bias-Korrektur der Prä-Post-Effekte (Stichprobe P3) .....	118
5.4.11	Verlaufsmessung auf dem multiplen Ergebnisindikator (Stichprobe P4) .....	119
5.4.12	Explorative Abschätzung der Kosten-Nutzen-Relation anhand der Effektgrößen .....	119
6	Entscheidungsregeln im Modellvorhaben (Mesoebene) .....	121
6.1	Entscheidungsregeln zu Beginn sowie im Verlauf .....	121
6.2	Evaluation der Entscheidungsregeln .....	127
6.2.1	Patienten mit und ohne Komorbidität bei den Eingangsdiagnosen .....	127
6.2.2	Psychometrische Diagnostik zu Therapiebeginn: Überprüfung der Grenzwerte.....	129
6.3	Verlaufserhebungen und Rückmeldungen.....	134
6.3.1	Rückmeldung zu Beginn der Therapie.....	134
6.3.2	Rückmeldung und Fortschritt im Therapieverlauf .....	135
6.3.3	Verlaufsprofile und Rückmeldung.....	137
6.3.4	Modifiziertes Gutachterverfahren im Therapieprozess und im Verlauf .....	144
7	Praktikabilität und Akzeptanz des Modellvorhabens (Mikroebene).....	145
7.1	Evaluation aus Therapeutensicht .....	145
7.1.1	Praktikabilitätseinschätzung durch die Therapeuten.....	145
7.1.2	Bewertung der eingesetzten Assessment-Instrumente durch die Therapeuten .....	147
7.1.3	Nutzung der Rückmeldungen durch die Therapeuten .....	147
7.1.4	Praktikabilitätseinschätzung durch die Therapeuten und Anzahl der Therapien .....	149
7.2	Evaluation aus Patientensicht .....	150
7.2.1	Bewertung des Modellvorhabens durch die Patienten .....	150
7.2.2	Akzeptanz der ambulanten Therapie durch die Patienten .....	152
7.3	Evaluation aus Gutachtersicht.....	155
8	Beantwortung der Studienfragestellungen.....	157
8.1	Hauptfragestellungen.....	157
8.2	Nebenfragestellungen .....	157
9	Diskussion.....	160
9.1	Effektivität .....	160
9.1.1	Angaben zum gesundheitlichen Befinden (EMEK-nonmonetär).....	160

9.1.2	Gesundheitsökonomische Daten (EMEK-monetär).....	160
9.2	Effizienz.....	160
9.2.1	Kosten-Nutzen-Verhältnis: Effektivität pro Therapiestunde.....	160
9.2.2	Dokumentationsaufwand für die Therapeuten.....	160
9.3	Nebenfragestellungen .....	161
9.3.1	Kategoriale Diagnostik.....	161
9.3.2	Ausgangsbelastung und Psychotherapiebedarf der Patienten .....	161
9.3.3	Verlaufsmessungen .....	161
9.3.4	Prä-Post-Effekte.....	162
9.3.5	Abschätzung der Kosten-Nutzen-Relation anhand der Prä-Post-Effektgrößen .....	163
9.3.6	Zufriedenheit der Patienten und Therapeuten .....	164
9.3.7	Therapiedauer und Varianzen .....	164
9.4	Methodische Aspekte.....	165
9.5	Fazit .....	166
10	Literaturverzeichnis.....	167
11	Anhang: Fragebogen Nichtteilnehmer - Sonstige Gründe für die Nichtteilnahme am Modellprojekt aus Therapeuten- und Patientensicht im Original-Wortlaut.....	170
12	Anhang: Offene Antworten zu den Evaluationsfragen aus Patienten- und Gutachtersicht im Original-Wortlaut.....	178
13	Anhang: Optimierung der Rückmeldungen im Projektverlauf.....	191
13.1	Anonymisierte Beispiel-Auswertung (grafische Rückmeldung) vor Modifikation der Rückmeldungen.....	193
13.2	Anonymisierte Beispiel-Auswertung (grafische Rückmeldung) nach Modifikation der Rückmeldungen.....	198

---

# 1 Zusammenfassung

## 1.1 Methodisches Vorgehen

Der vorliegende Abschlussbericht dokumentiert die Ergebnisse des Modellvorhabens der Techniker Krankenkasse (TK) nach § 63 Abs. 1 SGB V „Qualitätsmonitoring in der ambulanten Psychotherapie“ (MVPT). Es wurde im Jahr 2003 initiiert, zum 01.05.2005 gestartet, am 05.11.2007 (Bescheid des BVA) verlängert und zum 31.05.2011 mit dem vorliegenden Bericht abgeschlossen.

Zugrunde liegt ein clusterrandomisiertes Design mit 200 Therapeuten der Interventionsgruppe (IG) und 200 Therapeuten in der Kontrollgruppe (KG) in drei Modellregionen: Kassenärztliche Vereinigung (KV) Westfalen-Lippe, KV Hessen und KV Baden-Württemberg / Bezirk Südbaden.

Die 400 per Ausschreibung ermittelten, freiwillig am Modellvorhaben teilnehmenden Therapeuten wurden zunächst auf die IG und KG randomisiert (jeweils 200). Diese Therapeuten offerierten die (freiwillige) Studienteilnahme den insgesamt 4.452 TK-Versicherten (Grundgesamtheit), die sich bei ihnen während der Feldphase (01.05.2005 bis 30.06.2009) in psychotherapeutische Behandlung begaben und die Einschlusskriterien erfüllten. 1.708 Patienten (IG: 1.031, KG: 677) konnten für das Modellvorhaben gewonnen werden. Das Indikationsspektrum ist charakteristisch für die ambulante psychotherapeutische Versorgung. Am häufigsten sind depressive Störungen vertreten, gefolgt von Angsterkrankungen.

Zu Therapiebeginn (A), bei Therapieende (E) und zur 1-Jahres-Katamnese (K) wurden beim Patienten psychometrische Daten erhoben, u.a. BSI, IIP-D sowie ein störungsspezifisches Instrument. Letzteres wurde im Rahmen der Probatorik aufgrund der Hauptdiagnose für jeden Patienten individuell festgelegt (im MVPT meist der BDI aufgrund der Prävalenz depressiver Störungen). Alle Therapeuten füllten außerdem bei A und E eine Basisdokumentation aus.

Auf der Basis dreier psychometrischer Instrumente (BSI für allgemeine Symptomatik, IIP-D für interpersonale Probleme sowie einem störungsspezifischen Instrument je Patient) wurden zur Beurteilung der Krankheitslast bei Therapiebeginn Entscheidungsregeln ("klinisch auffällig" bzw. "unauffällig") sowie zur Veränderung der psychischen Problematik während der Therapie entsprechende Rückmeldungen für die Therapeuten der IG erstellt und evaluiert.

Bei Versicherten der IG genehmigte die TK im Falle klinisch auffälliger Werte eine Langzeittherapie ohne Einschaltung eines Gutachters. Nur in unklaren Fällen wurde eine (modifizierte) Begutachtung eingeleitet. Zudem erfolgten in der IG zu festgelegten Messzeitpunkten Verlaufsmessungen (V), deren Ergebnisse grafisch aufbereitet und den Therapeuten der IG umgehend rückgemeldet wurden. Optional bestand die Möglichkeit, die Ergebnisse auch mit den Patienten zu besprechen. Den grafischen Reports konnten die Therapeuten entnehmen, ob und wenn ja welche Veränderung der psychischen Problematik vorlag.

Bei Versicherten der KG war für die Bewilligung einer Langzeittherapie die traditionelle Begutachtung gemäß Psychotherapie-Richtlinien erforderlich. V-Messungen bzw. Rückmeldungen der Testergebnisse an die Therapeuten der KG (und damit auch der Patienten) fanden nicht statt.

Es bestand immer die Möglichkeit, das Therapieende schon vor Erreichen der bewilligten Sitzungszahl einzuleiten, in der IG z.B. bei Rückmeldung einer positiven Entwicklung (stabile Abnahme der psychischen Problematik).

Für den Therapiebeginn (A), fünf Verlaufsmessungen (V), die Abschlüsse (E) sowie die Katamnesen (K) liegen genügend Daten vor, um stabile Auswertungen mit einer hinreichenden Teststärke vornehmen zu können. Zusätzlich zur summativen Ergebnisevaluation auf Makro-, Meso- und Mikroebene (vgl. Evaluationsplan) wurde eine formative Evaluation zur Akzeptanz des Modellvorhabens aus Patienten-, Therapeuten- und Gutachtersicht sowie der Entscheidungsregeln für die grafischen Rückmeldungen durchgeführt. Die Auswertung zeigt eine gute Akzeptanz bei den Patienten sowie eine etwas zurückhaltendere Bewertung der Therapeuten (wenngleich die grafischen Reports als nützlich bewertet werden). Eher kritisch bewerteten hingegen die Gutachter das Modellvorhaben.

---

## 1.2 Ergebnisse

Die zentralen Fragestellungen des Projektes waren:

1. Führt der Einsatz von Qualitätsmonitoringinstrumenten des TK-Modells in ambulanten Psychotherapien zu einer höheren Ergebnisqualität (Effektivität) im Vergleich zur traditionellen Richtlinienpsychotherapie ?
2. Kann das TK-Modell die Effizienz der ambulanten Psychotherapie nachhaltig verbessern?

Es zeigen sich auch bei Berücksichtigung möglicher Confoundervariablen keine statistisch signifikanten Unterschiede zwischen IG und KG auf den zur Bewertung der Ergebnisqualität herangezogenen Skalen und multiplen Ergebniskriterien. Die erste Studienfrage ist somit zu verneinen. Auch die zweite Studienfragestellung muss mit nein beantwortet werden.

Bei Therapiebeginn sind die Mittelwerte in den Messinstrumenten für alle Instrumente vergleichbar mit den klinischen Normstichproben. Für 93% der Fälle ergeben sich mittlere bis schwere klinische Beeinträchtigungen. Lediglich bei 7% der Patienten lag eine geringe Beeinträchtigung vor, so dass bei Patienten der IG das modifizierte Gutachterverfahren eingeleitet wurde.

Bei den Zwischenmessungen (V) verzeichnen bereits über 30% Patienten eine Abnahme der Problematik, 20% eine Zunahme, bei rund 50% der Patienten kann (noch) keine Veränderung gemessen werden.

Bei den Abschlussmessungen (E) und Katamnesen (K) war bei ca. 65% der dokumentierten Fälle eine Abnahme der Problematik sichtbar, bei 28% keine signifikante Veränderung, bei 7% eine Verstärkung.

Insgesamt zeigen die explorativ berechneten, biaskorrigierten Prä-Post-Vergleiche große Effekte beim Vergleich zwischen Anfangs- (A) und Abschlussmessung (E) in einer Größenordnung von  $d = 1,0$ , die zudem in der 1-Jahres-Katamnese (K) stabil sind. Eine Kosten-Nutzen-Analyse unter Verwendung nichtmonetärer Ergebnismaße auf der Grundlage der Gleichung von Brogden, Cronbach und Gleser (Brogden, 1949; Cronbach & Gleser, 1965) erbrachte einen Nutzen zwischen 2 und 4 Euro pro investiertem Euro. Diese Ergebnisse dokumentieren einen beachtlichen Effekt der ambulanten Psychotherapie im Gesundheitswesen.

Mit Blick auf die externe Validität und der Interpretation aller Ergebnisse ist einschränkend zu beachten, dass nur etwa die Hälfte der von den Therapeuten im Studienzeitraum behandelten TK-Versicherten Patienten am Modellvorhaben teilgenommen hat. Zudem ist die Datenausfallrate zum E- und K-Zeitpunkt erheblich. So konnten für die Patientenangaben zum E-Zeitpunkt lediglich 597, für den K-Zeitpunkt lediglich 468 vollständige Datensätze gewonnen werden. Es ergaben sich allerdings keine Hinweise auf systematische Selektionseffekte in Abhängigkeit von der Gruppenzugehörigkeit IG/KG.

Fazit: Mit Blick auf die Prozess- und Ergebnisqualität psychotherapeutischer Behandlungen ergaben sich keine empirisch belegbaren, statistisch signifikanten Unterschiede zwischen IG und KG, weder durch die psychometrischen Messungen bei den Patienten bzw. Befragungen von Therapeuten und Patienten, noch durch den weitgehenden Verzicht auf das traditionelle Gutachterverfahren bei der Genehmigung von Langzeittherapien in der IG. Geht man davon aus, dass es sich beim Gutachterverfahren ebenso wie beim TK-Modell um eine Form der Qualitätsbeurteilung handelt, lassen sich im Vergleich beider Vorgehensweisen keine Unterschiede in den Behandlungsergebnissen feststellen.

Das Modellvorhaben liefert wertvolle Daten zur Versorgungsforschung, die auf eine hohe Wirksamkeit der ambulanten Psychotherapie hinweisen und damit ihre gesamtgesellschaftliche Bedeutung unterstreichen.

---

## 2 Evaluationsplan vom 04.11.2003

Die Hintergründe zur Entwicklung des Modellvorhabens "Qualitätsmonitoring in der ambulanten Psychotherapie", die Inhalte, die angewendeten Methoden und die Erläuterung der Durchführung sind in der nachfolgenden Projektbeschreibung ausführlich dargelegt (§ 63 SGB V). Der Gegenstand dieses Evaluationsplans nach § 65 SGB V sind die Fragestellungen und Hypothesen, das Evaluationsdesign und die Datenanalysen.

### 2.1 Ziele der Evaluation

Die beiden Kernziele des Projektes bestehen darin, im Rahmen der Richtlinienpsychotherapie verschiedene diagnostische Maßnahmen zur Qualitätssicherung zu erproben und zu vergleichen. Dies soll dazu beitragen, wissenschaftlich fundierte Befunde zur Prozess- und Ergebnisqualität der Psychotherapie im ambulanten Bereich zu erheben. Die Hauptfragestellung zielt auf die Erprobung neuer Qualitätssicherungsinstrumente und -maßnahmen in der ambulanten Psychotherapie und der Evaluation von deren Qualität und deren Wirtschaftlichkeit. Dabei wird das traditionelle QS-Verfahren mit dem QS-Verfahren nach dem TK-Modell verglichen.

Die QS-Maßnahmen des TK-Modells sehen die Kombination der folgenden vier Komponenten vor (vgl. Abbildung 1):

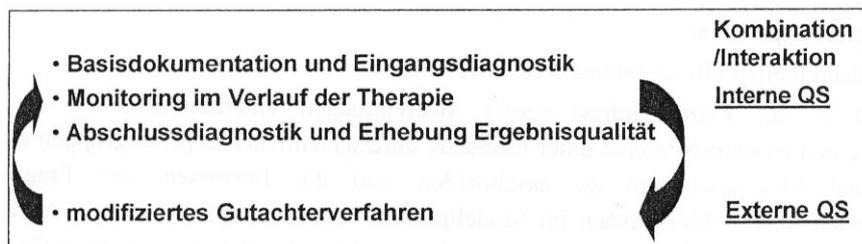


Abbildung 1. TK-Modell

Das vorliegende TK-Modell besteht aus einer Kombination von Maßnahmen der internen Qualitätssicherung (Eingangsdiagnostik, Verlaufsmonitoring und Abschlusserhebung) mit Instrumenten der externen Qualitätssicherung (modifiziertes Gutachterverfahren) und geht von einer positiven wechselseitigen Beeinflussung aus. Ein zeitnahe Austausch zwischen Therapeut (ggf. auch Patient) und Gutachter bzw. der unabhängigen Auswertungsstelle ist ein wichtiger Bestandteil des QS-Projektes.

Die zentralen Fragestellungen des Projektes sind:

1. Führt der Einsatz von Qualitätsmonitoringinstrumenten des TK-Modells in ambulanten Psychotherapien zu einer höheren Ergebnisqualität (Effektivität) im Vergleich zur traditionellen Richtlinienpsychotherapie ?
2. Kann das TK-Modell die Effizienz der ambulanten Psychotherapie nachhaltig verbessern?

Diese Hauptfragestellungen lassen sich nach der Makro-, Meso- und Mikroebene differenzieren, die im Rahmen der Abschlussevaluation untersucht werden und in Kapitel 1.6 (Hypothesen) im Einzelnen beschrieben werden.

## 2.2 Evaluationsdesign

Für das gesamte Evaluationsdesign wird die „Konzeption der fünf Datenboxen“ herangezogen, die eine Synthese unterschiedlicher Evaluationsschulrichtungen in sich vereint. Diese Konzeption hat sich bereits in vielen Anwendungsfeldern bewährt (Schmidt, 1991; Wittmann, 1990; Wittmann & Walach, 2001; Wittmann, Nübling, & Schmidt, 2002).

### 2.2.1 Vorbemerkung

Evaluationen sind Bewertungen, die naturgemäß bei allen beteiligten Gruppen Vorbehalte, Missbrauchsbedürfnissen, Hoffnungen und Ängste auslösen. Es ist deshalb wichtig, klar nach den verschiedenen Gruppen, den sogenannten Stakeholdern, zu unterscheiden und die verwendeten Kriterien, die die einzelnen Gruppen betreffen, fair und nachvollziehbar zu gestalten.

Die wichtigsten Stakeholdergruppen sind:

- die Versicherten (Psychotherapienutzer und Nichtnutzer)
- die ärztlichen und psychologischen Psychotherapeuten
- die Kassenärztlichen Vereinigungen
- die Hochschulambulanzen
- die Gutachter
- die wissenschaftlichen Fachvertreter aus Klinik und Forschung
- die Techniker Krankenkasse (TK)
- die Gesundheitspolitiker
- die gesundheitspolitisch interessierte Öffentlichkeit

Die TK hat in der Planungsphase große Anstrengungen unternommen, die zu verwendenden Kriterienmaße und Messinstrumente unter Einbezug unterschiedlicher Expertengruppen auf die einzelnen Therapien und Störungsformen zuzuschneiden. Damit werden die Interessen und Fragestellungen der unterschiedlichen Stakeholdergruppen im Modellprojekt berücksichtigt. Die aus dem Modellprojekt zu erwartenden Antworten dürften daher für alle Gruppen, gerade angesichts der Debatten um unser Gesundheitswesen, von hoher Relevanz sein.

Randomisierte Kontrollgruppenpläne (RCT's) sind einerseits in der Regelversorgung durch gesetzliche Rahmenbedingungen erschwert. Einzelne Stakeholdergruppen zeigen andererseits aus den unterschiedlichsten Interessen Vorbehalte gegenüber RCT's, meist mit Argumenten, die die externe Validität favorisieren. Andererseits wird das Potential von RCT's gerade in der Praxis, die durch Modellvorhaben nach § 63 Abs. I SGB V ermöglicht werden, nicht voll ausgeschöpft. Der hier unterbreitete Evaluationsplan versucht, beiden Gesichtspunkten gerecht zu werden. Es wird wo immer möglich randomisiert, dort wo dies nicht möglich ist, werden Kontrollstrategien entwickelt und gleichzeitig ein Maximum an Generalisierbarkeit der Ergebnisse angestrebt.

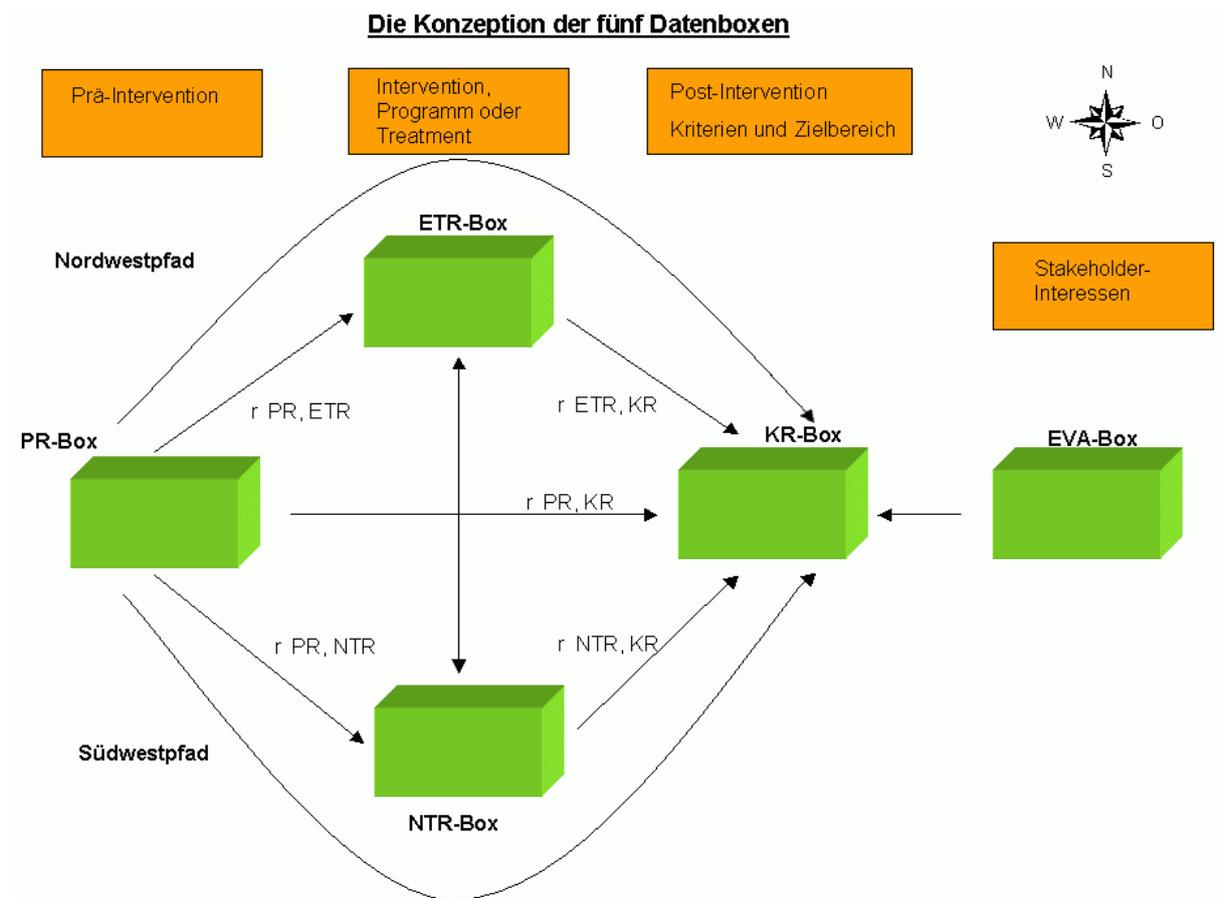
Das TK-Modellprojekt hat ausdrücklich diesen Fokus auf die reale Regelversorgung. Es wird aber erwartet, dass sich auch in diesem Bereich positive Effekte nachweisen lassen. Die ambulant tätigen Therapeuten wurden ja von denjenigen Therapeuten ausgebildet, die hinter den obengenannten erfolgreichen psychotherapeutischen Interventionen stehen. Break-Even-Point Analysen, d.h. „Wie groß muss ein Effekt sein, an dem sich Kosten und Nutzen gerade aufwiegen?“ weisen darauf hin, dass dieser Punkt bereits bei relativ kleinen Effekten liegt. Die Berechnung der Fallzahlen berücksichtigt diesen Sachverhalt und sieht einen Stichprobenumfang vor, der auch kleine Effekte mit genügend großer statistischer Teststärke entdecken kann, wenn sie vorhanden sein sollten.

Der Hauptfokus des Modellprojektes liegt vor allem darin, wie durch neue qualitätssichernde Maßnahmen (mit kurzen Reaktionszeiten) in der realen Versorgung die Effektivität weiter gesteigert werden kann. Diagnostische Informationen, die zur Therapiezielfindung dienen und ein Monitoring der Zielerreichung erlauben, sind in diesem Zusammenhang von hohem potenziellem praktischem Nutzen.

---

## 2.2.2 Messinstrumente

Die Auswahl und Zuordnung der einzelnen Messinstrumentgruppen kann konzeptuell über die Konzeption der fünf Datenboxen veranschaulicht werden (Abbildung 2). Diese Konzeption unterscheidet Datenquellen die vor einer Intervention (PR: Prädiktorbox), während einer Intervention (ETR: experimentelle Treatmentbox und NTR: nichtexperimentelle Treatmentbox) und nach einer Intervention (KR: Kriterienbox) erhoben werden sollen. Die fünfte Datenbox (EVA: Evaluationsbox) umfasst die einzelnen Stakeholdergruppen mit ihren unterschiedlichen Interessen und Zielsetzungen. Informationen zu dieser Datenbox müssen zu Beginn jeder Evaluationsstudie erfasst werden, um die Konstruktvalidität der Messinstrumente und Akzeptanz der Ergebnisse schon im Vorfeld der eigentlichen Studie zu maximieren.



**Abbildung 2.** Ein umfassendes Forschungskonzept zur systematischen Programmevaluation. PR-Box: Prädiktor-Box. ETR-Box: Experimentelle Treatment-Box. NTR-Box: Nichtexperimentelle Treatment-Box. KR-Box: Kriterien-Box.  $r_{PR, ETR}$ : Korrelation zwischen PR- und ETR-Box etc.

Die Anordnung der vier erstgenannten Datenboxen folgt einer klaren zeitlichen Reihenfolge und erlaubt auch für nichtexperimentelle korrelative Datenkonstellationen kausale Aussagemöglichkeiten über die statistische Methode der Pfadanalyse. Diese Methode kann entweder klassisch über hierarchische Regressionsanalysen oder über die modernen Methoden der Strukturgleichungsverfahren (SEM: structural equation modeling) durchgeführt werden.

### 2.2.2.1 Datenquellen zu Beginn der ambulanten Psychotherapie (PR-Box)

Tabelle 1 und Tabelle 2 geben einen Überblick über die diagnostischen Verfahren, die von Patienten bzw. Therapeuten bearbeitet werden.

**Tabelle 1. Diagnostische Verfahren und Erhebungen für Patienten**

	Beginn	Zwischen	Ende	1-Jahres-Katamnese
Basisdokumentation Anfang	x			
Brief Symptom Check List (BSI)	x	(x)	x	x
Interpersonale Probleme (IIP-D)	x	(x)	x	x
Störungsspezifische Verfahren (BDI, AKV, HZI, SOMS oder EDI)	x	(x)	x	x
Therapeutische Arbeitsbeziehung (HAQ-II)	x	(x)	x	
Gesundheitsbezogene Lebensqualität (SF-12)	x	(x)	x	x
Basisdokumentation Ende			x	

Anmerkungen. (x) Nur bei der Interventionsgruppe erhoben.

**Tabelle 2. Diagnostische Verfahren und Erhebungen für Therapeuten**

	Beginn	Zwischen	Ende
Therapeutenbasisbogen (einmalig)	x		
Basisdokumentation Beginn	x		
SKID Screening	x		
ICD-Diagnostik nach ICDL	x		
OPD Achsen Konflikt und Struktur (TP/AT)	x		
Beeinträchtigungsschwere (BSS)	x	(x)	x
Problemrelevante Lebensbereiche	x	(x)	x
Beeinträchtigung Funktionsniveau (GAF)	x	(x)	x
Therapeutische Arbeitsbeziehung (HAQ-I)	x	(x)	x
Therapieziele (VT)	x	(x)	x
Heidelberger Umstrukturierungsskala (TP/AT)		(x)	x
Basisdokumentation Ende			x

Anmerkungen. VT – Verhaltenstherapie, TP – Tiefenpsychologische Therapie, AT – Analytische Therapie. (x) Nur bei der Interventionsgruppe erhoben

### 2.2.2.2 Datenquellen zur Operationalisierung der Treatment-Boxen

In der ETR-Box operationalisiert ein einfacher Dummy-Vektor den Kontrast der Experimentalgruppe (Wert 1) zu Kontrollgruppe 1 (Wert 0). Bei erfolgreicher Randomisierung ist die Korrelation  $r_{PR, ETR}$  zwischen allen PR-Box Variablen und diesem Dummy-Vektor gleich Null, d.h. Experimental- und Kontrollgruppe unterscheiden sich nicht voneinander. Die Kontrastvektoren zur Operationalisierung der Kontrollgruppen 2 und 3 in Relation zueinander und zur Experimentalgruppe und der Kontrollgruppe 1 müssen der NTR-Box zugewiesen werden, da die Teilnahme am Modellvorhaben insgesamt ja freiwillig erfolgt (unabhängig davon, ob der betreffende Therapeut nun der Experimental- oder Kontrollgruppe zugewiesen wird). Hier kann über die Korrelationen  $r_{PR, NTR}$  untersucht werden, welche Selektionseffekte bei den beiden Studiengruppen gegenüber der Kontrollgruppe 2 und 3 zu berücksichtigen sind.

Neben den Dummy-Variablen zur Operationalisierung der Gruppenunterschiede können die Informationen die für die Experimentalgruppe erhoben werden zusätzlich in die ETR-Box eingebracht werden. Hierzu zählen alle Informationen die im Therapeutenbogen (siehe Projektbeschreibung) und im HAQ (Bassler, Potratz, & Krauthauser, 1995) erhoben werden. Diese Informationen können als individuelle Unterschiede der Therapeuten und als individuelle Unterschiede der Therapeut/Patientenbeziehung mit den Kriterienmaßen der KR-Box, sichtbar an den Korrelationen  $r_{ETR, KR}$  analysiert werden. Diese Variablen können als Therapiedosis oder Therapiequalitätsmerkmale aufgefasst werden, sie liefern Informationen wie die Höhe der Effekte möglicherweise mit solchen Merkmalen variieren.

Die Variationen des Gutachterverfahrens alt und neu können ebenfalls den Treatment-Boxen zugeordnet werden. Die Art der Variationen sind aber nicht experimentell variiert, sie werden so abgebildet wie sie in der Praxis entstehen und werden daher der NTR-Box zugeordnet. Die Gutachter für die drei Therapievarianten werden aber den zu begutachtenden Therapiefällen innerhalb jeder Therapierichtung per Zufall zugeordnet.

### 2.2.2.3 Datenquellen zur Operationalisierung der Kriterien-Box

In die KR-Box werden alle Variablen eingeordnet, für die Messwerte am Ende der Behandlung und zur 1-Jahres-Katamnese erhoben werden. Diese Messwerte können als Rohwerte oder als Differenzwerte zum Assessment in der PR-Box verwendet werden. Differenzwerte können für die Experimentalgruppe für alle Variablen die sowohl in der PR-Box als auch in der KR-Box erhoben werden berechnet werden. Für die Kontrollgruppe erfolgt die Prä-Messung (PR-Box) nur für BSI und IIP-D, so dass dort Differenzwerte nur für diese Variablen berechenbar sind. Weiterhin werden diese gemeinsamen Variablen verwendet, um den Randomisierungserfolg auch tatsächlich überprüfen zu können.

Die Vielzahl der Kriterienmaße kann entweder über theoriegeleitete Aggregation oder über faktorenanalytische Verfahren zu sogenannten einmaligen, messzeitpunktspezifischen, multiplen Ergebniskriterien (EMEK) verarbeitet werden. Die meisten der einzelnen Ergebnismaße korrelieren positiv untereinander, was eine solche Strategie rechtfertigt. Solche EMEK's zeichnen sich dann durch höhere Reliabilität und Validität aus. In den obengenannten Programmevaluationsstudien (Zauberberg und EQUA) wurden solchermaßen aggregierte Kriterien, die zudem noch differenziert nach den Interessen verschiedener Stakeholdergruppen aggregiert werden können, mit großem Erfolg eingesetzt. Zeigen sich z.B. vergleichbare Effekte auf einem rein monetär bewertbaren EMEK, ebenso wie auf einem EMEK das nur psychosoziale Kriterien oder nur Befindlichkeitsmerkmale umfasst, so steigen Glaubwürdigkeit der Ergebnisse über Effektivität und Effizienz eines Programms. Kontroversen über den Ertrag eines Modellprojektes oder Programms werden dann abgeschwächt und die Akzeptanz über das Umsetzungspotential steigen.

In die monetäre Analyse gehen neben den Behandlungskosten für die Psychotherapiesitzungen die Kosten ein, die im Rahmen des klassischen Gutachterverfahrens und des modifizierten Gutachterverfahrens entstehen. In der Abschlussbetrachtung ist zu berücksichtigen, dass die Stückkosten der einzelnen diagnostischen Instrumente bei einem Routineeinsatz deutlich sinken dürften und dass dann der Break-

---

Even-Point (Effektgröße  $d$ ) in der Routineversorgung nach Abschluss des Modellprojektes noch weiter absinkt.

#### 2.2.2.4 Datenquellen zur Operationalisierung der EVA-Box

Die Evaluationsbox ganz rechts in Abbildung 2 enthält die Interessen und Zielsetzungen der Stakeholder. Bei der Auswahl der Messinstrumente und der Wahl der Varianten des Evaluationsplanes wurden diese Interessen bereits in der Planungsphase berücksichtigt (siehe oben). Entscheidend ist jedoch auch wie das gesamte Modellprojekt mit seinen qualitätssichernden Maßnahmen von der Gruppe der Patienten und Therapeuten bewertet und eingeschätzt wird. Zum Abschluss der Therapie (Therapeut und Patient) und in der Katamnese (Patient) werden deshalb Fragebogen zur Zufriedenheit und Akzeptanz des Qualitätsmonitorings eingesetzt. Hierzu werden Fragebogen wie ZUF-8 (Schmidt, Lamprecht, & Wittmann, 1989) und das amerikanische SERVQUAL-System, speziell adaptiert auf das Modellprojekt, eingesetzt. Diese Adaptionen müssen noch erfolgen und werden vom Evaluationsteam in der Anfangsphase des Projektes entwickelt.

#### 2.2.2.5 Variationen des Datenerhebungsmodus

Die gesamte psychometrische Datenerhebung der Selbstbeurteilung soll nach Möglichkeit internetbasiert über das Hogrefe Testsystem erfolgen. Für diejenigen Teilnehmer, welche der klassischen Papiervariante den Vorzug geben, soll auch die Möglichkeit bestehen, die ausgefüllten Fragebögen an die unabhängige Datenstelle zu senden, wo diese dann eingegeben und der Studiendatenbank hinzugefügt werden. Die Art der Datenerhebung wird dokumentiert und als mögliche Kovariate im Hinblick auf alle Datenboxen analysiert.

#### 2.2.2.6 Übersicht der Messinstrumente

Die Operationalisierung der verschiedenen Datenboxen ist in der Tabelle 3 im Überblick dargestellt.

Tabelle 3. Überblick zur Operationalisierung der Datenboxen

PR-Box	ETR-Box	NTR-Box	KR-Box	EVA-Box
<b>Prädiktoren</b>	<b>Experimentelle Treatment-Box (Intervention)</b>	<b>Nichtexperiment. Treatment-Box (Kontrollgruppe)</b>	<b>Kriterien nach der Intervention</b>	<b>Evaluationsbox</b>
Basisdokumentation Patient und Therapeut	Dummy-Vektor =1/0	Dummy-Vektoren für zusätzliche Kontrollgruppen	Basisdokumentation Patient und Therapeut	Fragestellungen der verschiedenen Stakeholder
Diagnose			Diagnose	
OPD			OPD	
Umstrukturierungsskala			Umstrukturierungsskala	
BSI			BSI	
IIP-D			IIP-D	
Spezifischer Test			Spezifischer Test	
HAQ			HAQ	
SF-12			SF-12	

### 2.2.3 Vergleichsstrategien

Die Evaluation wird soweit wie möglich als randomisiertes Kontrollgruppendesign (RCT= randomized clinical trial) durchgeführt. Bestimmte Rahmenbedingungen, die durch die Einbettung des Modellprojektes in die reale Versorgung mit ambulanter Psychotherapie bedingt sind, verhindern jedoch eine perfekte Randomisierung für alle Fragestellungen. Bei diesen Problemen werden jedoch quasiexperimentelle Kontrollstrategien eingesetzt, die ein Höchstmaß an interner Validität und damit an kausalen Aussagemöglichkeiten erlauben. Forschungsleitend sind insgesamt betrachtet die Gütestandards der Versuchsplanung nach Cook und Campbell (1979) bzw. Shadish, Cook und Campbell (2002):

- interne Validität
- statistische Schlussfolgerungsvalidität
- externe Validität
- Konstruktvalidität

Durch umfangreiche Rekrutierungsbemühungen wird versucht, möglichst viele Therapeuten und damit implizit auch ihre TK-Patienten zur Teilnahme am Modellprojekt zu motivieren. Auf der Grundlage der statistischen Fallzahlenberechnung wurden drei Kassenärztliche Vereinigungen (Hessen, Südbaden und Westfalen-Lippe) ausgewählt, in denen die Rekrutierung von Therapeuten stattfinden soll.

#### 2.2.3.1 Randomisierungsstrategien zur kausalen Wirksamkeit des Qualitätsmonitorings

Die Auswahl der Therapeuten, die das Qualitätsmonitoringsystem anwenden sollen, erfolgt durch eine Zufallsauswahl der Therapeuten, die ihre Teilnahmebereitschaft bekundet haben. Die Liste der teilnahmebereiten Therapeuten wird dem Evaluationsteam von der KV in Form einer Liste anonymisierten Therapeutencodes übermittelt, so dass eine Verblindung sichergestellt ist. Sodann wird die Randomisierung vom Evaluationsteam in zwei Gruppen durchgeführt.

Die eine Gruppe stellt die Experimentalgruppe (EX-Grp) dar, nachfolgend auch einfach als IG bezeichnet. Die Kontrollgruppe 1 (KON-Grp 1) umfasst die teilnahmebereiten Therapeuten, die nicht der Experimentalgruppe zugewiesen wurden (vgl. Abbildung 3) und wird nachfolgend auch einfach als KG bezeichnet. Bei dieser Gruppe wird die psychometrische Diagnostik ebenfalls zu Beginn, am Ende der Therapie und bei der 1-Jahres-Katamnese in einer reduzierten Form (Basiserhebung mit BSI und IIP-D) durchgeführt. Dadurch ergeben sich zwei Vergleichsmöglichkeiten von Mittelwertsunterschieden auf den verwendeten Kriterienmaßen. Dieser Vergleich von IG und KG erlaubt einen kausalen Nachweis der Wirksamkeit der modifizierten Vorgehensweise im Modellprojekt mit der bisherigen Praxis. Durch erfolgreiche Randomisierung werden alle rivalisierenden Alternativverklärungen der Effekte ausgeschlossen. Der Randomisierungserfolg wird im Sinne eines Manipulationsschecks über Informationen aus der Eingangsdiagnostik (PR-Box), die für Experimental- und Kontrollgruppe vergleichbar erhoben werden (BSI und IIP-D), geprüft.

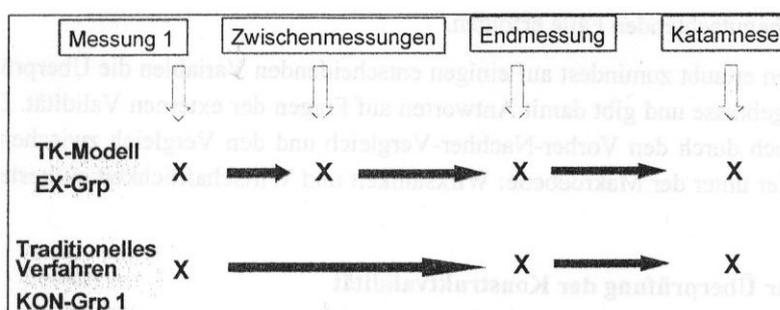


Abbildung 3. Randomisierungsstrategien.

Weiterhin werden Informationen aus dem Therapeuten-Praxisbogen und Kennwerte der Abschlussdokumentation am Behandlungsende, die sich jedoch auf den Behandlungsbeginn beziehen, wie z.B. Diagnosen, herangezogen. Durch diese Maßnahmen kann die Vergleichbarkeit von Therapeuten und deren teilnahmebereiten Patienten zwischen beiden Gruppen ausführlich überprüft und sichergestellt werden. Diese Strategie hilft, die interne Validität des Evaluationsplanes zu kontrollieren und zu maximieren.

### **2.2.3.2 Nichtexperimentelle Kontrollstrategien zur Maximierung der Generalisierbarkeit**

Es ist aber nicht auszuschließen, dass sich die teilnahmemotivierten Therapeuten und deren teilnahmemotivierten Patienten systematisch von den nichtteilnahmebereiten Gruppen unterscheiden. Teilnahmebereite Therapeuten könnten kompetenter sein und/oder ihre teilnahmebereiten Patienten eine höhere Psychotherapiemotivation aufweisen. Dies gefährdet wiederum vor allem die Generalisierbarkeit (externe Validität) der Ergebnisse auf das gesamte reale Versorgungssystem ambulanter Psychotherapie. Eine Randomisierung aller niedergelassenen Psychotherapeuten und ihrer Patienten ist jedoch nicht möglich und durchsetzbar. Voraussetzung für alle Modellprojekte ist die Einverständniserklärung (informed consent) von Therapeut und Patient.

Um diese Gefährdung der externen Validität abschätzen und zumindest partiell überprüfen zu können, werden zwei weitere Kontrollgruppen eingesetzt. Die TK zieht in Abstimmung mit dem TK-Datenschützer eine nach Therapierichtungen (VT, TP, AT) geschichtete Zufallsstichprobe aus der Gruppe derjenigen Therapeuten, die nicht an der Studie teilgenommen haben. Sie liefert die Kernkriterien der Versicherten, die von dieser Gruppe behandelt wurden, ebenfalls in anonymisierter Form an das Evaluationsteam. Diese Gruppe von Versicherten sollte zu gleich großen Teilen aus Versicherten der teilnehmenden KV'en (KON-Grp 2) und aus nicht ausgewählten KV'en (KON-Grp 3) stammen. Diese beiden Gruppen bezeichnen wir als Kontrollgruppen 2 und 3. Für diese beiden Gruppen stehen keine Daten aus dem QS-System zur Verfügung. Die TK verfügt jedoch über personenbezogene Leistungsdaten, z. T. in EDV-Form (Data-Warehouse), für alle vier Gruppen, d.h. Experimentalgruppe und die drei Kontrollgruppen. Diese Daten lassen sich nach direkten Kosten wie medikamentöse Therapie (Arzneimittel), stationäre Behandlung (Krankenhaus), Kuren, Krankengeld und indirekte Kosten (Arbeitsunfähigkeitstage) gliedern. Alle kostenrelevanten Daten der vier Gruppen werden für den Zeitraum ein Jahr vor der ambulanten Psychotherapie und den Katamnese-Zeitraum ein Jahr nach Ende der Behandlung personenspezifisch aggregiert. Diese Daten werden dem Evaluationsteam in anonymisierter Form zur Analyse zur Verfügung gestellt.

Beim Randomisierungsprozess Gutachten erhält jeder Gutachter einen Code, und die Randomisierung wird vom Evaluator geschichtet nach der Therapierichtung vorgenommen. Bei jedem notwendigen Gutachten wird der Gutachter durch einen Zufallsgenerator bestimmt. Diese Reihenfolge kann a-priori über die Laufnummer (1-xxxx) der zu begutachtenden Fälle erfolgen.

Der Vergleich der vier Gruppen erlaubt zumindest für einige entscheidende Variablen die Überprüfung der Generalisierbarkeit der Ergebnisse und gibt damit Antworten auf Fragen der externen Validität. Diese Datenbasis ermöglicht aber auch durch den Vorher-Nachher-Vergleich und den Vergleich zwischen den Gruppen, die Beantwortung der unter der Makroebene (Wirksamkeit und Wirtschaftlichkeit) aufgestellten Fragestellungen.

### **2.2.3.3 Kontrollstrategien zur Überprüfung der Konstruktvalidität**

Konstruktvalidität kann bezüglich Ursachen und Effekten unterschieden werden und bezieht sich vor allem auf die eingesetzten Messinstrumente. In der Planungsphase wurden Experten aus den unterschiedlichen Therapierichtungen eingesetzt, um die Inhaltsvalidität der Verfahren zu optimieren. Diese Instrumente stellen einen Kompromiss dar, der von allen Gruppen im Konsensverfahren akzeptiert wurde. Die Messinstrumente werden bisher nicht routinemäßig in der Regelversorgung eingesetzt. Vor allem aus psychoanalytischer Stakeholderperspektive wurden in der Planungsphase des Modellvorhabens Bedenken gegen quantifizierende Instrumente dahingehend geäußert, dass eine Dominanz psychomet-

---

rischer, quantitativer Verfahren die Konstruktvalidität psychodynamischer Ursachen und Effekte nicht adäquat abbilden könne. In unseren eigenen Programmevaluationsstudien erwies sich dieser Einwand bislang als unbegründet (Wittmann, Nübling, & Schmidt, 2002), so konnten der Stellenwert und die Konstruktvalidität der psychometrischen Messinstrumente für psychoanalytische Interventionen über beachtliche Effektgrößen mehrfach nachgewiesen werden. Diese Programmevaluationen untersuchten stationäre psychoanalytische Behandlungsprogramme, wie sie in psychosomatisch ausgerichteten Großkliniken mit sechs bis achtwöchigen Aufenthalt durchgeführt werden.

#### **2.2.3.4 Implementation und Evaluation des modifizierten Gutachterverfahrens**

Das gesamte Modellprojekt besteht aus einer Kombination zweier qualitätssichernder Innovationen, erstens der Einführung eines umfassenden Inventars an psychometrischen Messinstrumenten zum Monitoring des Therapieverlaufs und der Zielerreichung, sowie zweitens einer darauf aufbauenden Neuorientierung des Gutachterwesens.

Hier erwarten wir Antworten auf den Zielbereich der Mesoebene: Diagnostische und adaptive Entscheidungsstrategien. Die Entscheidungsregeln werden in einer ersten formativen Evaluationsphase explorativ entwickelt und optimiert, um dann in der zweiten Phase summativ evaluiert zu werden.

##### **2.2.3.4.1 Formative Evaluation der Entscheidungsregeln, des grafischen Rückmeldemodells und des Gutachterverfahrens**

In einem ersten Evaluationsschritt der sogenannten formativen Evaluation sollen die spezifischen empirischen Entscheidungsregeln des TK-Projektes (vgl. Lutz et al., 2005) evaluiert und optimiert werden. Dies erfolgt sobald 200 Patienten das TK-Projekt mit mindestens 20 Sitzungen durchlaufen haben. Die Daten dieser Patienten werden eingesetzt, um die Normen und Berechnungen, welche den Entscheidungsregeln zugrunde lagen, anhand des TK-Datensatzes anzupassen. Die Grundlage der Entscheidungsregeln für die Stabilitätsabschätzungen war bisher ein Datensatz der Psychotherapeutischen Ambulanz der Universität Bern, Schweiz.

Die klinischen Grenzwerte sowie die Entscheidungsregeln hinsichtlich der Entwicklungsverläufe wurden auf der Grundlage dieses Datensatzes entwickelt. Wenn auch dieser Datensatz in vielen Punkten repräsentativ für ambulante Psychotherapie in der Bundesrepublik Deutschland sein dürfte, so sind doch auch Abweichungen bei diesem Klientel zu erwarten. Aus diesem Grund sollten die Grenzwerte, klinischen Normen sowie die Entscheidungsregeln hinsichtlich des bis zu diesem Zeitpunkt zur Verfügung stehenden Datensatzes angepasst werden.

##### **2.2.3.4.2 Summative Evaluation der Entscheidungsregeln, des grafischen Rückmeldemodells und des Gutachterverfahrens**

Die folgenden vier summativen Evaluationsschritte werden im Anschluss an die formative Evaluation erarbeitet:

- a) Evaluation der Entscheidungsregeln und Erstellung eines endgültigen Entscheidungsalgorithmus. Das forschungsstrategische Vorgehen entspricht dem Vorgehen der formativen Evaluation. Die Regeln werden anhand des Gesamtdatensatzes angepasst. Wichtig ist, dass nun auf der Basis sämtlicher zur Verfügung stehender Patientenangaben, die Grenzwerte, klinischen Normen sowie die Entscheidungsregeln evaluiert werden können und die Erfahrungen für eine routinemäßige Anwendung nutzbar gemacht werden können. Zur Durchführung bedarf es einer kontinuierlichen Prüfung und fortlaufender Updates des Datensatzes durch das Forschungsteam.
  - b) Statistischer Vergleich zur Bedarfsgerechtigkeit der Indikationsentscheidungen. Neben der im vorherigen Abschnitt beschriebenen Erhebung zu den Einstellungen der Therapeuten sowie der Gutachter zur QS und dem neuen System wird zusätzlich ein Vergleich anhand der psychometrischen Informationen zu dem alten und neuen Gutachterverfahren durchgeführt werden. Dazu
-

erfolgt ein Vergleich hinsichtlich der Behandlungsdauer unter Berücksichtigung der Ausgangslage der Patienten sowie den Behandlungsbedingungen. Zentrales Ziel ist eine optimale Indikation für die Patienten. Das alte Gutachtersystem verteilt bestimmte Behandlungskontingente relativ unabhängig von der konkreten (empirisch nachvollziehbaren) Ausgangslage der Patienten. Daher muss analysiert werden, inwieweit sich durch das modifizierte Gutachtersystem eine Optimierung der Indikationsentscheidungen dahingehend ergibt, dass bedürftige Patienten eine aufwendigere Therapie erhalten und weniger bedürftige Patienten eine weniger aufwendigere Therapie.

- c) Befragung zur Evaluation von QS-Maßnahmen und des Gutachtersystems an Therapeuten und Gutachtern. In einem weiteren Evaluationsschritt wird eine Befragung zur Einschätzung des neuen Gutachterverfahrens und der QS-Maßnahmen an den Therapeuten sowie den Gutachtern durchgeführt. Dazu wird ein speziell zur Evaluation von Qualitätssicherungsmaßnahmen in der Psychotherapie entwickelter Fragebogen modifiziert und eingesetzt (Lutz et al., 1997). Dieser Fragebogen wird durch Fragen zu dem im Modellprojekt der TK neuartig eingesetzten Gutachterverfahren, den Qualitätssicherungsmaßnahmen, den Handlungskonsequenzen, der Übereinstimmung zwischen der klinischen Sicht und den empirischen Handlungsempfehlungen sowie zur Durchführung und Praxisrelevanz ergänzt. Zentral soll in diesem Schritt die Sicht der Therapeuten zu den Stärken und Schwächen sowie der Effektivität und Effizienz (altes versus neues Verfahren) herausgearbeitet werden. Aus diesen Angaben lassen sich wiederum wertvolle Schlussfolgerungen für ein zukünftig optimal einsetzbares Qualitätssicherungssystem und Gutachterverfahren ziehen (siehe Anlage B: Einstellungen zur Qualitätssicherung).
- d) Auswertung der Evaluationsfragen aus der Patientensicht. In die Erhebungen zum Therapieende werden zusätzliche Fragen zu den Effekten der Erhebungen, zur Qualität der Fragebogen, der Ergebnisabschätzung durch die Patienten sowie der Patientenzufriedenheit integriert. Anhand der Analysen zu diesen Fragen können wertvolle Informationen zur Patientensicht hinsichtlich des TK-Projektes bzw. des neuen Evaluationssystems ermittelt werden (siehe Anlage C: Evaluationsfragen aus Patientensicht).
- e) Effekte der grafischen Rückmeldung und Handlungsempfehlungen für die Therapeuten und Gutachter. In diesem Schritt wird über die globale Abschätzung zu den Effekten aus Therapeuten- und Gutachtersicht eine gezielte Evaluation am jeweiligen Patienten erfolgen. Die Angaben, welche in diesem Evaluationsschritt ausgewertet werden, befassen sich mit den Effekten, welche das grafische Rückmeldesystem sowie die darin integrierten Handlungsempfehlungen für das konkrete Handeln der Therapeuten haben. Angepasste Fragen werden in das neue Gutachterverfahren integriert und erlauben so die Abschätzung von konkreten Effekten, welche die Rückmeldungen für die Gutachter haben bzw. welche Konsequenzen daraus gezogen werden. Weiterhin spielen hier Fragen zur Konsistenz zwischen empirischen Handlungsregeln und klinischem Eindruck bzw. der Nützlichkeit beider Datenquellen eine zentrale Rolle (siehe Anlage D: Evaluation der Rückmeldungen und Handlungsempfehlungen).

Tabelle 4 zeigt das Analyseschema der Mesoebene mit den drei Stakeholdern und der Konsistenz zwischen der klinischen und der empirischen Beurteilung.

**Tabelle 4. Analyseschema Mesoebene: Qualität der Indikationsentscheidung in der ambulanten Psychotherapie (Beginn/Verlauf/Ende)**

Stakeholder	Studiengruppe	Klinischer Eindruck	Empirische Bewertung	Konsistenz	Stärken / Schwächen
<b>Patient</b>	TK-Modell	x	x	x	x
	Gutachterverfahren	x			x
<b>Therapeut</b>	TK-Modell	x	x	x	x
	Gutachterverfahren	x			x
<b>Gutachter</b>	TK-Modell	x	x	x	x
	Gutachterverfahren	x			x

### 2.2.3.5 Kontrollstrategien zur Überprüfung der Akzeptanz des Qualitätsmonitorings, des alten und des neuen Gutachterverfahrens

Die Messinstrumente ZUF-8 und SERVQUAL werden für die EX-Grp und KON-Grp 1 eingesetzt. Diese Messinstrumente erlauben die Analyse von Mittelwertsunterschieden zwischen den drei Therapierichtungen in Bezug auf Ertrag und Akzeptanz. Korrelative Analysen zu PR-, ETR- und KR-Box Variablen erlauben die Beantwortung von Fragen, in welchem Ausmaß Akzeptanz und Zufriedenheit mit Eingangs-, Therapie- und Erfolgsmerkmalen korrelieren. Dieser Teil des Evaluationsdesigns beantwortet die Fragestellungen des Zielbereiches Mikroebene: Praktikabilität und Akzeptanz.

Tabelle 5. Messinstrumente auf Meso- und Mikroebene

	Patient	Therapeut	Gutachter
Einstellungen zur Qualitätssicherung, Anlage B		x	(x)
Evaluationsfragen aus Patientensicht, Anlage C	x		
Evaluation der Rückmeldungen und Handlungsempfehlungen, Anlage D		x	(x)
ZUF-8, Fragebogen zur Messung der Patientenzufriedenheit	x		
SERVQUAL, Einstellungen und Zufriedenheit bei Dienstleistungen (Parasuraman, Zeithaml, Berry 1985)	x	x	

## 2.3 Statistische Datenanalyse

Das Modellprojekt fordert und ermöglicht den Einsatz eines breiten Spektrums an einfachen und komplexen Datenanalyseverfahren, adaptiert an die jeweilige Fragestellung, um damit eine möglichst hohe statistische Schlussfolgerungsgültigkeit zu erzielen.

### 2.3.1 Teststärkeanalyse und Fallzahlenberechnung

Ein zentraler Aspekt der statistischen Schlussfolgerungsgültigkeit ist die a-priori Planung einer ausreichenden Fallzahl für die vier verschiedenen Gruppen, um praktisch relevante Effektgrößen mit hoher Teststärke (Power = 1 - Fehler 2. Art) bei gegebenen Alpha-Niveau (Fehler 1. Art) auch entdecken zu können, falls vorhanden. Die Ableitung der erforderlichen Fallzahlen findet auf der Basis der Berechnung der Break-Even-Effektgröße und Variationen und Sensitivitätsanalysen für monetär bewertbare Kriterien statt (siehe Anlage A).

Als Effektgrößenmaße werden Cohen's d und das Glass'sche delta für alle Prä-Post-Vergleiche und alle Mittelwertsunterschiede zwischen den vier Gruppen eingesetzt. Bei Cohen's d werden Mittelwertsunterschiede im Prä-Post-Vergleich durch die gepoolte Standardabweichung prä-post dividiert. Bei Mittelwertsunterschieden wird diese Differenz durch die gepoolte Standardabweichung der beiden zu vergleichenden Gruppen dividiert. Beim Glass'schen delta wird die Differenz im Prä-Post-Vergleich nur durch die Standardabweichung (SD) prä und beim Vergleich zweier unabhängiger Gruppen nur durch die Standardabweichung der Kontrollgruppe dividiert. Bei Cohen's d kann ein differentieller Effekt die berechnete Effektgröße im Vergleich zum Glass'schen delta erhöhen, wenn die SD-post kleiner als die SD-prä ausfällt und im Vergleich zum Glass'schen delta erniedrigen wenn die SD-post größer ausfällt als die SD-prä. Vergleichbare Effekte ergeben sich, wenn die SD's von EX-Grp und KON-Grp unterschiedlich ausfallen. Erhöhungen und Reduzierungen von SD's sind Anzeichen für differentielle Effekte, die zusätzlich zu Mittelwertsunterschieden auftreten können, wenn Subgruppen von Patienten stärker von einer Intervention profitieren als andere.

Beide Effektgrößenvarianten liefern jedoch Ergebnismaße, die unabhängig von der Skalierung des verwendeten Kriteriums sind. Sie erlauben einen direkten Vergleich unterschiedlicher Kriterienmaße. Sie sind metaanalytisch aggregierbar und können synthetisiert werden. Prä-Post Effektgrößenmaße lassen sich auch für jede einzelne Person berechnen, was über die Analyse von Mittelwertsunterschieden hin-

aus korrelative Analysen zur Prognose und Erklärung der individuellen Variabilität der Effekte über multiple Regressionen und Pfadanalysen zulässt.

Legt man die gewünschte Effektgröße, das Alpha-Niveau und die Power fest, so lässt sich die benötigte Fallzahl ausrechnen. Was ist jedoch ein gewünschter Effekt? Cohen stellte ein heuristisch aus der Forschungspraxis abgeleitete Klassifikation vor, so bezeichnete er  $d=.20$ ,  $.50$  bzw.  $.80$  als kleine, mittlere bzw. große Effekte. Diese Klassifikation gilt jedoch nur für gruppenstatistische Mittelwertsunterschiede. Für individuelle Effekte wurde das Konzept der klinisch relevanten oder bedeutsamen Effekte vorgeschlagen, das auf individuellen Veränderungswerten, dem Standardmessfehler (Reliabilität) des Messinstrumentes und der Prä-Posttestkorrelation beruht. Dieses Konzept wird bei der Indikation zum möglichen Therapieende und dem neuen Gutachterverfahren eingesetzt. Für die Fallzahlenberechnung müssen wir jedoch das gruppenstatistische Cohen  $d$  heranziehen.

Für die Stakeholdergruppen TK und die gesundheitspolitische Öffentlichkeit sind kostenrelevante Kriterien von hoher Bedeutung. Welche Effekte sind unter dieser Perspektive relevant und bedeutsam? In Wittmann, Nübling und Schmidt (2002) ist eine Strategie vorgestellt, wie groß ein solcher Effekt am Break-Even Punkt sein muss, ein Effekt an dem sich Kosten und Nutzen gerade aufwiegen. Wir haben dabei Ideen und Konzeptionen weiterentwickelt, die bereits von Schmidt, Hunter und Pearlman (1982) vorgelegt wurden. Ein Effekt der etwas größer als der Break-Even-Effekt ist, lohnt sich aus einer ROI-Perspektive (ROI= return on investment). Zu unserer Überraschung liegt der Break-Even-Effekt im Bereich von  $d=.15$ , d.h. unterhalb von  $d=.20$ . Bereits kleine Effekte beginnen sich unter einer rein monetären Betrachtung zu lohnen (siehe Anhang A zu den Berechnungsgrundlagen und den getroffenen Annahmen).

Wir haben deshalb die Fallzahlenberechnung auf dieser Basis durchgeführt und kommen zu einer Fallzahl von je 1.400 Patienten für Experimental- und Kontrollgruppe, für  $d=.15$ ,  $\alpha=.01$  und  $\text{power}=.95$ . Liberalisiert man diese strengen Kriterien in  $d=.20$ ,  $\alpha=.05$  und  $\text{power}=.95$  benötigt man 540 Personen pro Gruppe. Eine Teilnehmerzahl von ca. 1.100 Versicherten erlaubt also bereits in Cohen's Terminologie kleine Effekte auf monetären Kriterien statistisch abzusichern, wenn sie vorhanden sein sollten.

Wie bei jedem längsschnittlich angelegten Design werden unweigerlich sogenannte Dropouts entstehen. Patienten könnten die Befragungsinstrumente unvollständig ausfüllen, die Therapie abbrechen oder Informationen könnten bei einzelnen Patienten verloren gehen. Für solche fehlenden Daten innerhalb der Befragungsinstrumente gibt es Schätzverfahren (EM-Algorithmen), die dann eingesetzt werden. Kompletter Datenausfall führt zur Reduzierung der Stichprobe. Wir rechnen damit, dass der Datenausfall nicht größer ist als 20%. Es wird geprüft, ob der Datenausfall als ein Zufallsprozess zu betrachten ist oder ob eine systematische Selektion vorliegt. Weiterhin werden ex-post-facto Power-Analysen durchgeführt. Diese ermöglichen dann bei Vorliegen der tatsächlichen verwertbaren Stichprobengröße, dem festgesetzten Alpha-Fehlerniveau und dem empirisch erzielten Effekt die Berechnung der Teststärke. Sollte diese Teststärke nicht ausreichen, kann entschieden werden, wie lange das Modellvorhaben noch weiterlaufen müsste, um die gewünschte Teststärke zu erzielen.

In die Planung der Fallzahlen sind bereits Sicherheitspuffer eingebaut, die aus heutiger Sicht ein solches Szenario als unwahrscheinlich erscheinen lassen. Diese Puffer liegen in der gewählten Effektgröße und der Teststärke ( $\text{Power}=.95$ ). Erhöht man den erwarteten Effekt am Break-Even-Point von  $d=.15$  auf  $d=.20$ , so liegen wir mit den Erwartungen immer noch im Bereich kleiner Effekte, bei denen sich die Qualitätssicherung bereits zu lohnen beginnt. Senkt man die Teststärke von  $.95$  auf  $.90$  so benötigt man ebenfalls deutlich weniger Personen. In der akademischen Forschungspraxis begnügt man sich sogar häufig mit dem Wert  $.80$ ,

Die Teststärke wurde bewusst höher gesetzt, um in diesem sensiblen Bereich Konflikten über die Bewertung der erzielten Effekte vorzubeugen. Eine Stichprobengröße von 1.000 Personen reicht nach unseren Berechnungen (siehe Anhang A) aus, um allen beteiligten Stakeholdergruppen eine faire Bewertungsgrundlage zu geben. Sollte der tatsächlich erzielte Effekt, den wir a-priori noch nicht kennen, in einem Grenzbereich liegen der zu Bewertungskonflikten führt, so kann die Stichprobenfallzahl durch Verlängerung der Laufzeit gegebenenfalls erhöht werden. Ob ein solcher Fall eintreten wird, lässt sich jedoch

---

bereits nach der Hälfte der Laufzeit des Modellvorhabens abschätzen, da dann die bis dahin erzielten Effektgrößen insgesamt und für bestimmte Teilgruppen berechenbar sind.

Effekte im Spiegel nichtmonetär bewertbarer Kriterien sind ebenso wichtig. Aus der Psychotherapieforschung und den Programmevaluationsstudien zur stationären Psychotherapie wissen wir, dass diese Effekte im Bereich von mittleren und großen Effekten liegen, deshalb können solche Effekte auf jeden Fall mit der geplanten Fallzahl gesichert werden. Die vorgeschlagene Fallzahl entspricht daher allen unterschiedlichen Stakeholdergruppen.

Wie lassen sich diese Fallzahlen praktisch realisieren? Unter der begründeten Annahme, dass ein Psychotherapeut ca. vier neue TK-Versicherte pro Jahr behandelt und der Tatsache, dass in der KV-Hessen 2.010 und in der KV-Südbaden 531 Psychotherapeuten zugelassen sind, werden etwa 8.040 bzw. 2.124 Versicherte der TK in den beiden Regionen eine Behandlung aufsuchen. Bei einer Therapeuten-Teilnahmequote von geschätzten 10% könnten damit insgesamt etwa 1.000 Personen pro Jahr in die Studie aufgenommen werden, was einer Teilnehmerzahl von etwa 2.000 Patienten im Laufe von zwei Jahren entsprechen würde. Mit dieser Teilnehmerzahl könnten sogar kleine Effekte mit einer guten Teststärke nachgewiesen werden.

**Tabelle 6. Zugelassene Psychotherapeuten und Schätzung teilnehmender TK-Patienten pro Jahr.**

	KV Hessen	KV Südbaden	KV Westfalen-Lippe	Summe
<i>Zugelassene Psychotherapeuten</i>	2.010	531	1.307	3.848
Teilnehmerquote 10%	804	212	523	1.539
Teilnehmerquote 20%	1.608	425	1.046	3.078
Teilnehmerquote 30%	2.412	637	1.568	4.618

**Anmerkung.** Die Schätzung erfolgt unter der Annahme, dass ein Psychotherapeut pro Jahr vier TK-Versicherte behandelt.

Die Laufzeit des Modellprojektes ist vor allem durch die maximale bewilligbare Stundenzahl und die Therapiedichte (Anzahl der Sitzungen pro Woche) mitbedingt. Zusätzlich zur Laufzeit der Therapie muss noch das Jahr bis zur Katamnese berücksichtigt werden. Die Gruppe der analytischen Langzeittherapien ist naturgemäß der entscheidende Faktor für den Abschluss des Modellprojektes.

Ausgehend von einer Therapiefrequenz von einer Sitzung pro Woche (TP und VT), wären alle Daten einer VT mit dem maximalen Kontingent von 80 Stunden einschließlich der Katamneseerhebung (52 Wochen) nach 132 Wochen bzw. eineinhalb Jahren verfügbar. Bei einer TP-Langzeitbehandlung von 100 Stunden und 52 Wochen Katamnese beträgt der Beobachtungszeitraum 152 Wochen. Bei der AT (hier ausgehend von 3 Sitzungen pro Woche) wären ebenfalls 152 Monate bzw. knapp 3 Jahre vorzuhalten (jeweils ohne Urlaub und sonstige Unterbrechungen kalkuliert).

Falls die Teilnahmebereitschaft im ersten Jahr bei wenigstens 10% oder darüber liegt, könnte das Modellprojekt in einem überschaubaren Rahmen von drei Jahren in den wichtigsten Aspekten abgeschlossen werden.

Die Daten werden beginnend mit der Eingangsdiagnostik kontinuierlich an das Evaluationsteam zur Auswertung weitergeleitet. Für alle 500 Fälle wird eine Zwischenauswertung vorgenommen. Diese Substichproben können wie in einer Metaanalyse einzeln und kontinuierlich über den Projektverlauf aggregiert analysiert und dargestellt werden. Es wird erwartet, dass die Größe der Effekte des Qualitätsmonitorings mit zunehmender Laufzeit und größerer Vertrautheit der Therapeuten mit dem System ansteigt. Sollten die Effekte deutlich größer als die a priori errechneten Break-Even-Effekte sein, so kann die notwendige Fallzahl und damit die Laufzeit deutlich verkürzt werden. Eine maximale Laufzeit von drei Jahren müsste dann zumindest für VT und TP nicht überschritten werden.

### 2.3.2 Datenaufbereitung und Datenreduktion

Die einzelnen Datenquellen, die wir den Datenboxen zugeordnet haben, umfassen eine Fülle an Variablen und Merkmalen, die über theoriegeleitete Aggregationen und exploratorische Faktorenanalysen reduziert werden, um der Kapitalisierung von Zufallseffekten vorzubeugen. Auf der Kriterienseite werden sogenannte multiple Ergebniskriterien für monetäre und nichtmonetäre Ergebnisaspekte gebildet (Schmidt, Bernhard, Wittmann, & Lamprecht, 1987; Wittmann, Nübling, & Schmidt, Evaluationsforschung und Programmevaluation im Gesundheitswesen, 2002). Ein EMEK-monetär wird als Summe der Reduktion von Leistungen in den Bereichen Krankengeld, Krankenhaus, Kuren und Medikamentenkonsum gebildet, das EMEK-nonmonetär aus den Veränderungen im BSI, IIP-D und SF-12. Die Rechtfertigung der Bildung solcher EMEK-Aggregate wird über die interne Korrelation der Indikatoren überprüft. Beide Varianten können aber auch theoriegeleitet begründet werden, da das monetäre EMEK den Ressourcenverbrauch in Geldwerteinheiten abbildet und das nichtmonetäre EMEK einen Index für die Verbesserung der Lebensqualität darstellt.

### 2.3.3 Deskriptive statistische Analysen

Für Patienten und Therapeuten werden die wichtigsten soziodemographischen Variablen hinsichtlich Häufigkeiten, Proportionen oder Prozentsätze für kategoriale Variablen dargestellt. Für stetige Variablen werden die Maße der zentralen Tendenz (Mittelwerte und Median) und die Streuungsmaße Standardabweichung, Schiefe und Rangemaße berichtet.

### 2.3.4 Inferenzstatistische Analysen

Grundlage für die inferenzstatische Analyse ist das allgemeine lineare Modell im Sinne von Cohen, Cohen, West, & Aiken (2003), das varianzanalytische Datenanalysen und Methoden der Regression/Korrelation miteinander verknüpft. Alle Analysen werden über Methoden der Regressionsdiagnostik auf Ausreißer und Verteilungsannahmen hin überprüft. Bei Verletzungen der Verteilungsannahmen werden die entsprechenden verteilungsfreien Analyseverfahren eingesetzt.

### 2.3.5 Exploratorische Analysen

Der Datensatz ist von der Anzahl der erhobenen Variablen äußerst umfangreich, so dass sich neben den a-priori spezifizierten Hypothesen viele explorativ interessante Fragestellungen bearbeiten lassen. Hauptziel der exploratorischen Analyse ist die Varianzaufklärung ausgewählter multipler Kriterienmaße durch die Prädiktor- und die Treatmentboxvariablen. Die Suchstrategien lauten beispielsweise:

- Welche Kombination von PR- und ETR/NTR-Box Variablen führen zu einer maximalen multiplen Korrelation?
- Gibt es Therapieziele die sich leichter erreichen lassen als andere? Hier wird eine qualitativ orientierte Inhaltsanalyse der Ziele verfolgt.
- Resultieren unterschiedliche Effektgrößen für unterschiedliche Diagnosegruppen?

Die empirisch gestützten Entscheidungsregeln des grafischen Rückmeldesystems werden in der formativen Evaluationsphase (siehe 2.2.3.4.1) ebenfalls exploratorisch gewonnen.

## 2.4 Zentrale Hypothesen

Die Hypothesen orientieren sich an univariaten, bivariaten und multivariaten Verteilungskennwerten der Kriterienvariablen und dienen zur Beurteilung des Gesamterfolgs des Modellvorhabens.

---

## 2.4.1 Mittelwertsunterschiede

### 2.4.1.1 Makroebene - Wirksamkeit und Wirtschaftlichkeit

- a) Das Qualitätsmonitoring führt bei allen nicht monetären Kriterienmaßen zu signifikant besseren Ergebnissen als bei der Kontrollgruppe 1, für die diese Kriterien ebenfalls erhoben werden (für die Kontrollgruppen 2 und 3 werden diese Informationen nicht erhoben).
- b) Das Qualitätsmonitoring, das bei der Experimentalgruppe verwendet wird, führt zu höheren Effektgrößen auf dem EMEK-nonmonetär bei allen Vergleichen von EX-Grp. mit KON-Grp. 1, und auf dem EMEK-monetär bei der Katamnese zu allen Kontrollgruppen 1-3.
- c) Wir erwarten für alle vier Gruppen im EMEK-monetär Effektgrößen  $d > .20$  im Vorher-Nachher-Vergleich, d.h. für den Vergleich des Einjahreszeitraumes vor der Therapie und den der Katamnese.

### 2.4.1.2 Mesoebene - Diagnostische und adaptive Entscheidungsstrategien

- a) Die Vorgehensweise nach dem TK-Modell (EX-Grp.) führt zu einem bedarfsgerechteren Einsatz der therapeutischen Ressourcen (Indikator: Varianz der Behandlungsdauer in der EX-Grp. ist größer als in der KON-Grp. 1).
- b) Die Indikationsstellung bei der Aufnahme von Psychotherapie kann in der EX-Grp. durch den Einsatz verschiedener diagnostischer Instrumente gegenüber der KON-Grp. I verbessert werden.
- c) Die Therapeuten in der EX-Grp. nutzen die Handlungsempfehlungen in unterschiedlicher Weise.
- d) Die beteiligten Gutachter der EX-Grp. nutzen die Handlungsempfehlungen in unterschiedlicher Weise.
- e) Die Konsistenz der eingesetzten Entscheidungsregeln und Handlungsoptionen unterscheidet sich aus Patienten-, Therapeuten- und Gutachtersicht.
- f) Die in der EX-Grp. eingesetzten Messinstrumente decken die zentralen Problembereiche aus Patientensicht ab.
- g) Der Umfang des Erhebungsinventars muss für einen praktikablen Einsatz in der Routine angepasst werden.

### 2.4.1.3 Mikroebene - Praktikabilität und Akzeptanz

- a) Die Patienten in der Experimentalgruppe erleben die Rückmeldung als eine Stärkung ihrer Befähigung (Empowerment).
- b) Es wird eine relativ hohe Akzeptanz der ambulanten Psychotherapie bei Experimental- und Kontrollgruppe auf den Messinstrumenten ZUF-8 und SERVQUAL erwartet.
- c) Experimental- und Kontrollgruppe unterscheiden sich im Spiegel der Patientensicht nicht signifikant voneinander.
- d) Die Praktikabilitätseinschätzung des Qualitätsmonitorings aus Therapeutesicht steigt mit der Anzahl der Therapien, die mit diesem System durchgeführt werden.

## 2.4.2 Korrelative Zusammenhänge und Strukturgleichungsverfahren

Zwischen PR-Box, ETR/NTR-Box Variablen sowie zur KR-Box wird, gestützt auf frühere Forschungsergebnisse, eine Reihe korrelativer Zusammenhänge erwartet.

Modellhypothesen („latent growth“) zum Prozess des Verlaufs der Zwischen- und Endmessungen. Die Zwischenmessungen der psychometrischen Messinstrumente bei der Experimentalgruppe werden hinsichtlich Ausgangsniveau (Intercept) und Prozessverlauf (slope) hin analysiert.

---

### 2.4.3 Gesamtbewertung

Insgesamt stellt sich die Frage, unter welchen Bedingungen die im Modellvorhaben eingesetzten Instrumente und Prozesse für den Einsatz in der Routineversorgung empfohlen werden können. Der Erfolg des Modellvorhabens wird aus der Perspektive der unterschiedlichen Stakeholdergruppen unterschiedlich bewertet werden, da sie sich in der Gewichtung der Erfolgsparameter unterscheiden werden. Eine Gruppe wird besonders den Erfolg in Geldwerteinheiten bewerten, eine andere Gruppe wird Aspekte der Lebensqualität, gesundheitliches Wohlbefinden, Symptomreduktion und Zufriedenheit am höchsten gewichten.

Wir haben einen breiten Kriterienkatalog von monetär bewertbaren Parametern und von nichtmonetären Kriterien vorgesehen, die zu getrennten Indexskalen aggregiert werden. Erstere werden einer Kosten-Nutzen-Analyse unterzogen und letztgenannte einer Kosteneffektivitäts-Analyse. Im Anhang A haben wir eine Form der Kosten-Nutzen-Analyse vorgestellt, die es erlaubt, einen Erfolgsparameter  $d$  am Break-Even-Point zu berechnen, an dem sich monetäre Kosten und Nutzen gerade aufwiegen. Dieser Wert liegt nach unseren Modellrechnungen auch nach Variationen im Sinne einer Sensitivitätsanalyse eher um den Wert  $d=.20$ , der in der Fachliteratur als kleiner Effekt betrachtet wird. Sollte der tatsächliche Erfolg im monetären Index um die Hälfte größer sein, d.h.  $d=.30$ , so wäre das Modellprojekt alleine an diesem Parameter bereits ein großer Erfolg. Betriebswirtschaftlich wäre dann der Index  $ROI = 1,5$ . Bereits deutlich kleinere ROI's werden in anderen Bereichen schon als ein Erfolg betrachtet.

Für den nichtmonetären Index können die Effektgrößenmaße  $d$  ebenfalls als standardisierte Mittelwertsdifferenz vor und nach der Behandlung und im Vergleich zu Kontrollgruppen berechnet werden. Zur vergleichenden Bewertung werden hier als Benchmarks einerseits die bisher berichteten Maße aus Meta-Analysen der internationalen Psychotherapieforschung herangezogen, die im Schnitt um  $d=.80$  liegen, andererseits werden zum Vergleich auch die bisher im deutschen Sprachraum berichteten Effekte stationärer Psychotherapie wie sie im Bereich der psychosomatischen Rehabilitation erforscht wurden herangezogen. Auch in diesem Bereich sind die berichteten Effekte beachtlich, sie streuen um 0,60 bis über 1,00 (Wittmann, Nübling & Schmidt, 2002). Da die Meta-Analysen meist auf gut kontrollierten Studien beruhen, erwarten wir in der realen Versorgungspraxis eine gewisse Reduktion des Wirkungsgrades, da dort die Rahmenbedingungen nicht so gut kontrollierbar sind.

Von einem Erfolg sprechen wir, wenn ein  $d>.40$  erreicht wird und dieser Effekt im Verlauf des Projektes vom ersten zu dritten Jahr signifikant ansteigt. Ein solcher Anstieg wäre ein Nachweis der internen Qualitätssicherung, der auch durch einen Anstieg in der Zufriedenheit der Patienten und Therapeuten begleitet werden sollte. Wir erwarten weiterhin eine stärkere Nutzung der Daten des Qualitätsmonitorings für Ausbildung und Supervision, insbesondere die beteiligten Hochschulambulanzen dürften hierbei eine Vorreiterrolle spielen.

---

## 2.5 Qualitätssicherung der Daten

Die Qualitätssicherung der psychometrisch erhobenen Daten wird durch das Hogrefe-Apparatezentrum sichergestellt, die der kassengespeicherten Daten durch die EDV-Abteilung der TK. Eine Substichprobe bestehend aus den ersten 50 Fällen wird zusätzlich vom Evaluationsteam auf Plausibilität, Vollständigkeit und Ausreißer geprüft. Weiterhin soll die Rückmeldung der Daten an die Therapeuten bezüglich Zeitintervall, Vollständigkeit und Verständlichkeit überprüft werden.

## 2.6 Durchführung der Evaluation

### 2.6.1 Zeitplan

Die Evaluation soll zeitnah und parallel zum Projektverlauf und nicht erst nach Abschluss des Modellprojektes erfolgen. Tabelle 7 gibt den Zeitplan des Modellvorhabens wieder.

Tabelle 7. Zeitplan

	2004	2005	2006	2007
<b>Einrichtung Datenerfassung</b>	x			
<b>Bestandsaufnahme Literatur</b>	x x x			
<b>Auswertungsplanung</b>	x x x			
<b>Dokumentation</b>	x x x x x x x x x x x x x x			
<b>Standardauswertung</b>		x x x x x x x x x x x x x x		
<b>Jahresauswertung</b>		x x	x x	
<b>Abschlussauswertung</b>				x x x

### 2.6.2 Aufgaben des Evaluationsteams

- Randomisierung der Therapeuten zu Experimental- und Kontrollgruppe zu Projektbeginn und fortlaufend bis die notwendigen Fallzahlen erreicht sind sowie die Randomisierung der Gutachten.
- Qualitätskontrolle der vom Trust Center gelieferten Daten an einer ersten Substichprobe von 50 Patienten.
- Kurze strukturierte telefonische Rückfragen bei ausgewählten Therapeuten bezüglich Praktikabilität der Internet- bzw. Papier und Bleistift-Version und bezüglich der Rückkopplung der psychometrischen Eingangs- bzw. Verlaufsdiagnostik, vor allem im ersten halben Jahr des Modellprojektes. Gegebenenfalls kann diese Nachfrage auch über E-Mail internetbasiert erfolgen.
- Kontinuierliches Effektmonitoring über die wichtigsten Ergebnisparameter für abgeschlossene Subgruppen von jeweils 500 Fällen zur Frühevaluation des Modellprojektes.
- Anpassung, Adaptation und Optimierung der Berner Entscheidungsregeln in der formativen Phase.
- Kontinuierliche summative Evaluation der Entscheidungsregeln im Abstand eines halben Jahres im ersten Projektjahr, dann im Einjahresabstand.
- Anpassung und Vorbereitung der Fragebogenerhebungen „Einstellungen zur Qualitätssicherung und zum Gutachtersystem“.
- Jahresberichte und Abschlussbericht des gesamten Modellprojektes.

## 2.7 Anlage A: Berechnung der Break-Even-Effektgröße

Unter der Perspektive des Kosten-Nutzens haben wir die Strategie, a priori einen so genannten Break-Even-Point zu berechnen, als besonders hilfreich eingeschätzt und vorgeschlagen (Wittmann, Nübling, & Schmidt, 2002). Grundlage dieser Betrachtung ist eine von Brogden [Brogden, 1949; Cronbach & Gleser, 1965] entwickelte Gleichung, die den ökonomischen Nutzen einer Intervention abzuschätzen gestattet. Diese Gleichung wurde von Schmidt, Hunter und Pearlman (1982) aufgegriffen und in ihrer Bedeutung für die Bewertung von Ausbildungs- und Trainingsprogrammen vorgestellt. Sie kann jedoch in gleicher Weise für psychotherapeutische und medizinische Interventionen angewandt werden.

$$(1) \quad U = N \cdot T \cdot d \cdot a \cdot SD_{prod} - N \cdot K$$

Die einzelnen Parameter diese Gleichung sind wie folgt definiert:

U = Der Nettonutzen einer Intervention in Geldeinheiten, z.B. Euro,

N = Anzahl der therapierten Patienten,

T = Zeitdauer, wie lange der Therapieeffekt anhält, in Jahren,

d = Die Effektgröße, gemessen als standardisierte Mittelwertsdifferenz, entweder berechnet im Vergleich zu einer unbehandelten Kontrollgruppe, oder als standardisierte Differenz nachher-vorher,

a = Der proportionale Anteil der durch die Therapie verbesserten Kriterien an der Produktivität,

$SD_{prod}$  = Die Standardabweichung der Produktivität einer Vergleichsgruppe die diese Intervention nicht benötigte bzw. erhalten hat auf der Berechnungsgrundlage eines Jahres in Euro,

K = Die Gesamtkosten (direkt plus indirekt) der Intervention pro Patient in Euro.

Vor Durchführung einer Evaluationsstudie kennen wir natürlich die Effektgröße d noch nicht. Wir können aber abschätzen, wie groß der Effekt mindestens sein muss, wenn Bruttonutzen und Kosten sich gerade aufwiegen. Dieser Wert ist der so genannte Break-Even-Point einer monetären Investition. An diesem Punkt ist der Nettonutzen U gleich Null. Setzen wir in Gleichung 1  $U = 0$  und lösen nach d auf, so erhalten wir die Effektgröße am Break-Even-Point (Gleichung 2):

$$(2) \quad d_{break-even} = K / (T \cdot a \cdot SD_{prod})$$

Wir benötigen nur noch die Gesamtkosten K, Dauer des Effektes T, den proportionalen Beitrag der Kriterien an der Produktivität und deren Standardabweichung, um die Effektgröße am Break-Even-Point berechnen zu können. Diese Größe kann als Vergleichsmaßstab für den später zu erhebenden tatsächlichen empirischen Effekt d empirisch herangezogen werden. Bei allen Investitionen hofft man einen so genannten Return On Investment (ROI) zu erhalten, der größer als Eins ist. Niemand ist zufrieden, wenn die Investition nach einem Jahr gerade wieder eingebracht wird. Dieser ROI-Koeffizient ist der Quotient aus der empirischen Effektgröße und derjenigen am Break-Even-Point, oder anders ausgedrückt (Gleichung 3):

$$(3) \quad d_{empirisch} = ROI \cdot d_{break-even}$$

Der Hauptgrund, weshalb diese seit über 50 Jahre bekannte Gleichung nicht angewendet werden konnte, lag in der Schwierigkeit, den Parameter  $SD_{\text{prod}}$  vernünftig zu schätzen. Man würde erwarten, dass die Betriebswirtschaftslehre hierfür eine Lösung erarbeitet hätte, was aber nicht der Fall war. Interessanterweise haben die Psychologen Frank Schmidt und Jack Hunter dieses Problem auf eine verblüffend einfache Weise gelöst und brauchbare Schätzungen entwickelt. Produktivität ist ein Merkmal das durch viele Faktoren beeinflusst wird. Es gibt viele unterschiedliche Wege produktiv zu sein. Solche Merkmale sind normalverteilt. Der Bereich vom 15ten bis zum 85ten Perzentil umfasst hierbei zwei Standardabweichungen, einen Bereich der durch den Mittelwert solcher symmetrischer Verteilungen gerade halbiert wird. Sie befragten daher Experten aus vielen Organisationen und Berufsfeldern, was die Arbeitsleistung eines unterdurchschnittlich produktiven Mitarbeiters am 15ten Perzentil wert sei, was am 50ten Perzentil (Durchschnitt) und was am 85ten Perzentil. Letzteres entspräche einer überdurchschnittlichen Produktivität. Am 85ten Perzentil haben nur noch 15 Prozent der Mitarbeiter eine höhere Produktivität und am 15ten Perzentil nur 15 Prozent eine schlechtere Arbeitsleistung. Zur Stützung der Schätzung empfahlen sie, für die Schätzungen den finanziellen Aufwand heranzuziehen, wenn die Arbeitsleistung auf dem freien Markt eingekauft werden müsste. Die Differenz zwischen dem Mittelwert und dem 85ten Perzentil sollte ungefähr genauso groß sein, wie die Differenz zwischen Mittelwert und dem 15ten Perzentil. Die aggregierten Ergebnisse bestätigten diese Annahme. Das Gehalt wiederum sollte ein Spiegel der Produktivität auf einem freien Markt sein, da Organisationen, die zu viel bezahlen, Bankrott gehen, solche die zu wenig bezahlen, ihre Mitarbeiter verlieren werden. Sie setzten deshalb die auf zwei Wegen geschätzte Standardabweichung in Relation zum bezahlten Jahresgehalt und fanden als Hauptergebnis, dass  $SD_{\text{prod}}$  in dem Intervall 40-70% des Jahresgehaltes liegt. Neuere Untersuchungen weisen eher auf die 70% Marke hin.

Wir haben nun die Grundlagen für die Planung des zu erhebenden Stichprobenumfangs für das Projekt geschaffen. Der Versuchsplan sollte auf jeden Fall einen Effekt mit genügend großer Wahrscheinlichkeit entdecken können, wenn dieser tatsächlich vorhanden sein sollte. Diese Wahrscheinlichkeit nennt man die Power eines Tests. Diese Power sollte wenigstens .80 sein. Die Effektgröße am Break-Even-Point ist für das Gesundheitswesen insgesamt, aber auch für die Unternehmen in denen die TK-Versicherten arbeiten, besonders interessant. Effekte die größer als  $d_{\text{Break-Even}}$  sind sollten deshalb mit dieser Power entdeckt werden können.

Zur Berechnung dieses Effektes nach Gleichung benötigen wir die Parameter aus Gleichung (2). Die Kosten der Therapie pro Versicherten  $K$  setzen sich zusammen aus der Anzahl bewilligter Stunden mal deren Kosten plus die Kosten des Qualitätsmonitoring bei einem späteren Routineeinsatz. Wir setzen 50 Euro pro Stunde an plus 100 Euro für das Qualitätsmonitoringsystem. Bei 30 Therapiestunden wären dies insgesamt 1.600 Euro. Für  $T$  nehmen wir an, dass der Effekt ein Jahr lang anhält. Parameter  $a$  setzen wir gleich .50, d.h. die Verbesserung des Befindens und die Beschwerdereduktion korreliert wenigstens zu .50 mit der Produktivität. Das durchschnittliche Gehalt eines TK-Versicherten liegt bei ca. 37.000 Euro pro Jahr, die Standardabweichung der Produktivität liegt damit nach der 40-70% Regel von Schmidt und Hunter zwischen 14.800 und 25.900 Euro. Setzen wir diese Werte in Gleichung (2) ein, so resultieren für die konservative 40%-Schätzung der  $SD_{\text{prod}}$  ein Effekt von .22 und für die optimistischere 70%-Regel ein Effekt von .12 am Break-Even-Point. Solche Effekte werden nach Cohen als kleine Effekte bezeichnet. Die Konsequenz für die Planung des Stichprobenumfangs lautet deshalb: Bereits bei kleinen Effekten beginnt sich die Therapie auch ökonomisch zu lohnen. Effekte die in diesem Bereich liegen sollten also auf jeden Fall mit genügend großer Teststärke entdeckt werden können, falls sie tatsächlich vorhanden sein sollten. Bei zu geringer Teststärke können solche Effekte nicht mit genügend hoher Wahrscheinlichkeit entdeckt werden. In diesem Fall ist man gezwungen, die Nullhypothese beizubehalten. Als praktische Schlussfolgerung wird dann der Fehlschluss der Ineffektivität gezogen, die Intervention nicht mehr weiter durchgeführt und als Resultat entstehen hohe Opportunitätskosten.

Um einen kleinen Effekt von  $d = .20$  bei einem Alpha-Fehler von .05 mit einer Power von .80 entdecken zu können, benötigt man für den Prä-Post-Test Vergleich wenigstens 310 Personen. Will man bei gleichen Rahmenbedingungen eine Power von .90, benötigt man 430 Personen und für eine Power von .95 benötigt man 540 Personen.

Der ökonomische Nutzen, der unter dieser Perspektive der Produktivitätssteigerung entsteht, ist ein Nutzen, der unmittelbar den Betrieben der Versicherten und dem Gesundheitswesen insgesamt zugute kommt und nicht direkt der TK. Indirekt wird es sich aber auch positiv für die TK auswirken, wenn die Krankenkassen die Effektivität der von ihnen finanzierten Maßnahmen nachweisen können. Der direkte Nutzen für die TK ist natürlich vor allem derjenige, der über die Reduktion unnötiger Leistungsausgaben zustande kommt.

Nehmen wir das Kriterium der Leistungsausgaben als Kriterium zur Berechnung einer Effektgröße am Break-Even-Point, der die TK direkt am stärksten interessiert. Der Parameter aus Gleichung 2 ist nun  $a=1$ , die Kosten für eine Therapie von 30 Stunden plus Qualitätsmonitoring seien wiederum 1.600 Euro. Die Standardabweichung der Leistungsausgaben für ganzjährig Versicherte ist nach Angaben des TK-Finanzcontrollings 3.569 Euro, da die Verteilung relativ schief ist könnte diese Standardabweichung durch Ausreißer erhöht sein und wir reduzieren sie auf 3.000 Euro. Wir nehmen an  $T=2$  Jahre, dann ist  $d_{\text{Break-Even}} = 1.600 / (2 * 3.000) = .27$ , was ebenfalls einem eher kleinen Effekt entspricht. Auch die reine Perspektive der TK deutet daraufhin, dass der Versuchsplan auf jeden Fall genügende Teststärke besitzen muss, um bereits kleine Effekte entdecken zu können.

## 2.8 Anlage B: Einstellungen zur Qualitätssicherung

Ausschnitt aus dem Fragebogen „Einstellungen zu Wirkungen und Einflüssen der Qualitätssicherung in der Psychotherapie“ (modifiziert aus Lutz, 1997):

### Einstellungen zu Wirkungen und Einflüssen der Qualitätssicherung in der Psychotherapie

W. Lutz & S. Tholen, Universität Bern

In diesem Fragebogen geht es um die Effektivität von Qualitätssicherungsmaßnahmen. Zentral sind Ihre Ideen und Erfahrungen zu den Wirkungen und Einflüssen der Qualitätssicherung in der Psychotherapie. Neben den standardisierten Fragen, befinden sich am Ende des Bogens mehrere freie Zeilen, auf denen Sie zusätzliche Ideen zu den Wirkungen oder Risiken der Qualitätssicherung beschreiben können.

Versuchen Sie bitte die folgenden Fragen möglichst vollständig zu beantworten.

Vielen Dank für Ihre Mitarbeit!

#### Welche Wirkungen und Einflüsse haben Ihrer Meinung nach Qualitätssicherungsmaßnahmen ?

		ja	eher ja	weder noch	eher nein	nein
1.	Sie ermöglichen eine sinnvolle Zuteilung von Ressourcen/Sach- und Personalmitteln.	<input type="radio"/>				
2.	Sie sind eine Hilfe bei der Identifikation von problematischen Behandlungsergebnissen.	<input type="radio"/>				
3.	Sie führen zu einer Steigerung des Problembewusstseins.	<input type="radio"/>				
4.	Sie fördern die Unehrlichkeit (z.B. beim Ausfüllen bestimmter Fragen).	<input type="radio"/>				
5.	Sie verstärken die Angst vor Kontrolle.	<input type="radio"/>				
6.	Sie führen zu einer Verbesserung der Arbeitszufriedenheit /Motivation.	<input type="radio"/>				
7.	Sie schränken die therapeutische Freiheit ein.	<input type="radio"/>				
8.	Sie sind eine gute Grundlage für die Forschung.	<input type="radio"/>				
9.	Sie bedeuten mehr Arbeit.	<input type="radio"/>				
10,	Sie liefern Argumente für den Ausbau psychotherapeutischer Versorgung.	<input type="radio"/>				
11.	Sie beinhalten als zentrales Ziel die Kostenkontrolle.	<input type="radio"/>				
12.	Sie liefern einen empirischen Nachweis der Wirksamkeit unseres psychotherapeutischen Vorgehens	<input type="radio"/>				

**Welche Wirkungen und Einflüsse haben Ihrer Meinung nach Qualitätssicherungsmaßnahmen ?**

		ja	eher ja	weder noch	eher nein	nein
13.	Sie decken Defizite bzgl. Struktur, Prozess und den Ergebnissen von Patientenbehandlungen auf.	<input type="radio"/>				
14.	Sie sind ein Beitrag zur optimalen psychotherapeutischen Versorgung der Bevölkerung.	<input type="radio"/>				
15.	Sie machen den Behandlungsverlauf und die Behandlungsergebnisse durchsichtiger.	<input type="radio"/>				
16.	Die Effektivität psychotherapeutischen Handelns ist wissenschaftlich gesichert; dennoch halte ich es für wichtig, durch kontinuierliche Beobachtungen in der Praxis, weitere Informationen zu Wirkungen, Patienteneinschätzungen und Nebenwirkungen zu erhalten.	<input type="radio"/>				
17.	Sie verbessern das diagnostische Verhalten.	<input type="radio"/>				
18.	Sie liefern einen Beitrag zur Verbesserung des Indikationsverhaltens.	<input type="radio"/>				
19.	Sie geben Anregungen das Behandlungskonzept ggf. während der Behandlung besser auf die Patientin bzw. den Patienten abzustimmen.	<input type="radio"/>				
20.	Sie führen zu einer höheren Patientenzufriedenheit.	<input type="radio"/>				
21.	Sie führen zu einer besseren Kommunikation zwischen Leistungserbringern und Kostenträgern.	<input type="radio"/>				
22.	Sie fördern das Kostenbewusstsein und damit die Zielorientierung von Psychotherapeutinnen.	<input type="radio"/>				

**Was bewirkt Qualitätssicherung Ihrer Ansicht nach noch?**

(Sie können hier sowohl positive als auch negative Aspekte anführen)

---



---



---

## 2.9 Anlage C: Evaluationsfragen aus Patientensicht

(Lutz & Tholen, 2003b)

1. Für wie wichtig halten Sie es, dass psychotherapeutische Behandlungsergebnisse erhoben werden (z.B. durch Fragebögen wie diesen)?

wichtig       eher wichtig  
 eher unwichtig       unwichtig

2. War der Zeitaufwand, den Sie im Laufe Ihrer Therapie dafür einsetzen mussten, vertretbar?

ja       eher ja  
 eher nein       nein

3. Waren für Sie die Fragen, die Sie gerade beantwortet haben, verständlich und gut zu beantworten?

gut       eher gut  
 weder noch       eher schwer  
 schwer

4. Hatten Sie Mühe die Fragen zu beantworten, da diese Ihnen zu nahe gingen?

ja       eher ja  
 eher nein       nein

5. Haben die Fragen insgesamt alle Bereiche die Ihnen wichtig erscheinen abgedeckt?

ja       eher ja  
 eher nein       nein

Welcher Bereich hat Ihrer Meinung nach gefehlt?

---

---

---

VIELEN DANK FÜR IHRE MITARBEIT!

---

## 2.10 Anlage D: Evaluation der Rückmeldungen u. Handlungsempfehlungen

(Lutz & Tholen, 2003b)

- **Wie gut stimmten die grafischen Rückmeldungen und Handlungsempfehlungen (Reports) mit Ihrem klinischen Eindruck für diese Patientin/diesen Patienten überein?**

sehr gut       gut       mäßig       wenig       gar nicht

- **In welchem Ausmaß haben die grafischen Rückmeldungen und Handlungsempfehlungen (Reports) ihre Behandlungsstrategie beeinflusst?**

sehr ausgeprägt       ausgeprägt       mäßig       wenig       gar nicht

- **Welche Behandlungsstrategien haben sie auf Grund der grafischen Rückmeldungen und Handlungsempfehlungen (Reports) verstärkt eingesetzt?** Mehrfachantworten sind möglich:

- keine, die Rückmeldungen entsprachen nicht meiner klinischen Einschätzung.
  - keine, die Rückmeldungen haben meine Behandlungsstrategie bestätigt.
  - Ich habe rechtzeitig das Therapieende vorbereitet.
  - Ich habe die therapeutische Beziehungsgestaltung überdacht und verändert.
  - Ich habe versucht die Ressourcen der Patientin/des Patienten gezielt anzusprechen.
  - Ich habe gezielt versucht die Therapiemotivation der Patientin/ des Patienten anzusprechen und zu verbessern.
  - Ich habe gezielt versucht die interpersonalen Probleme der Patientin/ des Patienten anzusprechen und zu verbessern.
  - Ich habe die Ergebnisse der Fragebögen mit der Patientin, dem Patienten besprochen.
  - Ich habe die Patientin/den Patienten mit Kolleginnen/Kollegen diskutiert (Supervision/Intervision).
  - Ich habe gezielt in der Fachliteratur nach Informationen zu der Problematik/der Störung oder den Konflikten dieser Patientin/dieses Patienten gesucht.
  - Ich habe die Behandlungsmodalität geändert (z.B. Hinzunahme des Partners, Gruppentherapie).
  - Ich habe neue Themenbereiche angesprochen.
  - Ich habe neue Hausaufgaben erprobt.
  - Ich habe die Sitzungsintervalle erhöht.
  - Ich habe die Sitzungsintervalle verringert.
  - Ich habe an einer spezifischen Weiterbildung teilgenommen.
- andere: \_\_\_\_\_

**VIELEN DANK FÜR IHRE MITARBEIT!**

---

### 3 Realisierung des Projekts

Nach Wittmann (1995) sind die drei wichtigsten Säulen, auf denen die Evaluationsforschung beruht:

- **Techniken der Versuchsplanung**, hier also das gewählte clusterrandomisierte Studiendesign mit drei Messzeitpunkten Prä, Post und 1-Jahres-Katamnese sowie zusätzlichen Zwischenmessungen bei der Interventionsgruppe.
- **Verfahrensweisen des Assessments**, hier also die gewählte Fragebogenmethode und einbezogenen standardisierten bzw. selbstentwickelten Instrumente mit Selbst- und Fremdbeurteilungen.
- **Bewertungs- und Entscheidungshilfen**, hier also die gewählte standardisierte Ergebnisdarstellung mit Effektgrößen sowie die Aggregation singulärer Bewertungsaspekte zu multiplen Ergebniskriterien, die eine Berücksichtigung unterschiedlicher Stakeholder-Interessen erlaubt. Wo immer möglich, erfolgt eine einfach interpretierbare Darstellung der Ergebnisse mit Prozentwerten bzw. lineartransformierten Wertebereichen von 0-100, Bei der Analyse von kostenrelevanten Aspekten wird, wo immer möglich, eine Transformation der Daten in Geldwerteinheiten (Euro) vorgenommen.

Weiter werden in diesem Kapitel die Randomisierung der Studientherapeuten, die projektbegleitende Unterstützung (Therapeutenschulung, Hotline) sowie die Datenerhebung (Fragebogenrücklauf, Entwicklung der Teilnehmerzahlen) beschrieben.

#### 3.1 Evaluationsdesign und Messzeitpunkte

Zur Beantwortung der Studienfragestellungen aus dem Evaluationsplan wurde ein clusterrandomisierter Kontrollgruppenplan gewählt. Die Zuweisung zur IG und KG fand also auf Ebene der Therapeuten statt, d.h. alle Patienten des betreffenden Therapeuten gehören entweder zur IG oder aber zur KG, um Konfundierungseffekte innerhalb der Praxen zwischen IG und KG zu vermeiden. Dem traditionellen Gutachterverfahren in der Kontrollgruppe wird ein modifiziertes Vorgehen in der Interventionsgruppe gegenübergestellt. Hier wird auf das bislang erforderliche positive Votum eines externen Gutachters verzichtet, wenn zu Therapiebeginn die Patientenbefragung und Therapeutendiagnose eindeutig psychopathologische Ergebnisse liefern. Dabei wurden nur in der IG zur Unterstützung der Diagnosefindung die IDCL-Checklisten (Hiller et al., 1995) eingesetzt, während sich die Therapeuten der KG wie herkömmlich alleine an den ICD-10-Kriterien des diagnostischen Manuals bzw. ihrer bisherigen Praxis orientierten. Die Patienten der Interventionsgruppe werden zu Therapiebeginn, in festgelegten Abständen während der Behandlung, bei Therapieabschluss sowie ein Jahr nach Therapieende mit standardisierten Diagnostikinstrumenten befragt (siehe Abbildung 4). Deren Ergebnisse werden EDV-gestützt (HTS) grafisch aufbereitet und den Therapeuten zeitnah zurück gemeldet (HTS-Rückmeldungen). Ein Teil der Rückmeldung beschreibt - empirisch gestützt - zusammenfassend, wie sich die Problematik des Patienten verändert hat. Die Ergebnisse der Interventionsgruppe werden mit denen einer Kontrollgruppe verglichen, die zu Therapiebeginn, Therapieabschluss und zur 1-Jahres-Katamnese mit den gleichen standardisierten Diagnostikinstrumenten untersucht wird, die Ergebnisse aber nicht zurückgemeldet bekommt.



& Keller, 1994), Fragebogen zu Körperbezogenen Ängsten, Kognitionen und Vermeidung AKV, (Ehlers & Margraf, 2001), Hamburger Zwangsinventar HZI (Zaworka, Hand, Jauernig, & Lünenschloß, 1993), Eating Disorder Inventory EDI (Paul & Thiel, 2005) das Screening für Somatoforme Störungen SOMS (Rief et al., 1997).

Tabelle 8. Psychometrische Testverfahren im TK-Modell

## Testverfahren

### Störungsübergreifend (alle Patienten)

- Brief Symptom Inventory (BSI)
- Inventar für Interpersonale Probleme (IIP-D)
- Fragebogen zum Gesundheitszustand (SF-12)
- Helping Alliance Questionnaire (HAQ)

### Störungsspezifisch (nur bei Patienten mit entsprechender Hauptdiagnose)

- Beck Depressionsinventar (BDI)
- Fragebogen zu körperbezogenen Ängsten, Kognitionen und Vermeidung (AKV)
- Hamburger Zwangsinventar (HZI)
- Screening für Somatoforme Störungen (SOMS)
- Eating Disorder Inventory (EDI)

### Fremdbeurteilungen durch die Therapeuten

- Basisdokumentation
- Therapieeingangsdagnostik
- Global Assessment of Functioning (GAF) und Beeinträchtigungsschwere-Score (BSS)
- Helping Alliance Questionnaire (HAQ)
- Rating zur Problembelastung in verschiedenen Lebensbereichen
- Fragebogen zur Medikation mit Psychopharmaka
- Optional: Operationalisierte Psychodynamische Diagnostik (OPD)

Mit den gewählten fünf störungsspezifischen Inventaren können nicht alle möglicherweise relevanten Diagnosen abgedeckt werden. Aus Gründen der Ökonomie hat sich im Rahmen der Projektvorbereitung diese Auswahl der Inventare für die häufigsten Störungen aber als praktikabel erwiesen. Bei allen anderen Diagnosen wird standardmäßig der BDI eingesetzt. Der BSI, der IIP-D und das jeweilige störungsspezifische Inventar bilden die zentralen Selbstbeurteilungsinstrumente, während SF-12 und HAQ zur ergänzenden Information für die Therapeutinnen und Therapeuten herangezogen wurden. Aus den beiden störungsunspezifischen Instrumenten BSI und IIP-D sowie dem jeweiligen störungsspezifischen Instrument wird nach dem Ausfüllen die Rückmeldung generiert (Lutz et al., 2004; Lutz et al., 2005).

Zu den Fremdbeurteilungen gehören eine Basisdokumentation mit demographischen Angaben und Anamnese, Therapieeingangsdagnostik (nur für die IG) mit dem Berner Inventar für Therapieziele BIT-CT (Grosse Holtforth & Grawe, 2002) sowie eine Abschlussdokumentation mit Fragen zum Therapieverlauf, Modus der Therapiebeendigung, Therapiezielerreichung, Status bei Therapieende, ein Fragebogen zur psychopharmakologischen Behandlung sowie Ratings zur aktuellen Problembelastung in verschiedenen Lebensbereichen. Das Funktionsniveau schätzen die Therapeuten anhand der Global Assessment of Functioning Scale GAF (Endicott, Spitzer, Fleiss, & Cohen, 1976) und anhand des Beeinträchtigungsschwere-Score BSS (Schepank, 1995) ein. Auch die Therapeuten bewerten die Qualität der therapeutischen Beziehung aus ihrer Sicht anhand des Helping Alliance Questionnaire HAQ (Bassler, Potratz, & Krauthauser, 1995). Von den psychodynamisch arbeitenden Therapeuten wird darüber hinaus die Operationalisierten Psychodynamische Diagnostik OPD (Arbeitskreis OPD, 1996) verwendet.

### 3.3 Bewertungs- und Entscheidungshilfen

Zur standardisierten Bewertung der Unterschiede zwischen IG und KG bzw. Prä- und Post-Messung werden Effektgrößen mitgeteilt. Hierzu wird einheitlich und durchgängig die Effektgröße  $d$  mitgeteilt. Nach (Cohen, 1992) lässt sich ein  $d$  von 0,20 als kleiner, ein  $d$  von 0,50 als mittlerer und ein  $d$  von 0,80 als großer Effekt klassifizieren.

Zur zusammenfassenden Bewertung des Behandlungserfolges werden einzelne Beurteilungsaspekte (Items oder Skalen) zu multiplen Ergebniskriterien verdichtet. Vor Aggregation werden die Einzelitems einer Lineartransformation auf einen Wertebereich von 0-100 unterzogen, was die Interpretation der Mittelwerte deutlich vereinfacht, ohne dabei den Nachteil eines Informationsverlustes in Kauf nehmen zu müssen.

#### 3.3.1 Standardisierte Ergebnisbewertung mit Effektgrößen

Die Berechnung der Effektgröße  $d$  erfolgt bei allen Vergleichen von kontinuierlichen Variablen zwischen IG und KG anhand der gepoolten Streuungen nach (Glass, McGaw, & Smith, 1978) mit der folgenden Gleichung (zit. nach Lipsey & Wilson, 2001):

$$d = \frac{M_{IG} - M_{KG}}{\sqrt{\frac{(N_{IG} - 1) \cdot SD_{IG}^2 + (N_{KG} - 1) \cdot SD_{KG}^2}{N_{IG} + N_{KG} - 2}}}$$

Wo Prä/Post-Vergleiche vorgenommen werden, erfolgt die Berechnung von  $d$  grundsätzlich anhand der gepoolten Streuung der Prä-Messung aller Patienten von IG und KG:

$$d = \frac{M_{PRÄ} - M_{POST}}{\sqrt{\frac{(N_{PRÄIG} - 1) \cdot SD_{PRÄIG}^2 + (N_{PRÄKG} - 1) \cdot SD_{PRÄKG}^2}{N_{PRÄIG} + N_{PRÄKG} - 2}}}$$

An einigen Stellen ist dabei der Sonderfall zu berücksichtigen, dass nur Prä/Postmessungen für die IG vorgenommen wurden. Hier erfolgt die Standardisierung der Mittelwertsdifferenzen dann grundsätzlich anhand der Streuung aller Patienten der Prä-Messung der IG:

$$d = \frac{M_{PRÄ} - M_{POST}}{SD_{PRÄ}}$$

Auch kategoriale Variablen werden zwischen IG und KG verglichen. Es werden dichotome Merkmale (2 x 2-Tabellen), aber auch mehrstufig kategoriale Variablen (k x 2-Tabellen) getrennt für IG und KG berichtet. Nachdem der Phi-Koeffizient (bei 2 x 2-Tabellen) bzw. Cramer's V (bei k x 2-Tabellen) hier als Zusammenhangsmaße Verwendung finden, lässt sich  $d$  anhand der folgenden Gleichung (Lipsey & Wilson, 2001) berechnen, indem Cramer's V bzw. Phi für die Produkt-Moment-Korrelation  $r$  in die Gleichung eingesetzt wird:

$$d = \frac{2 \cdot r}{\sqrt{1 - r^2}}$$

Umgekehrt gilt:

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

Näherungsweise gilt somit die Beziehung  $d = 2 r$  bzw.  $r = \frac{1}{2} d$ . Im Ergebnisteil wird im Interesse einer einheitlichen Berichtsgestaltung durchgängig nur die Effektgröße  $d$  mitgeteilt.

Als Signifikanztest wird beim Gruppenvergleich kontinuierlicher Variablen der t-Test für unabhängige Stichproben (IG/KG-Vergleich) bzw. für abhängige Stichproben (Prä/Post-Vergleich) verwendet.

Beim Gruppenvergleich kategorialer Variablen zwischen IG und KG wird der Chi<sup>2</sup>-Test verwendet. Effektgrößen und inferenzstatistische Kennwerte werden hier allerdings nur dann berechnet, wenn die für den Chi<sup>2</sup>-Test geltende Voraussetzung  $N > 4$  in beiden Gruppen erfüllt ist.

### 3.3.2 Multiple Ergebniskriterien

Um bei der Vielzahl einzelner singulärer Bewertungsaspekte nicht den Überblick zu verlieren und um der Gefahr von Fehlinterpretationen aufgrund des Phänomens der Alpha-Fehler-Inflation vorzubeugen, werden aus den Einzelaspekten multiple Ergebniskriterien für die Therapeuten- und Patientenangaben gebildet. Diese erlauben eine zusammenfassende Bewertung des Gesamtstatus der Patienten vor und nach der Therapie bzw. der Veränderung des gesundheitlichen Befindens unter Berücksichtigung unterschiedlicher Stakeholder-Interessen (Abbildung 5).

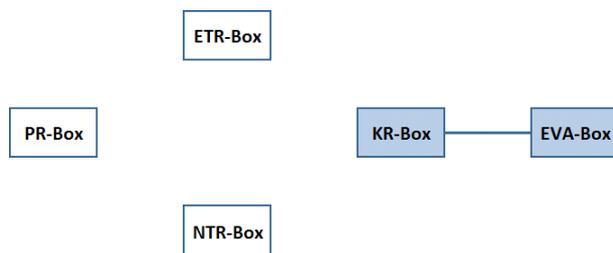


Abbildung 5. Auswahl der Kriterien zur Evaluation der Ergebnisqualität der durchgeführten Therapien (KR-Box) unter Berücksichtigung der Interessen verschiedener Seiten (EVA-Box).

Bei der Bildung der multiplen Ergebniskriterien sind hier folgende Überlegungen ausschlaggebend vgl. Schmidt, Bernhard, Wittmann, & Lamprecht, 1987; Wittmann, Nübling, & Schmidt, 2002; Schmidt, Steffanowski, Nübling, Lichtenberg, & Wittmann, 2003):

- Während bei der traditionellen Skalenkonstruktion eher die Bildung von inhaltlich homogenen Skalen favorisiert wird, ist bei den multiplen Ergebniskriterien die Einbeziehung einer heterogenen Palette von Einzelaspekten ausdrücklich erwünscht, um Einseitigkeiten bei der Evaluation vorzubeugen, welche deren Akzeptanz gefährden könnten.
- Durch die Aggregation erfolgt gegenüber den Einzelitems eine Reliabilitätssteigerung, da sich zufällige Fehler nach beiden Seiten gegenseitig ausgleichen, systematische gleichsinnige Veränderungen mehrerer Bereiche in eine bestimmte Richtung aber deutlicher hervortreten.
- Auch unterschiedlich skalierte Items mit unterschiedlichem Wertebereich lassen sich nach dem Allgemeinen Linearen Modell zu einem multiplen Indikator integrieren (Cohen, Cohen, West, & Aiken, 2003)
- Die gesamte verfügbare Varianz auf Itemebene soll dabei ausgeschöpft werden, d.h. auf eine künstliche Dichotomisierung kontinuierlicher Variablen vor Aufsummierung der Einzelaspekte zu den drei neuen Skalen wird verzichtet, um einem Verlust von Information und Differenzierungsfähigkeit vorzubeugen (Cohen, 1983).
- Alle Einzelaspekte werden hierzu auf eine einheitliche Metrik lineartransformiert. Dies geschieht, indem jedes Item auf einen rechnerisch möglichen Wertebereich von 0 – 100 transformiert wird. Damit bieten diese bzw. die daraus berechneten IG/KG bzw. Prä/Post-Differenzen eine einheitliche, auch für statistische Laien intuitiv verständliche Bewertungshilfe bieten. Diese Bewertungshilfe basiert auf dem Konzept der Qualitätspunkte (QP) der Deutschen Rentenversicherung, das zur standardisierten Bewertung der Struktur, Prozess- und Ergebnisqualität von Rehabilitationseinrichtungen verwendet wird. 100 QP sind dabei als das maximal erreichbare Ergebnis definiert (Beckmann, Klosterhuis, & Lindow, 2007). Alle Items werden vor der Aggregation so gepolt, dass hohe Werte für wünschenswerte Ausprägungen im Sinne eines guten Befindens bzw. Beschwerdefreiheit stehen.

- Die Lineartransformation ist methodisch zulässig, da inferenzstatistische Kennwerte, Korrelationskoeffizienten und Effektgrößen gegenüber der Lineartransformation invariant sind.
- Bei allen monetär bewertbaren Aspekten erfolgt alternativ eine direkte Darstellung bzw. Umrechnung in Euro-Einheiten.
- Das multiple Bewertungs- bzw. Ergebniskriterium wird berechnet, indem bei jedem Patienten die Summe aus den gültig beantworteten Items durch die Anzahl der gültig beantworteten Items geteilt wird. Es wird somit eine individuelle Missing-Data-Korrektur unter der Annahme vorgenommen, dass bei einzelnen fehlenden Items der gemittelte Wert über alle Items des betreffenden Patienten der beste Schätzer für einzelne ausgelassene Items ist. Auch das multiple Ergebniskriterium weist somit einen möglichen Wertebereich von 0-100 Punkte auf.
- Fehlen bei einem Patienten mehr als 20% aller Items, so wird kein Skalenwert für den betreffenden Patienten mehr berechnet, sondern dieser ebenfalls als fehlend codiert. An dieser Stelle ist anzumerken, dass in der zeitgenössischen Forschung häufig eine multiple Schätzung fehlender Item-Werte mittels EM-Algorithmus durchgeführt wird. Eine entsprechende Überprüfung wurde für die Items der Therapeuten- und Patientenskalen vorgenommen. Es ergab sich, dass die EM-korrigierten Skalenwerte nur bei der dritten bzw. vierten Nachkommastelle von den mittels hier beschriebenen Verfahren der individuellen Missing-Data-Korrektur abwichen. Aus Gründen der Nachvollziehbarkeit wurde daher von einer Manipulation der Originaldaten durch eine Schätzung fehlender Item-Werte abgesehen und der hier beschriebenen klassischen Methode der individuellen Missing-Data-Korrektur auf Skalenebene der Vorzug gegeben. Bei einer Verwendung von geschätzten Item-Werten würde man sich ansonsten der Kritik aussetzen, dass hier einfach hypothetische Werte Verwendung finden, die zudem auf Algorithmen basieren, die für den Leser intransparent sind, was den Vorwurf einer gewissen Willkürlichkeit und somit auch Zweifel an der Richtigkeit der berichteten Ergebnisse nach sich ziehen könnte.
- Bei allen monetär bewertbaren Aspekten erfolgt die Skalenbildung alternativ durch einfache Aufsummierung aller im Beobachtungszeitraum angefallenen Verbrauchskosten in Euro.
- Alle Items und Skalen ermöglichen die Berechnung von standardisierten Effektgrößen zum Gruppenvergleich (IG/KG-Vergleiche) bzw. Prä/Post-Vergleiche nach den aktuellen wissenschaftlichen Forschungsstandards im Sinne der metaanalytischen Methodik.
- Im Rahmen der Datenanalysen wurden multiple Bewertungs- bzw. Ergebniskriterien gebildet zur zusammenfassenden Bewertung des Behandlungserfolges. Durch diese Datenaggregation wird dem Argument der Alpha-Fehler-Inflation bei Verwendung mehrerer Einzelvergleiche (z.B. t-Tests) begegnet und zudem eine Reliabilitätssteigerung erzielt. Diese Vorgehensweise entspricht dem Procedere bei metaanalytischen Studien, bei denen beim Vorliegen mehrere Einzeleffekte diese zunächst zu einer Studieneffektgröße aggregiert werden (Lipsey & Wilson, 2001).

### 3.4 Randomisierung der Studientherapeuten

Ziel der Randomisierung war eine zufällige Zuweisung der freiwillig zur Teilnahme an der Studie gemeldeten Therapeuten zur Experimental- (IG) und Kontrollgruppe (KG). Die Randomisierung sollte dabei so erfolgen, dass nicht nur das Verhältnis von Männern zu Frauen in der IG und KG identisch ist, sondern auch die Proportionen in den Therapierichtungen gewahrt werden. Die Randomisierung wurde für jede der drei Modellregionen getrennt vorgenommen. Für jede Modellregion wurde eine Kreuztabelle erstellt, getrennt nach Männern und Frauen sowie nach den drei Therapierichtungen Verhaltenstherapie (VT), Tiefenpsychologisch orientierte Therapie (TP) und Analytische Psychotherapie (AT). Aus jeder der Zellen wurden per Zufall die teilnehmenden Therapeuten zur IG oder KG zugelost. Tabelle 9 zeigt die resultierende Häufigkeitsverteilung für die Gesamtstichprobe der 403 teilnehmenden Therapeuten, getrennt für die Interventionsgruppe (IG) und Kontrollgruppe (KG).

---

Tabelle 9. Häufigkeitsverteilung der 403 teilnehmenden Therapeuten nach Geschlecht und Therapieverfahren.

Geschlecht	Interventionsgruppe (IG)			Kontrollgruppe (KG)			Gesamt		
	M	W	Gesamt	M	W	Gesamt	M	W	Gesamt
<b>VT</b>	54 26,6%	53 26,1%	<b>107</b> <b>52,7%</b>	51 25,5%	55 27,5%	<b>106</b> <b>53,0%</b>	105 26,1%	108 26,8%	<b>213</b> <b>52,9%</b>
<b>TP</b>	26 12,8%	50 24,6%	<b>76</b> <b>37,4%</b>	25 12,5%	51 25,5%	<b>76</b> <b>38,0%</b>	51 12,7%	101 25,1%	<b>152</b> <b>37,7%</b>
<b>AT</b>	1 0,5%	1 0,5%	<b>2</b> <b>1,0%</b>	3 1,5%	0 0,0%	<b>3</b> <b>1,5%</b>	4 1,0%	1 0,2%	<b>5</b> <b>1,2%</b>
<b>VT/TP</b>	3 1,5%	1 0,5%	<b>4</b> <b>2,0%</b>	4 2,0%	2 1,0%	<b>6</b> <b>3,0%</b>	7 1,7%	3 0,7%	<b>10</b> <b>2,4%</b>
<b>TP/AT</b>	7 3,4%	5 2,5%	<b>12</b> <b>5,9%</b>	6 3,0%	2 1,0%	<b>8</b> <b>4,0%</b>	13 3,2%	7 1,7%	<b>20</b> <b>5,0%</b>
<b>VT/TP/AT</b>	1 0,5%	1 0,5%	<b>2</b> <b>1,0%</b>	1 0,5%	0 0,0%	<b>1</b> <b>0,5%</b>	2 0,5%	1 0,2%	<b>3</b> <b>0,7%</b>
<b>Gesamt</b>	<b>92</b> <b>45,3%</b>	<b>111</b> <b>54,7%</b>	<b>203</b> <b>100,0%</b>	<b>90</b> <b>45,0%</b>	<b>110</b> <b>55,0%</b>	<b>200</b> <b>100,0%</b>	<b>182</b> <b>45,2%</b>	<b>221</b> <b>54,8%</b>	<b>403</b> <b>100,0%</b>

Bei einem Ausstieg von Teilnehmern nach erfolgter Randomisierung, allerdings noch vor Studienbeginn (Start der Datenerhebungen), konnte auf eine Warteliste zurückgegriffen werden. Von dieser wurde ein entsprechender Teilnehmer mit der gleichen Merkmalskombination (Geschlecht und Verfahrenszugehörigkeit) - sofern vorhanden - per Zufall ausgewählt. Ein ursprünglich zur KG randomisierter Teilnehmer in Westfalen-Lippe musste dabei zur Vermeidung von Konfundierungseffekten bei Studienbeginn der IG zugeordnet werden, da es sich um den Ehepartner eines IG-Teilnehmers handelte. Nach Beginn der Datenerhebungen wurden bei Drop-Outs keine neuen Therapeuten mehr nachnominiert. Abbildung 6 gibt die Zahl der teilnehmenden Therapeuten in der IG und KG nach dem Consort-Standard (CONSORT, 2010) wieder.



signifikanten Unterschiede zwischen erwarteten und beobachteten Häufigkeiten ( $\chi^2 = 0,435$ ;  $df = 1$ ;  $p = .509$ ), ebensowenig zwischen männlichen (22,5%) und weiblichen (27,6%) Therapeuten ( $\chi^2 = 1,04$ ;  $df = 1$ ;  $p = .293$ ) und der Verfahrenszugehörigkeit (VT: 21,1%; TP: 30,9%; AT und Kombination mehrerer Verfahren: 26,3%;  $\chi^2 = 4,524$ ,  $df = 2$ ,  $p = .104$ ). Bei den Modellregionen liegt die Quote der Aussteiger in Südbaden mit 40,3% allerdings höher als in Westfalen-Lippe mit 22,9% bzw. Hessen mit 22,3% ( $\chi^2 = 8,759$ ;  $df = 2$ ;  $p = .013$ ).

### 3.5 Schulung der Therapeuten

Vor Beginn der Datenerhebungen wurden die teilnehmenden Therapeuten der Interventionsgruppe in Rahmen von eintägigen Schulungen über die Kerngedanken, das Prozedere und Implikationen des Projekts informiert, um effektiv an der Studie mitarbeiten zu können. An diesen Seminaren konnten die Therapeuten freiwillig einen Samstag teilnehmen und erhielten dafür als zusätzlichen Anreiz acht Weiterbildungspunkte. Die Weiterbildungen sind im Zeitraum vom 30.04. bis 03.12.2005 erfolgt, wobei zwei Schulungen in Dortmund, eine in Münster, zwei in Frankfurt und zwei in Freiburg stattfanden. Die Schulungen wurden von der TK in Kooperation mit dem Evaluationsteam der Universität Trier (Professor Lutz; bis 2007 Universität Bern) vorbereitet und durchgeführt. Die Schulungen (Abbildung 7) haben in der entsprechenden Kassenärztlichen Vereinigung oder lokal in einem Kongresshotel stattgefunden.



Abbildung 7. Schulung der Therapeuten der Interventionsgruppe zu Beginn des Modellvorhabens

Folgende Schulungsthemen wurden am Vormittag behandelt:

- Allgemeine Einführung in das Thema Qualitätssicherung
- Vorstellung der Zeitplanung des Projektes
- Erläuterung des Projektordners. Dieser enthält sämtliche zur Studiendurchführung erforderlichen Unterlagen (Informationen, Dokumente und Fragebögen)
- Datenerhebungswege (computergestützt oder traditionell)
- Auswahl der für die eigenen Bedürfnisse optimalen Variante durch die Therapeuten
- Zeitpunkt der jeweils erforderlichen Testungen und deren Vergütung
- Kategoriale Diagnostik sowie Selbst- und Fremdbeurteilungsverfahren der Studie
- Übung des Umgangs mit dem HTS zur computergestützten Datenerhebung.

Folgende Schulungsthemen wurden am Nachmittag behandelt:

- Antrag zur Bewilligung der Therapieverlängerung
- Modifiziertes Gutachterverfahren zur Therapiebewilligung sowie bei Verlängerung
- Handlungsempfehlungen und grafische Rückmeldungen anhand von konstruierten Beispielfällen möglicher Behandlungsverläufe
- Statistische Einführung in die Grundlagen der Handlungsempfehlungen
- Evaluationsfragebögen, die nach der Studie von den Patienten, Therapeuten und Gutachtern beantwortet werden
- Organisation und Durchführung der Katamnese-Erhebung ein Jahr nach Therapieabschluss.

Für die Mitglieder der KG wurde keine Schulung durchgeführt. Sie erhielten lediglich einen Studienordner, dem der Ablauf und die Unterlagen für die Datenerhebungen zum Aufnahmezeitpunkt, bei Entlassung und zur Einjahres-Katamnese zu entnehmen waren.

Insgesamt 159 Therapeuten der IG (78,3% von 203) nutzten das Schulungsangebot, das an den vier Veranstaltungsorten Münster, Dortmund, Frankfurt und Freiburg am Ende des eintägigen Seminars auch evaluiert wurde. Der Mittelwert des Items „Gesamtbewertung der Veranstaltung“ aus dem Fragebogen der Landespsychotherapeutenkammer Baden-Württemberg zur Evaluation akkreditierter Fortbildungsveranstaltungen lag über alle Teilnehmer und Schulungsorte gemittelt bei 2,3 (SD = 0,6). Auch Praxisrelevanz (2,3), Inhalte (2,2), Didaktik (2,3) und Arbeitsatmosphäre (2,3) wurden positiv bewertet (Codierung aller fünf Items nach dem Schulnotensystem: 1 = sehr gut, 2 = gut, 3 = befriedigend, 4 = ausreichend, 5 = mangelhaft, 6 = ungenügend). Für weitere Details sei an dieser Stelle auf den Zwischenbericht 2006 verwiesen.

### **3.6 Hotline für Therapeuten**

Seit der ersten Schulung bis zum Ende des Modellvorhabens wurde durch das Team um Prof. Lutz (Universität Trier/ Bern) eine Telefon- sowie eine E-Mail-Hotline für Fragen und Anmerkungen der Therapeuten angeboten. Die Hotline wurde regelmäßig vonseiten der Mitarbeiter des Teams gepflegt, so dass die telefonischen und schriftlichen Anfragen unmittelbar beantwortet werden konnten.

Im Fall von vertiefenden Fragen oder Anmerkungen konnte so im Rahmen des Projektes ein dauerhafter Informationsdienst angeboten werden. Die Hotline stand sowohl der Interventions- sowie der Kontrollgruppe offen und wurde von beiden in Anspruch genommen. Insbesondere zu Anfang des Modellprojektes wurde die Hotline intensiv von den Therapeuten genutzt. Es handelte sich häufig um Fragen nach dem Ablauf einzusetzender Instrumente sowie näheren Detailfragen zu Skalenaufbau, Normierungen und Cut-off-Kriterien. Im Verlauf des Modellprojektes konnte eine inhaltliche Verschiebung der Fragen beobachtet werden. Während zu Anfang des Projektes vermehrt Fragen zu Abläufen und Kriterien für die Aufnahme von Patienten in das Projekt gestellt wurden, kamen im Verlauf Fragen zu den Zwischenmessungen sowie dem Umgang mit Therapieabbrüchen hinzu. Gegen Ende des Projektes lag der Schwerpunkt auf der Durchführung der Abschluss- und Katamnese-Erhebungen. Daneben wurden über den gesamten Verlauf des Projektes hinweg Fragen zur Dokumentation von patientenspezifischen Informationen sowie administrative Fragen gestellt, zum Beispiel bezüglich des Ablaufs der Nachbestellung von Fragebögen und Formularen oder zum Versand von Unterlagen. Tabelle 10 bietet eine Übersicht der gängigsten Fragen und Anmerkungen der Therapeuten.

Je nach dem, an welchem Punkt die Therapeuten im Modellprojekt standen, waren ausführlichere Antworten notwendig. So konnte bei Therapeuten, die ihren ersten Patienten im Rahmen des Projektes behandelten, die Beantwortung ihrer Fragen schnell mehrere Seiten E-Mail oder lange Telefongespräche bedeuten.

---

Generell lässt sich anmerken, dass die Hotline nicht nur eine klärende Funktion inne hatte, sondern auch motivational förderlich schien. In vielen Fällen konnten über die Hotline auch motivationale und zielbindende Ratschläge vermittelt werden.

**Tabelle 10. Hotlineanfragen der Therapeut/innen**

Anfragetyp	Häufigkeit der Anfrage
Ablauf der Messungen/ des Projekts	42
Dokumentation (z.B. Code-Zuweisung, Therapieabbruch, Psychopharmaka)	25
Diagnostik	18
Patientenakquise und Aufnahmekriterien	15
Abschluss- und Katamneseerhebung	14
Auswertungsschlüssel/ Normierung/ Interpretation	13
Verweis an andere Ansprechpartner	13
Administration (z.B. Nachbestellung Fragebögen, Versendung von Unterlagen)	12
Fragen und Anregungen zum Projekt allgemein	9
Abrechnung von Therapieleistungen	9
Gutachterverfahren bzw. -kurzbericht	7
Beantragung Therapieverlängerung	6
Projektende	5
Softwareprobleme, verzögerte Datenweitergabe	5
Rückmeldungen/ Feedback	5
Datenschutz	2
Persönliche Themen des Therapeuten (z.B. Umzug)	2

### 3.7 Datenerhebung und Fragebogenrücklauf

Wie in Abschnitt 3.5 erwähnt, konnten sich die Therapeuten bei Studienbeginn für einen von zwei Testwegen entscheiden. Teilnehmer mit selbstadministrierter Testung führten die Datenerhebung direkt am PC mit einem für das MVPT speziell konfigurierten Hogrefe Testsystem<sup>®</sup> durch und sandten die Daten an das Trustcenter. Therapeuten mit fremdadministrierter Testung sandten die auf herkömmliche Weise ausgefüllten Fragebögen an das Trustcenter, wo diese dann manuell erfasst und der Studiendatenbank hinzugefügt wurden. Die Studiendatenbank wurde in monatlichem Abstand vom Team des Trustcenters an die Evaluatoren in Mannheim bzw. Trier gesandt. Bei fehlenden Dokumentationen wurden die betreffenden Therapeuten durch das Trustcenter an die Bearbeitung und Nachlieferung der Daten erinnert.

Nachfolgend wird der Fragebogenrücklauf der Therapeuten und Patienten dargestellt. Dabei wird deutlich, dass die Entwicklung der Teilnehmerzahlen insgesamt hinter den Erwartungen zurückgeblieben ist. Am Ende dieses Abschnitts werden mögliche Gründe hierfür diskutiert.

### 3.7.1 Fragebogenrücklauf der Therapeuten

Abbildung 8 gibt die Verteilung der Patienten in der IG und KG sowie in den drei Modellregionen wieder. Es ist ersichtlich, dass in der IG mit 1.031 (60,4% von 1.708 bzw.  $M = 1.031 / 203 = 5,08$  Patienten pro Therapeut) mehr Patienten in das Projekt einbezogen wurden als in der KG mit 677 (39,6% von 1.708 bzw.  $M = 677 / 200 = 3,39$  Patienten pro Therapeut). Eine Ausnahme bildet die Modellregion Südbaden, wo mehr Patienten in der KG als in der IG sind ( $\chi^2 = 18,404$ ;  $df = 2$ ;  $p < .001$ ;  $d = .207$ ). Bis zum Stichtag am 30.06.2010 (Datenannahmeschluss) lagen 1.629 Therapeutenfragebögen zur Eingangsdokumentation bei Therapiebeginn (Stichprobe T1) sowie 934 Therapeutenfragebögen zur Abschlussdokumentation bei Therapieende (Stichprobe T2) vor.

Patienten mit Einwilligung zur Studienteilnahme bei Erstkontakt (N = 1.708)	
Westfalen-Lippe	828
Hessen	788
Südbaden	92
<b>Insgesamt</b>	<b>1708</b>



Zuweisung zur IG und KG anhand der Gruppenzugehörigkeit des Therapeuten, N = 1.708		
	Interventionsgruppe	Kontrollgruppe
Westfalen-Lippe	507	321
Hessen	488	300
Südbaden	36	56
<b>Insgesamt</b>	<b>1031</b>	<b>677</b>



Aufnahmemessung durch den Therapeuten liegt vor, T1 = 1.629		
	Interventionsgruppe	Kontrollgruppe
Westfalen-Lippe	485	313
Hessen	451	294
Südbaden	32	54
<b>Insgesamt</b>	<b>968</b>	<b>661</b>



Aufnahme- und Entlassmessung durch den Therapeuten liegt vor, T2 = 934		
	Interventionsgruppe	Kontrollgruppe
Westfalen-Lippe	292	163
Hessen	259	180
Südbaden	13	27
<b>Insgesamt</b>	<b>564</b>	<b>370</b>

Abbildung 8. Übersicht über die Anzahl teilnehmender Patienten (Therapeutenfragebögen, CONSORT Standard).

Abbildung 9 zeigt die Entwicklung des Rücklaufs der Fragebögen für beide Studiengruppen.

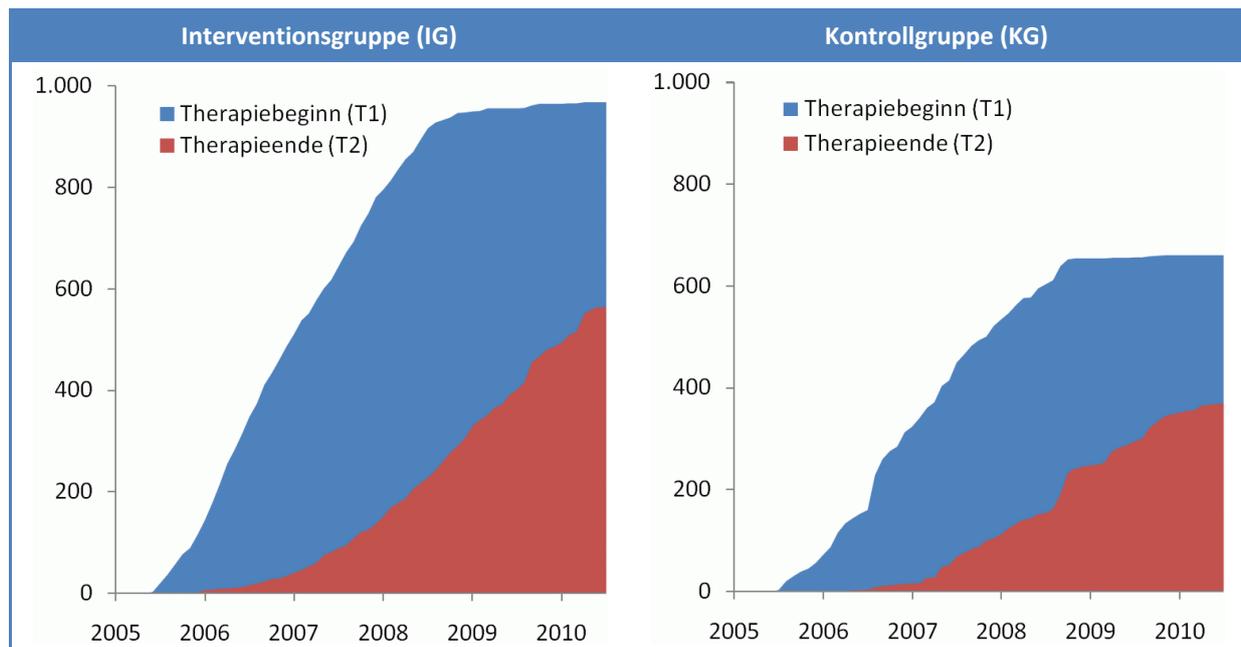


Abbildung 9. Rücklauf der Therapeutenfragebögen.

Die Stichprobe T2 mit 934 Abschlussdokumentationen entspricht einem Anteil von 57,3% an der Stichprobe T1 mit 1.629 Aufnahmedokumentationen. Tabelle 11 gibt die Verteilung der Fragebögen auf die beiden Studiengruppen wieder. Der Chi<sup>2</sup>-Wert beträgt .359 (d = 0,05), d.h. IG und KG unterscheiden sich im Hinblick auf den Anteil fehlender Abschlussdokumentationen nicht voneinander.

Tabelle 11. Vorhandene Abschlussdokumentationen der Therapeuten (T1 = 1.629 Aufnahmedokumentationen)

	IG		KG		Vergleich	
	N	%	N	%	d	p
Ja	564	58,3	370	56,0	0,05	.359
Nein	404	41,7	291	44,0		
Gesamt	968	100,0	661	100,0		

### 3.7.2 Fragebogenrücklauf der Patienten

Patienten mit Einwilligung zur Studienteilnahme bei Erstkontakt (N = 1.708)	
Westfalen-Lippe	828
Hessen	788
Südbaden	92
<b>Insgesamt</b>	<b>1708</b>

↓

↓

Zuweisung zur IG und KG anhand der Gruppenzugehörigkeit des Therapeuten, N = 1.708		
	Interventionsgruppe	Kontrollgruppe
Westfalen-Lippe	507	321
Hessen	488	300
Südbaden	36	56
<b>Insgesamt</b>	<b>1031</b>	<b>677</b>

↓

↓

Aufnahmemessung des Patienten liegt vor, P1 = 1.598		
	Interventionsgruppe	Kontrollgruppe
Westfalen-Lippe	497	278
Hessen	481	263
Südbaden	33	46
<b>Insgesamt</b>	<b>1011</b>	<b>587</b>

↓

↓

Aufnahme- und Entlassmessung des Patienten liegt vor, P2 = 597		
	Interventionsgruppe	Kontrollgruppe
Westfalen-Lippe	224	97
Hessen	198	102
Südbaden	41	15
<b>Insgesamt</b>	<b>463</b>	<b>214</b>

↓

↓

Aufnahme- und Katamnesemessung des Patienten liegt vor, P3 = 468		
	Interventionsgruppe	Kontrollgruppe
Westfalen-Lippe	154	67
Hessen	150	82
Südbaden	2	13
<b>Insgesamt</b>	<b>306</b>	<b>162</b>

↓

↓

Aufnahme-, Entlass- und Katamnesemessung des Patienten liegt vor, P4 = 300		
	Interventionsgruppe	Kontrollgruppe
Westfalen-Lippe	100	50
Hessen	89	50
Südbaden	1	10
<b>Insgesamt</b>	<b>190</b>	<b>110</b>

Abbildung 10. Übersicht über die Anzahl teilnehmender Patienten (Patientenfragebögen, nach CONSORT, 2010).

Abbildung 11 zeigt die Entwicklung des Rücklaufs der Fragebögen für beide Studiengruppen.

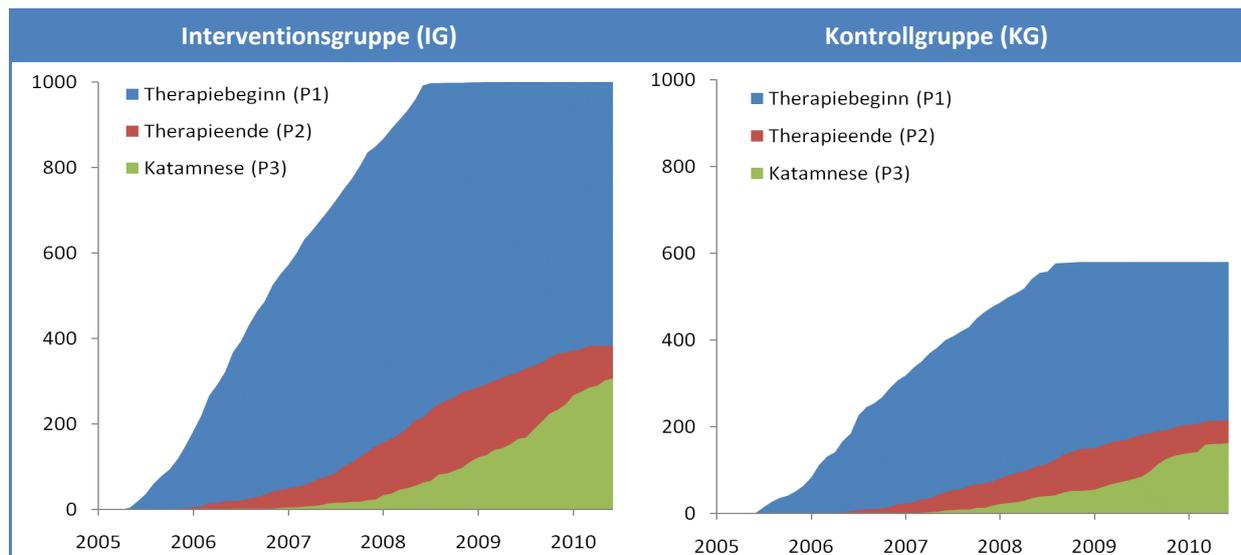


Abbildung 11. Rücklauf der Patientenfragebögen.

Die auf den ersten Blick überraschend geringe Anzahl an Dokumentationen zum Therapieende (P2) bzw. zur 1-Jahreskatamnese (P3), wenn man von der Ausgangsstichprobe P1 ausgeht, bedarf einer näheren Betrachtung und Erläuterung, um voreilige Schlüsse zu vermeiden.

Viele Behandlungen wurden vorzeitig beendet, sei es durch Therapieabbruch oder Wohnortwechsel der Patienten oder aber durch vorzeitigem Ausstieg des Therapeuten aus der Studie, so dass hier keine Daten zum Therapieende bzw. zur 1-Jahres-Katamnese mehr gewonnen werden konnten. Eine detaillierte Analyse zu den Gründen für fehlende Dokumentationen findet sich in Abschnitt 3.8.

Weiter ist anzumerken, dass bis zum Stichtag der letzten Datenannahme am 30.06.2010 viele Therapien noch nicht abgeschlossen waren und eine noch größere Anzahl der bereits abgeschlossenen Therapien nicht das 1-Jahres-Katamnese-Kriterium von 12 Monaten seit Therapieabschluss erfüllten.

Aus den Therapeutenabschlussdokumentationen (Stichprobe T2) sowie zwei von uns in Kooperation mit der TK durchgeführten Nachbefragungen (bei fehlenden Dokumentationen) geht hervor, dass bis zum Stichtag am 30.06.2010 insgesamt 863 Therapien definitiv regulär abgeschlossen wurden.

- Schränkt man die Stichprobe P2 auf diese Zahl von 863 regulär abgeschlossene Therapien ein, für die theoretisch in jedem Fall auch eine Abschlussmessung der Patienten vorliegen sollte, so resultiert eine Rücklaufquote von immerhin noch  $489 / 863 = 56,7\%$  für die Therapieabschlussdokumentationen der Patienten. Die Diskrepanz zu den tatsächlich dokumentierten 597 Patientenabschlussmessungen in P2 erklärt sich hier daraus, dass nicht alle Therapeuten eine Abschlussdokumentation abgegeben haben bzw. auf die Nachfassreaktion reagiert haben, hier aber dennoch in immerhin 108 Fällen noch eine Patientenabschlussdokumentation vorliegt.

Von den 863 regulär abgeschlossenen Therapien wurden 575 Behandlungen bis zum Stichtag am 30.06.2009 (12-Monatszeitraum bis Datenannahmeschluss zum 30.06.2010) auch tatsächlich beendet. Für alle diese Behandlungen sollte theoretisch auch eine Katamnesebewertung der Patienten zur Verfügung stehen.

- Schränkt man die Stichprobe P3 auf diese Zahl von 575 regulär bis zum 30.06.2009 abgeschlossenen Therapien ein, so resultiert eine Schätzung der katamnestischen Rücklaufquote von immerhin noch  $354 / 575 = 61,6\%$ . Die tatsächliche Zahl der dokumentierten Katamnesen liegt mit  $P3 = 468$  um 114 Messungen höher und die Diskrepanz lässt sich auch hier daraus erklären, dass zwar vom Patienten eine verifizierte Katamnesebewertung vorliegt, die Therapeutenabschlussdokumentation jedoch fehlt.

### 3.8 Gründe für fehlende Dokumentationen

Neben den in der Studie erfassten Patienten gibt es auch TK-Versicherte, die im Studienzeitraum von dem betreffenden Therapeuten zwar behandelt wurden, die aber nicht an der Studie teilnahmen. In solchen Fällen konnten die Therapeuten auf einem Kurzfragebogen die Gründe dafür vermerken.

Angesichts der hohen Zahl fehlender Abschluss- und Katamnesedokumentationen wurden in den Jahren 2009 und 2010 ursprünglich nicht vorgesehene Nachfassaktionen per Befragung all derjenigen Therapeuten durchgeführt, bei denen noch Datenlieferungen ausstanden und um Mitteilung des Studienstatus der betreffenden Patienten bzw. Nachreichung der fehlenden Dokumentationen gebeten.

#### 3.8.1 Nichtteilnahme von Patienten am Modellvorhaben

Neben den 1.629 Aufnahmedokumentationen liegen auch 469 ausgefüllte Fragebögen der Therapeuten vor, welche die Gründe für eine Nichtteilnahme von TK-Versicherten am Modellvorhaben dokumentieren. Auch hier wurden mehr Fragebögen von den IG-Therapeuten (N = 262; entspricht 55,9% von 469) abgegeben als von den KG-Therapeuten (N = 207; entspricht 44,1% von 469). Zwischen IG und KG ergeben sich keine statistisch signifikanten Unterschiede hinsichtlich der Geschlechts- und Altersverteilung der nichtteilnehmenden Patienten.

Tabelle 12. Geschlecht der nicht am Modellvorhaben teilnehmenden Patienten

	IG		KG		Vergleich	
	N	%	N	%	d	p
Männlich	78	31,0	58	29,3	0,04	.704
Weiblich	174	69,0	140	70,7		
Gesamt	252	100,0	198	100,0		

Tabelle 13. Geschlecht der nicht am Modellvorhaben teilnehmenden Patienten

	IG		KG		Vergleich	
	N	%	N	%	d	p
bis 30 Jahre	52	20,6	47	24,0	0,18	.457
30-39 Jahre	58	23,0	33	16,8		
40-49 Jahre	77	30,6	69	35,2		
50-59 Jahre	49	19,4	34	17,3		
ab 60 Jahre	16	6,3	13	6,6		
Gesamt	252	100,0	196	100,0		

In den Fällen, wo die Therapeuten einem TK-Versicherten keine Teilnahme an der Studie offeriert haben, wurde als Begründung in 59 Fällen das klinische Zustandsbild angegeben. In 23 Fällen wurden negative Auswirkungen auf die Therapie befürchtet und in 20 Fällen war eine mögliche Überforderung des Patienten durch die Testung der Grund für den Ausschluss. Unter der Rubrik "Sonstige Gründe" wurden von den Therapeuten sehr unterschiedliche Hintergründe für die Nicht-Teilnahme von Patienten gegeben. Wiederholt genannt wurden hier eine hohe Belastung der Patienten und andere Störungsbilder, die entweder Ausschlusskriterien des TK-Modellvorhabens waren oder aber mit dem Fragebogenmaterial nicht gut zu erfassen sind. In Tabelle 14 sind die Ergebnisse wiedergegeben. Signifikante Unterschiede zwischen IG und KG bestehen nicht. Die sonstigen Gründe sind im Originalwortlaut im Anhang aufgeführt.

Tabelle 14. Gründe, warum dem Patienten durch den Therapeuten die Teilnahme nicht offeriert wurde

Patient Teilnahme nicht offeriert...	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben waren möglich</i>						
wegen des klinischen Zustandsbildes	28	10,7	31	15,0	-0,13	.164
weil die Testbeantwortung den Patienten überfordern könnte	8	3,1	12	5,8	-0,13	.144
weil sich eine Teilnahme negativ auf die Behandlung auswirken könnte	11	4,2	12	5,8	-0,07	.426
weil der Aufwand für jeden TK-Patienten zu groß ist	3	1,1	7	3,4	-0,15	.096
weil der Patient eine negative Einstellung zu Tests hat	9	3,4	5	2,4	0,06	.519
aus sonstigen Gründen (offene Antwort)	58	22,1	42	20,3	0,04	.628
<b>Gesamt</b>	262	100,0	207	100,0		

Anmerkung. Angaben der Therapeuten.

Darüber hinaus gab es TK-Versicherte, die von sich aus eine Studienteilnahme ablehnten. 158 Patienten wollten keinen zusätzlichen Zeitaufwand für die Testung auf sich nehmen. 103 Patienten misstrauten dem Datenschutz und wollten daher keine persönlichen Angaben für die Studie machen. Tabelle 15 zeigt die Ergebnisse. Signifikante Unterschiede zwischen IG und KG bestehen nicht. Größere Gruppen in der Kategorie „andere Gründe“ waren Therapieabbrüche während der Probatorischen Sitzungen und generell die Ablehnung einer Mehrbelastung; alle Nennungen sind im Anhang aufgelistet.

Tabelle 15. Gründe, warum der Patient die offerierte Teilnahme abgelehnt hat

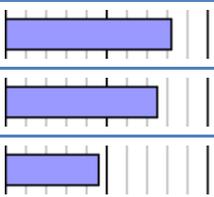
Patient lehnte Teilnahme ab...	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben waren möglich</i>						
mag keine Tests	51	19,5	37	17,9	0,04	.661
misstraut dem Datenschutz	60	22,9	43	20,8	0,05	.580
mag keinen zusätzlichen Zeitaufwand	86	32,8	72	34,8	-0,04	.656
fürchtet Ablehnung der Therapie (-verlängerung) wegen Psychometrie	6	2,3	4	1,9	-	-
hat bereits an Studien teilgenommen und daher momentan kein Interesse	7	2,7	8	3,9	-0,07	.466
fühlt sich nicht fähig, den Fragebogen auszufüllen	24	9,2	26	12,6	-0,11	.236
fehlende Bezahlung	2	0,8	1	0,5	-	-
sonstige Gründe (offene Antwort)	68	26,0	43	20,8	0,12	.190
<b>Gesamt</b>	262	100,0	207	100,0		

Anmerkung. Angaben der Therapeuten.

In den Sommermonaten des Jahres 2006 wurde im Auftrag der Techniker Krankenkasse eine Telefonumfrage durchgeführt, in deren Rahmen das Rekrutierungsverhalten der teilnehmenden Therapeuten der Interventionsgruppe genauer untersucht werden sollte. Aufgrund der eingeschränkten Erreichbarkeit der Therapeuten (viele Bandansagen mit ungenauer Angabe von Sprechstundenzeiten, außerdem fiel die Befragung in die Urlaubszeit), konnten 93 Therapeuten telefonisch befragt werden (vgl. Tabelle 16). Von diesen Therapeuten gaben 76 (82%) an, seit ihrer Einschreibung ins Modellvorhaben TK-Patienten in Behandlung zu haben. Jedoch konnten davon nur 35 Therapeuten alle TK-Patienten ins Modell auf-

nehmen. Insgesamt wurden von allen befragten Therapeuten seit ihrer Einschreibung ins Modellvorhaben 263 TK-Patienten behandelt, wovon 66 (25,1%) nicht ins Modellvorhaben aufgenommen wurden.

**Tabelle 16. Telefonbefragung im Jahr 2006: Behandlung und Aufnahme von TK-Versicherten ins Modellvorhaben**

	IG		
	N	%	
Haben Sie seit Ihrer Einschreibung ins Modellvorhaben TK-Patienten behandelt (bezogen auf alle 93 befragten Therapeuten)?	76	81,7%	
Anteil der ins Modellvorhaben aufgenommenen TK-Patienten (bezogen auf alle 263 behandelten TK-Versicherten)	197	74,9%	
Konnten Sie alle TK-Patienten ins Modellvorhaben aufnehmen (bezogen auf 76 Therapeuten, die TK-Versicherte behandelt haben)?	35	46,0%	

### 3.8.2 Fehlende Abschluss- und Katamnesedokumentationen

Tabelle 17 zeigt die Ergebnisse der Nachbefragung der Therapeuten hinsichtlich fehlender Abschlussdokumentationen. Zwischen IG und KG ergeben sich dabei keine signifikanten Verteilungsunterschiede.

Immerhin haben die Studienteilnehmer in  $\frac{3}{4}$  aller Fälle auf die Anfrage reagiert und die fehlende Abschlussdokumentation begründet, wenn keine Nachlieferung der Daten mehr möglich war. In 25% aller Fälle ist der Grund für die fehlende Abschlussdokumentation nicht bekannt, da die Befragten auf die beiden Erinnerungsschreiben nicht reagiert haben. Etwa 20% aller Patienten war zum Zeitpunkt der Erinnerungsaktion noch in Behandlung und etwa weitere 20% hatten ihren Wohnort gewechselt, weshalb keine Abschlussdokumentation möglich war. In 16% aller Fälle gaben die Therapeuten an, dass die Therapie regulär abgeschlossen wurde, ohne die fehlende Dokumentation weiter zu begründen oder die ausstehenden Daten nachzuliefern. In 5% aller Fälle war der Therapeut aus der Studie ausgestiegen und bei weiteren 5% aller Fälle handelte es sich um einen echten Therapieabbruch. Weitere Gründe bestanden darin, dass nach den probatorischen Sitzungen keine Therapie zustande kam, die Therapie unterbrochen wurde mit Option auf Wiederaufnahme oder aber andere Gründe eine Fortführung nicht möglich machten. So wechselten zwei Patienten im Studienzeitraum die Krankenkasse, zwei den Therapeuten und drei Patienten wurden in die stationäre Psychiatrie eingewiesen. Vier Patienten verstarben im Studienzeitraum. Insgesamt wurden zwei postalische Erinnerungsaktionen bei ausstehenden Abschluss- bzw. Katamnesedokumentationen durchgeführt, eine im Jahr 2009 und eine im Jahr 2010. Darüber hinaus gehende Kontaktierungen der Therapeuten bei fehlenden Datenlieferungen wurden angesichts der im Projektverlauf aufgetretenen, teilweise nicht unerheblichen Widerstände einiger Interessensgruppen gegen das Modellvorhaben, nicht durchgeführt, um eine Verärgerung der Teilnehmer zu vermeiden.

Tabelle 18 gibt die entsprechende Verteilung für die Patienten *mit* vorhandener Abschlussdokumentation (T2) wieder. Etwa vier von fünf Therapien wurden regulär abgeschlossen, statistisch signifikante Verteilungsunterschiede zwischen IG und KG bestehen nicht.

Auch bei überfälligen Katamnesedokumentationen wurden die Therapeuten im Rahmen der beiden Erinnerungsanschreiben darum gebeten, fehlende Katamnesen nachzureichen bzw. den Grund für das Fehlen der Daten zu dokumentieren. Tabelle 19 gibt die Details wieder. Insgesamt liegt die Rücklaufquote in der IG um 6,1% höher als in der KG.

Tabelle 17. Modus der Therapiebeendigung bei fehlender Abschlussdokumentation (T1 – T2 = 695)

	IG		KG		Vergleich	
	N	%	N	%	d	p
Keine Antwort / Unbekannt	93	23,0	77	26,5	0,24	.589
Therapie kam nicht zustande	15	3,7	8	2,7		
Patient ist noch in Therapie	90	22,3	52	17,9		
Therapie ist unterbrochen	10	2,5	7	2,4		
Therapie wurde regulär beendet	71	17,6	45	15,5		
Therapie vorzeitig beendet: Abbruch	21	5,2	14	4,8		
Therapie vorzeitig beendet: Umzug	77	19,1	65	22,3		
Therapie vorzeitig beendet: Sonstiges	8	2,0	4	1,4		
Therapeut aus der Studie ausgestiegen	18	4,5	19	6,5		
<b>Gesamt</b>	<b>404</b>	<b>100,0</b>	<b>291</b>	<b>100,0</b>		

Tabelle 18. Modus der Therapiebeendigung bei vorhandener Abschlussdokumentation (T2 = 934)

	IG		KG		Vergleich	
	N	%	N	%	d	p
Keine Antwort / Unbekannt	0	0,0	0	0,0	0,21	.257
Therapie kam nicht zustande	0	0,0	0	0,0		
Patient ist noch in Therapie	2	0,4	2	0,5		
Therapie ist unterbrochen	3	0,5	1	0,3		
Therapie wurde regulär beendet	433	76,8	311	84,1		
Therapie vorzeitig beendet: Abbruch	79	14,0	32	8,6		
Therapie vorzeitig beendet: Umzug	37	6,6	20	5,4		
Therapie vorzeitig beendet: Sonstiges	10	1,8	4	1,1		
Therapeut aus der Studie ausgestiegen	0	0,0	0	0,0		
<b>Gesamt</b>	<b>564</b>	<b>100,0</b>	<b>370</b>	<b>100,0</b>		

Tabelle 19. Katamnesestatus bei vorhandener Abschlussdokumentation (T2 = 934)

	IG		KG		Vergleich	
	N	%	N	%	d	p
Keine Antwort des Patienten / Unbekannt	251	44,5	164	44,3	0,22	.007
Patient verzogen	10	1,8	11	3,0		
Patient nicht erreichbar	7	1,2	10	2,7		
Patient möchte nicht teilnehmen	6	1,1	10	2,7		
Nichtteilnahme aus anderen Gründen	18	3,2	10	2,7		
Technische Probleme (Datenverlust etc.)	7	1,2	10	2,7		
Katamnese liegt vor	265	47,0	155	41,9		
<b>Gesamt</b>	<b>564</b>	<b>100,0</b>	<b>370</b>	<b>100,0</b>		

### 3.9 Abschließende Bewertung

Insgesamt fällt eine hohe Zahl von fehlenden Therapieabschluss- und Katamnesedokumentationen in der Studie auf. Hierfür lassen sich drei Gründe anführen:

1. **Überschreitung des Studienzeitraumes:** Gemäß § 63 Abs. 5 SGB V sind Modellvorhaben inklusive Planung, Realisierung und Auswertung im Regelfall auf längstens 8 Jahre zu befristen. Eine erneute Verlängerung des Erhebungszeitraumes kam daher nicht in Frage. Eine Reihe von Daten zu den Therapieabschlüssen und Katamnesen von Therapien, die im Studienzeitraum zwar begonnen, aber nicht mehr rechtzeitig beendet wurden, konnten daher bei der Auswertung keine Berücksichtigung mehr finden. Dies betrifft insbesondere die Langzeitbehandlungen, die dadurch in der statistischen Analyse etwas unterrepräsentiert sind. Dieser Kontext sollte bei der Bewertung der Rücklaufquoten berücksichtigt werden.
  2. **Technische Limitationen:** Verantwortlich für das Controlling der Datenerhebung war als Trustcenter das Apparatezentrum des Hogrefe Verlags in Göttingen. Trotz guter Zusammenarbeit mit dem Team des Trustcenters erwies sich das für die Datenerfassung im MVPT speziell adaptierte Hogrefe Testsystem<sup>®</sup> (HTS) als ungeeignet, um ein einwandfreies Controlling zu gewährleisten. Dies beginnt mit einer fehlenden eindeutigen Benutzerführung, wann - unter Ausschluss möglicher Fehleingaben - welche Messungen im Rahmen der Studie vorzunehmen sind und hört mit der Exportschnittstelle auf, die keine standardisierte Datenbankbindung vorsieht, d.h. alle Daten mussten zunächst mit entsprechendem Zeitaufwand von Hand in mehrere ASCII-Textdateien extrahiert werden. Darüber hinaus traten mehrfach Inkonsistenzen in den monatlich gelieferten Daten auf, d.h. einzelne ehemals vorhandene Fälle waren in den Folgelieferungen mitunter nicht mehr enthalten, so dass entsprechend recherchiert und nachgebessert werden musste. Wiederholte Messungen zu mehreren Zeitpunkten wurden in den Exportdateien vom HTS einfach aneinandergereiht und bei versäumten Testsitzungen ging der eindeutige Bezug zu den vom Studienprotokoll vorgesehenen Messzeitpunkten verloren, wenn der Therapeut übersah, die aktuelle Sitzungsnummer einzutragen, was häufig der Fall war. Darüber hinaus war bei den Patientenangaben nicht gekennzeichnet, ob es sich bei der letzten dokumentierten Messung noch um eine Verlaufsmessung oder bereits um die Abschlussmessung handelte, so dass dies erst aus dem Datum auf den Therapeuten- und Patientenfragebögen durch entsprechenden Abgleich ermittelt werden konnte. Auch die Katamnesemessung war nicht immer als solche gekennzeichnet (obwohl die entsprechende Option bestand und zum Teil auch genutzt wurde) und musste in solchen Fällen ebenfalls durch Abgleich der Datumsangaben mit den Therapieabschlussdokumentationen ermittelt werden. Fehlte die Therapieabschlussdokumentation durch den Therapeuten, so war dieser Abgleich natürlich nicht möglich. In solchen Fällen musste die letzte nicht als Katamnese gekennzeichnete Messung von uns grundsätzlich konservativ als Zwischenmessung behandelt werden. Die Entwicklungsabteilung des Hogrefe-Verlags sah sich nicht in der Lage, im Studienverlauf weitere für das Controlling benötigte Funktionalitäten der HTS-Software zu ergänzen. Ein Teil der fehlenden Abschluss- und Katamnesedokumentationen ist hier also durch technischen Limitationen der zur Datenerhebung eingesetzten Software bedingt, auf welche die Evaluatoren keinen Einfluss hatten.
  3. **Berufspolitische Vorbehalte:** Neben den erwähnten formalen und technischen Problemen gab es aus unserer Sicht aber auch berufspolitische Gründe, die zu einer verringerten Compliance bis hin zum Studienausstieg beigetragen haben dürften. So wurden bereits im Vorfeld der Studie mehrfach Bedenken hinsichtlich der Evaluation dahingehend geäußert, dass die Studienergebnisse auch zu anderen Aussagen verwandt werden könnten, als die Auswertung der mit der Studie zu überprüfenden Fragestellungen hergeben. (z.B. vergleichende Aussagen zu den drei Verfahren in der Richtlinien-Psychotherapie). Des Weiteren wurde als problematisch bezeichnet, dass Ergebnismessung und das Monitoring weitgehend symptombezogen erfolge, strukturelle und prozessbezogene Änderungskriterien (z.B. anhand der OPD) dabei nicht einbezogen wurden. Dieser Einwand konnte in der Planungsphase noch aufgegriffen werden und die OPD fand in der Interventionsgruppe des Modellvorhabens als optional einsetzbares Assessment-
-

Instrument noch Berücksichtigung. Die entsprechenden Auswertungsergebnisse finden sich in Kapitel 5.2. Auch die Gutachter äußerten massive Vorbehalte gegen das Modellvorhaben. Diese Einwände wurden – trotz Information und Vorbereitung der Studie unter Einbeziehung aller Beteiligten – auch während der Datenerhebungsphase von bestimmten Interessengruppen weiter vorgetragen und verbreitet, was möglicherweise zu einer Verunsicherung einiger Therapeuten und damit suboptimalen Entwicklung der Teilnehmerzahlen beigetragen haben mag. So stieg jeder vierte Therapeut in Projektzeitraum aus der Studie aus.

Als Fazit lässt sich aus unserer Sicht festhalten, dass für künftige Vorhaben dieser Art das Studiencontrolling und die Evaluation in einer Hand liegen sollten, um forschungsmethodische Kompetenzen auch bei der Datenerhebung einbringen zu können. Bereits aus Kostengründen und Gründen der Unabhängigkeit von privatwirtschaftlichen Interessen empfiehlt sich dabei der Einsatz von lizenzfreien Testverfahren und allgemein verfügbaren Datenbankanwendungen.

Andere Aspekte liegen außerhalb unseres Einflussbereiches, wie die berufspolitische Diskussion, die von den jeweiligen Interessengruppen zu führen ist. Auch (und gerade) bei Einbeziehung aller Beteiligten in eine Studie dieser Größenordnung gibt es in einem so komplexen Feld wie der psychotherapeutischen Versorgung kontroverse Standpunkte. Als Evaluatoren nehmen wir hier grundsätzlich eine unvoreingenommene und neutrale Position ein. Natürlich befürworten wir als Wissenschaftler eine Erhebung von empirischen Daten, wo immer dies einen Informationsgewinn und Steigerung der Transparenz für alle Beteiligten verspricht unter Wahrung der Privatsphäre und des Datenschutzes.

Trotz der erwähnten Probleme konnte ein beeindruckend großer Datensatz gewonnen werden. Die Stichprobengrößen zum Zeitpunkt Therapieende sind groß genug, um eine Auswertung der Studienfragestellungen aus dem Evaluationsplan mit ausreichender Teststärke zu gewährleisten. Auch die katamnestic Daten lassen sich dazu verwenden, um eine belastbare Aussage über die Stabilität der zum Therapieende gemessenen Ergebnisse treffen zu können.

## 4 Stichprobenbeschreibung und Überprüfung der Randomisierung

Es wird anhand der bei Therapiebeginn erhobenen Daten (PR-Box) überprüft, ob die Randomisierung (ETR-Box) erfolgreich war (Abbildung 12). Hierzu werden Gruppenvergleiche zwischen Interventions- (IG) und Kontrollgruppe (KG) durchgeführt. Bei metrischen Variablen geschieht dies durch Mittelwertvergleiche und t-Tests für unabhängige Stichproben. Bei kategorialen Variablen werden die Häufigkeitsverteilungen mitgeteilt und Chi<sup>2</sup>-Tests zur Identifikation möglicherweise vorhandener Unterschiede zwischen IG und KG berechnet. Die getrennt nach den beiden Gruppen mitgeteilten deskriptiven Statistiken dienen gleichzeitig der Beschreibung der Ausgangsstichprobe aller am Modellvorhaben teilnehmenden Patienten.

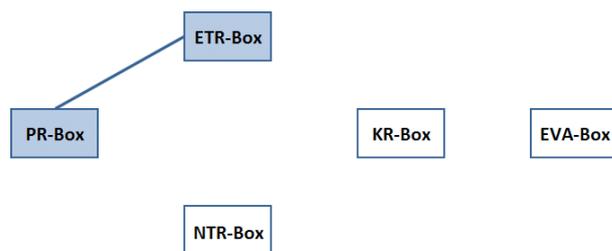


Abbildung 12. Untersuchung der Beziehungen zwischen PR-Box und ETR-Box zur Überprüfung, ob sich hinsichtlich der Prädiktorvariablen Unterschiede zwischen Interventions- und Kontrollgruppe ergeben.

### 4.1 Repräsentativität der Daten: Stichprobe TK

Wie im Evaluationsplan vorgegeben, steht aus der Datenbank der TK ein umfangreicher Referenzdatensatz für mehrere Kontrollgruppen zur Verfügung, um die Repräsentativität der im Modellvorhaben dokumentierten Behandlungen für alle TK-Versicherte im gleichen Zeitraum abschätzen zu können.

Einschlusskriterien gemäß Studienprotokoll:

- Ambulante Psychotherapie mit Behandlungsbeginn zwischen 01.04.2005 und 30.06.2008
- Diagnose aus der Gruppe F3-F6 liegt vor (Ausschlusskriterium: Diagnose F1-F2)
- Patient bei Therapiebeginn mindestens 18 Jahre alt

Folgende Stichproben konnten mit diesen Angaben aus dem Datenbanksystem der TK zum Stichtag am 30.06.2010 gewonnen werden:

- IG = 1.060 Behandlungen: Patient wurde in die Studie einbezogen. Therapeut nimmt am Modellvorhaben teil und gehört zur Interventionsgruppe.
- KG = 614 Behandlungen: Patient wurde in die Studie einbezogen. Therapeut nimmt am Modellvorhaben teil und gehört zur Kontrollgruppe.
- NT = 2.778 Behandlungen: Patient wurde nicht in die Studie einbezogen. Therapeut nimmt jedoch am Modellvorhaben teil, d.h. dem Patienten wurde eine Teilnahme nicht offeriert oder der Patient lehnte die Studienteilnahme ab.
- KG2 = 33.210 Behandlungen: Patient und Therapeut nehmen nicht Modellvorhaben teil, der Patient ist jedoch im Bereich der am Modellvorhaben beteiligten KVen versichert.
- KG3 = 27.563 Behandlungen: Patient und Therapeut nehmen nicht Modellvorhaben teil und der Patient ist in einem anderen KV-Bereich versichert (Bayern, Brandenburg und Hamburg)

Tabelle 20 gibt das durchschnittliche Alter in den Stichproben wieder. Eine Oneway-Varianzanalyse ist statistisch signifikant ( $p < .001$ ). Die Gruppe NT ist dabei im Durchschnitt am ältesten und unterscheidet sich im Einzelvergleich von allen anderen Gruppen. Statistisch signifikant wird auch der Altersunterschied um etwas mehr als ein Jahr zwischen IG und KG. Beide Studiengruppen unterscheiden sich jedoch nicht von KG2 und KG3 hinsichtlich des Alters.

Tabelle 20. Alter in den fünf Vergleichsstichproben.

	IG	KG	NT	KG2	KG3
<b>N</b>	1060	614	2778	33210	27563
<b>M</b>	40,2	41,5	42,4	41,2	40,6
<b>SD</b>	11,7	11,0	11,6	11,6	11,2

Anmerkung: Signifikante Einzelvergleiche ( $p < .05$ ): IG-KG; IG-NT; KG2-NT; KG3-NT.

Beim Geschlecht ergeben sich kaum Auffälligkeiten (Tabelle 21). Abgesehen von einem etwas größeren Anteil männlicher Patienten im Vergleich zwischen IG und KG3 sind die Einzelvergleiche nicht statistisch signifikant.

Tabelle 21. Geschlechtsverteilung in den fünf Vergleichsstichproben

	IG	KG	NT	KG2	KG3
<b>Männer</b>	344 32,5%	182 29,6%	841 30,3%	9618 29,0%	7711 28,0%
<b>Frauen</b>	716 67,5%	432 70,4%	1937 69,7%	23592 71,0%	19852 72,0%
<b>Gesamt</b>	1060 100,0%	614 100,0%	2778 100,0%	33210 100,0%	27563 100,0%

Anmerkung: Signifikante Einzelvergleiche ( $p < .05$ ): IG-KG3.

Tabelle 22 schlüsselt die dokumentierten Behandlungen nach den drei Verfahrensrichtungen auf. Verhaltenstherapeutische Behandlungen sind im MVPT sowohl in der IG als auch in der KG überrepräsentiert, tiefenpsychologisch orientierte und analytische Psychotherapien hingegen unterrepräsentiert. Zwischen IG und KG zeigen sich dabei keine Verteilungsunterschiede. Auch in der Nichtteilnehmer-Stichprobe NT ist der VT-Anteil gegenüber KG2 und KG3 noch erhöht, insgesamt aber geringer als in der IG und KG. Die größte Aufgeschlossenheit für das Modellvorhaben war unter den Patienten und Therapeuten also bei verhaltenstherapeutischer Behandlung zu finden.

Tabelle 22. Therapieverfahren in den fünf Vergleichsstichproben

	IG	KG	NT	KG2	KG3
<b>Verhaltens- therapie</b>	716 67,5%	413 67,3%	1599 57,6%	16350 49,2%	11166 40,5%
<b>Tiefenpsychol. Therapie</b>	329 31,0%	197 32,1%	1145 41,2%	15763 47,5%	14586 52,9%
<b>Analytische Therapie</b>	15 1,4%	4 0,7%	34 1,2%	1097 3,3%	1811 6,6%
<b>Gesamt</b>	1060 100,0%	614 100,0%	2778 100,0%	33210 100,0%	27563 100,0%

Anmerkung: Signifikante Einzelvergleiche ( $p < .05$ ): IG-KG3.

Auf den ersten Blick überrascht der Befund, dass in der IG im Vergleich zur KG im Durchschnitt 15 Sitzungen mehr bewilligt wurden (Tabelle 23). Dies ist jedoch der modifizierten Vorgehensweise im Modellvorhaben geschuldet und bedeutet nicht, dass diese Kontingente auch ausgeschöpft wurden (hierzu wird ausdrücklich auf die Ergebnisse in Abschnitt 6.3.3, Abbildung 40) verwiesen. Rechnet man den bewilligten Behandlungszeitraum in Monateinheiten um, so ergeben sich die in Tabelle 24 dargestellten Kennwerte. Über alle Gruppen hinweg zeigt sich dabei eine konstante Größe von 25 Monaten, lediglich in der KG liegt die Behandlungsdauer im Durchschnitt um 1,5 Monate darüber und unterscheidet sich von allen anderen Gruppen statistisch signifikant.

Tabelle 23. Bewilligte Sitzungen in den fünf Vergleichsstichproben.

	IG	KG	NT	KG2	KG3
<b>N</b>	1060	614	2778	33210	27563
<b>M</b>	50,4	36,8	35,3	39,7	45,4
<b>SD</b>	18,2	22,3	24,1	33,3	41,4

Anmerkung: Signifikante Einzelvergleiche ( $p < .05$ ): IG und alle anderen Gruppen; KG3 und alle anderen Gruppen. Die höhere Anzahl bewilligter Sitzungen in der IG bedeutet nicht, dass diese auch ausgeschöpft wurden, sondern sind durch die modifizierte Vorgehensweise im Modellvorhaben bedingt. Die Ergebnisse zu den tatsächlich abgerechneten Sitzungen in IG und KG finden sich in Abschnitt 6.3.3, Abbildung 40.

Tabelle 24. Behandlungsdauer in Monaten

	IG	KG	NT	KG2	KG3
<b>N</b>	1059	613	2778	33210	27563
<b>M</b>	25,2	26,5	24,9	25,0	25,0
<b>SD</b>	7,9	9,1	8,7	8,8	9,1

Anmerkung: Signifikante Einzelvergleiche ( $p < .05$ ): KG und alle anderen Gruppen.

## 4.2 Therapeutenangaben bei Behandlungsbeginn: Stichprobe T1

Insgesamt liegen für T1 = 1.629 Patienten die Angaben bei Behandlungsbeginn durch die Therapeuten vor (95,4% von 1.708). Nachfolgend werden die Verteilungsmerkmale getrennt für die IG und KG berichtet und Gruppenvergleiche (t-Tests für unabhängige Stichproben bei Mittelwerten bzw. Chi<sup>2</sup>-Tests bei kategorialen Variablen) zur Überprüfung der Randomisierung durchgeführt.

Die Angaben der Therapeuten bei Behandlungsbeginn bestehen aus vier Komponenten:

- Basisdokumentation (T1<sub>A</sub> = 1.618 ausgefüllte Fragebögen; 99,3% von 1.629)
- Therapieeingangsdagnostik (T1<sub>B</sub> = 1.572 ausgefüllte Fragebögen; 96,5% von 1.629)
- Psychopharmakologische Behandlung (T1<sub>C</sub> = 1.532 ausgefüllte Fragebögen; 94,0% von 1.629)
- Therapeuten-Ratings (T1<sub>D</sub> = 1.574 ausgefüllte Fragebögen; 96,6% von 1.629)

### 4.2.1 Basisdokumentation

Insgesamt sind T1<sub>A</sub> = 1.618 Fragebögen vorhanden (IG: 964, KG: 654).

#### A) Verwaltung /Organisation

Während sich IG und KG hinsichtlich der Wartezeit auf das Erstgespräch nicht voneinander unterscheiden, fand in der IG im Durchschnitt eine halbe probatorische Sitzung mehr als in der KG statt (Tabelle 25). Streng genommen handelt es sich bei der Anzahl der probatorischen Sitzungen allerdings um keine Prädiktorvariable, sondern bereits um eine Prozessvariable, die der NTR-Box zuzuordnen ist. So ist die deutliche Differenz um eine halbe Sitzung (3,25 versus 2,74) vermutlich durch die umfangreichere Diagnostik bei der IG (TK-Modell) im Vergleich zur KG (herkömmliches Vorgehen) bedingt. Dieser Zusatzaufwand im Vorfeld der eigentlichen Therapie sollte zum einen bei der ökonomischen Bewertung des Modellvorhabens Berücksichtigung finden, zum anderen stellt eine präzise Diagnostik und die damit verbundene Klärung des Behandlungsanliegens aber auch bereits einen wichtigen Bestandteil des therapeutischen Prozesses selbst dar. Hinsichtlich Zuweisungskontext (Tabelle 26) finden sich keine signifikanten Unterschiede bei den drei wichtigsten Zugangswegen der Patienten zum Therapeuten (Hausarzt, Eigeninitiative und Facharzt für Psychiatrie).

Tabelle 25. Wartezeit auf Erstgespräch und Anzahl der probatorischen Sitzungen

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Wartezeit auf Erstgespräch in Wochen	865	6,28	9,37	584	6,24	7,83	0,00	.937
Anzahl der probatorischen Sitzungen	751	3,25	1,30	498	2,74	1,26	0,40	<.001

Anmerkung. T1 = 1.629.

Tabelle 26. Zuweisungskontext

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Ohne Vermittlung / direkt	249	25,8	171	26,1	-0,01	.932
Hausarzt / Allgemeinarzt	302	31,3	184	28,1	0,07	.187
Facharzt für PT Medizin / Ärztl. Psychoth.	43	4,5	15	2,3	0,11	.030
Facharzt für Psychiatrie / Neurologie	159	16,5	103	15,7	0,02	.741
Anderer Facharzt	29	3,0	28	4,3	-0,07	.220
Psychologischer Psychotherapeut	26	2,7	17	2,6	0,01	.999
Kinder- und Jugendlichenpsychotherap.	3	0,3	0	0,0	-	-
Anderer Psychologe	11	1,1	1	0,2	-	-
Psychiatrische Klinik	18	1,9	12	1,8	0,00	.999
Psychosomatische Klinik / Kurklinik	15	1,6	7	1,1	0,04	.542
Anderer Klinik	8	0,8	6	0,9	-0,01	.999
Beratungsstelle / Psychosoz. Einrichtung	23	2,4	17	2,6	-0,01	.914
Kostenträger / Krankenkasse	32	3,3	21	3,2	0,01	.999
Anderer Patient	60	6,2	47	7,2	-0,04	.508

Anmerkung. T1 = 1.629.

## B) Soziodemographische Daten

Bei den nachfolgend aufgeführten soziodemographischen Daten zeigen sich keine gravierenden Unterschiede zwischen IG und KG. Bei einzelnen statistisch signifikanten Gruppenvergleichen ist aufgrund der großen Zahl von Einzelvergleichen auch immer das Problem der Alpha-Fehler-Inflation zu berücksichtigen. Dennoch lässt sich hier ein leichter Trend feststellen und zu einem inhaltlich konsistenten Gesamtbild zusammenfügen:

Die Patienten der IG im Vergleich zur KG...

- sind etwa ein Jahr jünger (M = 40,3 versus 41,2 Jahre; d = -0,10)
- sind eher ledig (37,3% versus 31,0%; d = 0,15)
- leben seltener mit Kindern (36,0% versus 42,7%; d = -0,13)
- sind noch eher in Ausbildung (13,7% versus 9,9%; d = 0,20)
- haben ein geringeres Haushaltsnettoeinkommen (2.135 versus 2.315 Euro; d = -0,14)
- leben seltener vom Einkommen des Ehepartners (23,8% versus 29,7%; d = -0,23).

Die aufgeführten Merkmale zur privaten, beruflichen und finanziellen Situation sind erfahrungsgemäß altersabhängig. Dies wurde mit einfaktoriellen Anovas (bzw. Produkt-Moment-Korrelation beim Einkommen) überprüft. In allen Fällen ist  $p < .001$ , d.h. die aus dem Alltag geläufigen Zusammenhänge mit dem Alter bestehen auch in der Studienstichprobe, weshalb das Alter als möglicherweise relevante Kovariate für die weiteren Analysen vorgemerkt werden sollte.

Tabelle 27. Alter der Patienten

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Alter in Jahren	968	40,16	11,38	659	41,27	11,03	-0,10	.052

Anmerkung. T1 = 1.629.

Tabelle 28. Geschlecht der Patienten

	IG		KG		Vergleich	
	N	%	N	%	d	p
Männlich	311	32,1	201	30,4	0,04	.463
Weiblich	657	67,9	460	69,6		
Gesamt	968	100,0	661	100,0		

Anmerkung. T1 = 1.629.

Tabelle 29. Staatsangehörigkeit

	IG		KG		Vergleich	
	N	%	N	%	d	p
Deutsch	943	97,8	628	96,0	0,11	.050
Andere Nationalität	21	2,2	26	4,0		
Gesamt	954	100,0	661	100,0		

Anmerkung. T1 = 1.629.

Tabelle 30. Familienstand

	IG		KG		Vergleich	
	N	%	N	%	d	p
Verheiratet, zusammen lebend	417	43,6	311	47,8	0,15	.068
Verheiratet, getrennt lebend	70	7,3	46	7,1		
ledig	357	37,3	202	31,0		
geschieden	98	10,3	76	11,7		
verwitwet	14	1,5	16	2,5		
Gesamt	956	100,0	651	100,0		

Anmerkung. T1 = 1.629.

Tabelle 31. Partnersituation

	IG		KG		Vergleich	
	N	%	N	%	d	p
In fester Partnerschaft	620	66,6	429	67,5	0,10	.373
Kurzfristig keine Partnerschaft	105	11,3	85	13,4		
Längerfristig keine Partnerschaft	165	17,7	103	16,2		
Wechselnde Partnerschaften	18	1,9	10	1,6		
Unbekannt	23	2,5	9	1,4		
Gesamt	931	100,0	636	100,0		

Anmerkung. T1 = 1.629.

Tabelle 32. Personen im Haushalt

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Personen, die ständig im Haushalt leben	922	2,47	1,25	636	2,55	1,27	-0,06	.183
Davon älter als 18 Jahre	767	1,96	0,91	509	1,96	0,86	0,00	.972

Anmerkung. T1 = 1.629.

Tabelle 33. Wohnsituation

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Alleine	238	24,7	151	23,1	0,04	.497
Mit Partner	514	53,3	372	56,9	-0,07	.173
Mit Kindern	347	36,0	279	42,7	-0,13	.008
Bei Eltern	88	9,1	44	6,7	0,09	.101
Bei anderen Verwandten	4	0,4	1	0,2	-	-
Wohngemeinschaft	24	2,5	13	2,0	0,03	.622

Anmerkung. T1 = 1.629.

Tabelle 34. Höchster allgemeinbildender Schulabschluss

	IG		KG		Vergleich	
	N	%	N	%	d	p
Abgang ohne Hauptschulabschluss	10	1,1	3	0,5	0,14	.236
Hauptschulabschluss (Volksschulabschluss)	109	11,7	92	14,3		
Realschule (Mittlere Reife)	283	30,3	190	29,5		
Polytechnische Oberschule	6	0,6	3	0,5		
Fachhochschulreife	114	12,2	91	14,1		
Allgemeine Hochschulreife	405	43,3	255	39,6		
Anderer Schulabschluss	8	0,9	10	1,6		
Gesamt	935	100,0	644	100,0		

Anmerkung. T1 = 1.629.

Tabelle 35. Höchster Berufsabschluss

	IG		KG		Vergleich	
	N	%	N	%	d	p
Noch in Ausbildung (Azubi oder Student)	99	10,6	53	8,2	0,21	.015
Kein beruflicher Abschluss	46	4,9	35	5,4		
Betriebliche Ausbildung (Lehre)	256	27,5	223	34,5		
Berufsfachschule, Handelsschule	110	11,8	60	9,3		
Fach-, Meisterschule, Berufsakademie	143	15,4	83	12,8		
Fachhochschulabschluss	96	10,3	63	9,7		
Hochschulabschluss	169	18,2	112	17,3		
Anderer Abschluss	12	1,3	18	2,8		
Gesamt	931	100,0	647	100,0		

Anmerkung. T1 = 1.629.

Tabelle 36. Derzeitige oder zuletzt ingehabte berufliche Stellung

	IG		KG		Vergleich	
	N	%	N	%	d	p
Selbständiger Landwirt	3	0,3	0	0,0	0,20	.031
Akademiker in freiem Beruf	25	2,7	17	2,7		
Selbständig in Handel, Gewerbe, Handwerk	45	4,8	31	4,9		
Beamter, Richter, Berufssoldat	5	0,5	4	0,6		
Angestellter	680	73,0	467	73,3		
Arbeiter	29	3,1	37	5,8		
In Ausbildung	128	13,7	63	9,9		
Mithelfender Familienangehöriger	16	1,7	18	2,8		
Gesamt	931	100,0	637	100,0		

Anmerkung. T1 = 1.629.

Tabelle 37. Derzeitige oder zuletzt ingehabte berufliche Stellung

	IG		KG		Vergleich	
	N	%	N	%	d	p
Vollzeit erwerbstätig (mindestens 35h)	410	43,2	272	42,1	0,10	.605
Teilzeit erwerbstätig (15-34h)	145	15,3	104	16,1		
Teilzeit unter 15h	79	8,3	61	9,4		
Mutterschafts-/Erziehungsurlaub	20	2,1	16	2,5		
Auszubildender	34	3,6	14	2,2		
Nicht erwerbstätig	260	27,4	179	27,7		
Gesamt	948	100,0	646	100,0		

Anmerkung. T1 = 1.629.

Tabelle 38. Monatliches Nettoeinkommen des Haushaltes

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Haushaltsnettoeinkommen in Euro	460	2135	1294	422	2315	1270	-0,14	.037

Anmerkung. T1 = 1.629.

Tabelle 39. Haupteinkommensquelle

	IG		KG		Vergleich	
	N	%	N	%	d	p
Eigene Arbeitstätigkeit	484	51,5	319	49,3	0,23	.009
Eigene Rente	23	2,4	21	3,2		
Einkommen des (Ehe-) Partners	224	23,8	192	29,7		
Unterstützung durch Eltern	66	7,0	39	6,0		
Arbeitslosengeld	46	4,9	19	2,9		
Unterhalt des Ehegatten (bei Trennung)	11	1,2	18	2,8		
BaföG	13	1,4	3	0,5		
Sozialhilfe	16	1,7	9	1,4		
Sonstiges	47	5,0	23	3,6		
Unbekannt	10	1,1	4	0,6		
Gesamt	940	100,0	647	100,0		

Anmerkung. T1 = 1.629.

### C) Arbeitsfähigkeit

Keine statistisch signifikanten Unterschiede ergeben sich zwischen IG und KG bei der Arbeitsfähigkeit zu Therapiebeginn. Etwa ¼ aller Patienten sind bei Therapiebeginn nach Angaben der Therapeuten arbeitsfähig. 47,6% der IG und 43,1% der KG gaben Arbeitsunfähigkeitszeiten im Jahr vor der Behandlung an (ebenfalls kein statistisch signifikanter Unterschied).

Tabelle 40. Arbeitsfähigkeit bei Erstkontakt

	IG		KG		Vergleich	
	N	%	N	%	d	p
Ja	715	76,2	485	75,9	0,01	.881
Nein	223	23,8	154	24,1		
Gesamt	938	100,0	639	100,0		

Anmerkung. T1 = 1.629.

Tabelle 41. Durchgehende Arbeitsfähigkeit bei Erstkontakt

	IG		KG		Vergleich	
	N	%	N	%	d	p
Ja	468	52,4	352	56,9	-0,09	.087
Nein	425	47,6	267	43,1		
Gesamt	893	100,0	619	100,0		

Anmerkung. T1 = 1.629.

### D) Krankheitsanamnese

Keine signifikanten Unterschiede zwischen IG und KG bestehen auch hinsichtlich Art und Umfang von ambulanten und stationären Vorbehandlungen in der Anamnese, ein Indikator für den Schweregrad und die Chronizität des Krankheitsbildes. Etwa 30% aller Patienten hat bereits Erfahrungen mit ambulanter Psychotherapie in der Anamnese gemacht.

Tabelle 42. Anzahl der ambulanten psychotherapeutischen / psychiatrischen Vorbehandler in der Anamnese

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Anzahl Vorbehandler	964	0,57	0,95	654	0,57	1,02	0,00	.913

Anmerkung. T1 = 1.629.

Tabelle 43. Ambulante psychische Vorbehandlungen in der Anamnese

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Ambulante Psychotherapie	313	32,5	189	28,9	0,08	.128
Facharzt Psychiatrie / Neurologie	203	21,1	115	17,6	0,09	.084
Psychosomatische Grundversorgung	30	3,1	10	1,5	0,10	.044

Anmerkung. T1 = 1.629.

Tabelle 44. Dauer der ambulanten Psychotherapie in der Anamnese (Sitzungen insgesamt)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Ambulante Psychotherapiesitzungen (Stunden) in der Anamnese	964	9,41	27,28	654	9,06	26,67	0,01	.801

Anmerkung. T1 = 1.629.

Tabelle 45. Stationäre psychische Vorbehandlungen in der Anamnese

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Psychosomatische Klinik	129	13,4	92	14,1	-0,02	.694
Psychiatrische Klinik	80	8,3	62	9,5	-0,04	.410

Anmerkung. T1 = 1.629.

Tabelle 46. Stationäre psychische Vorbehandlungen in der Anamnese – Anzahl der Aufnahmen insgesamt

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Aufnahmen Psychosomatische Klinik	964	0,17	0,51	654	0,19	0,58	-0,04	.444
Aufnahmen Psychiatrische Klinik	964	0,37	1,66	654	0,45	1,83	-0,05	.358

Anmerkung. T1 = 1.629.

Tabelle 47. Stationäre psychische Vorbehandlungen in der Anamnese – Behandlungstage insgesamt

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Behandlungstage Psychosomatische Klinik	964	4,84	29,53	654	5,50	23,23	-0,02	.630
Behandlungstage Psychiatrische Klinik	964	8,54	55,23	654	9,09	40,14	-0,01	.358

Anmerkung. T1 = 1.629. Behandlungstage = Im Fragebogen dokumentierte Behandlungsmonate x 30,5.

Tabelle 48. Ambulante somatische Vorbehandlung in den letzten 12 Monaten vor Therapiebeginn

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Ambulante Vorbehandlung Hausarzt	584	60,6	404	61,8	-0,02	.629
Ambulante Vorbehandlung Facharzt	366	38,0	244	37,3	0,01	.879

Anmerkung. T1 = 1.629.

Tabelle 49. Stationäre somatische Vorbehandlung in den letzten 12 Monaten vor Therapiebeginn

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Behandlung Akut-Klinik letzte 12 Monate	94	9,8	75	11,5	-0,06	.268
Behandlung Reha-Klinik letzte 12 Monate	45	4,7	42	6,4	-0,08	.125

Anmerkung. T1 = 1.629.

Tabelle 50. Stationäre somatische Vorbehandlung in den letzten 12 Monaten vor Therapiebeginn – Anzahl Aufnahmen

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Aufnahmen Akut-Klinik letzte 12 Monate	964	0,15	0,63	654	0,14	0,51	0,02	.891
Aufnahmen Reha-Klinik letzte 12 Monate	964	0,05	0,23	654	0,06	0,26	-0,04	.213

Anmerkung. T1 = 1.629.

Tabelle 51. Stationäre somatische Vorbehandlung in den letzten 12 Monaten vor Therapiebeginn – Behandlungstage

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Behandlungstage Akut-Klinik	964	1,84	12,29	654	2,35	10,89	-0,04	.389
Behandlungstage Reha-Klinik	964	1,67	10,69	654	2,26	9,78	-0,06	.253

Anmerkung. T1 = 1.629. Behandlungstage = Im Fragebogen dokumentierte Behandlungsmonate x 30,5.

#### 4.2.2 Therapieeingangsdagnostik

Insgesamt sind  $T1_B = 1.572$  ausgefüllte Fragebögen vorhanden (IG: 919, KG: 653).

Im Rahmen der probatorischen Sitzungen wurde von den teilnehmenden Therapeutinnen und Therapeuten eine diagnostische Einordnung der Patienten gemäß den Kriterien der ICD-10 (WHO, 2010) vorgenommen. Die Therapeuten konnten dabei bis zu vier Diagnosen vergeben. In der Interventionsgruppe erfolgte dies unter Zuhilfenahme der IDCL-Checklisten (Hiller, et al., 1995), während sich die Therapeuten der Kontrollgruppe wie herkömmlich alleine an den ICD-10-Kriterien des diagnostischen Manuals bzw. ihrer bisherigen Praxis orientieren. Bei der Anzahl und Art der vergebenen Diagnosen handelt es sich aufgrund der unterschiedlichen Vorgehensweise in IG und KG also nicht nur um ein reines Prädiktor-Merkmal der PR-Box, sondern auch um ein Prozess-Merkmal, das der NTR-Box zugeordnet werden kann.

Für 1.544 Patienten (98,2% von 1.572) liegt mindestens eine gültige ICD-10-Diagnose vor. Tabelle 52 gibt wieder, wie viele Diagnosen im Durchschnitt vergeben wurden. In der IG wurden statistisch signifikant mehr ICD-10-Diagnosen codiert als in der KG ( $d = 0,29$ ). Dies wird besonders deutlich, wenn man sich die Häufigkeitsverteilung anschaut: So wurde bei der IG in 48,8% aller Fälle mindestens eine weitere Diagnose vergeben, während dies bei der KG in lediglich in 34,8% aller Fälle geschah. Dritt- und Viertdiagnosen sind weniger häufig, werden in der IG aber insgesamt etwa doppelt so oft vergeben wie in der KG. Für die beobachteten Unterschiede sind zwei Erklärungen denkbar:

- Ordnet man die ICD-10-Diagnostik der PR-Box zu, so könnten die zu beobachtenden Unterschiede darauf zurückgeführt werden, dass in der IG vor Therapiebeginn tatsächlich eine höhere Komorbidität als in der KG besteht.
- Ordnet man die ICD-10-Diagnostik der NTR-Box zu, so führt das modifizierte Vorgehen im TK-Modell möglicherweise zu einer vertieften Auseinandersetzung mit der ICD-10 und damit zu einer umfangreicheren bzw. differenzierteren Diagnostik.

Die Anzahl der Eingangsdiagnosen als Maß für die Komorbidität sollte also neben dem Alter als weitere Kovariate für die Analysen der Ergebnisqualität Berücksichtigung finden.

Tabelle 52. Anzahl der Eingangsdiaagnosen nach ICD-10

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Anzahl der Eingangsdiaagnosen	919	1,71	0,92	653	1,46	0,77	0,29	<.001
<i>Häufigkeitsverteilung</i>								
	N	%		N	%			
Keine gültige Diagnose vorhanden	13	1,4		15	2,3			
Eine Diagnose	457	49,7		410	62,8			
Zwei Diagnosen	292	31,8		153	23,4			
Drei Diagnosen	95	10,3		63	9,6			
Vier Diagnosen	62	6,7		12	1,8			
Gesamt	919	100,0		653	100,0			

Anmerkung. T1 = 1.629. Als gültige Diagnosen wurden alle Codierungen definiert, bei denen neben dem führenden Buchstaben F aus dem ICD-10-Kapitel „Psychische und Verhaltensstörungen“ zumindest auch die ersten beiden Ziffern vorhanden waren (also beispielsweise F32 oder F32.1). Fehlte die Nachkommastelle, so wurde grundsätzlich die Ziffer 9 als Nachkommastelle „nicht näher bezeichnet“ angenommen. In 212 Fällen fehlte bei der Erstdiagnose allerdings der führende Buchstabe (d.h. es wurde beispielsweise 32.1 anstatt F32.1 vom Therapeuten angegeben). Da somatische Erkrankungen aus den Kapiteln A-E bzw. G-Z indikationsgemäß nur sehr selten vorkommen, wurde in solchen Fällen grundsätzlich davon ausgegangen, dass es sich um eine psychische Diagnose aus dem F-Kapitel handelte.

In Tabelle 53 ist die Verteilung der Erstdiagnosen (Hauptgruppen des ICD-10-F-Kapitels) wiedergegeben. Die Gruppen F3 und F4 machen mehr als 90% aller Hauptdiagnosen aus. Tabelle 54 bis Tabelle 56 zeigen die entsprechenden Verteilungen für die Zweit-, Dritt- und Viertdiagnosen.

Vergleicht man die Verteilungen der Erst- und Zweitdiagnosen inhaltlich, so wird deutlich, dass – wenn eine Zweitdiagnose gestellt wurde – der Anteil der Diagnosen aus dem Kapitel F4 gegenüber dem Kapitel F3 bei den Zweitdiagnosen deutlich höher ist als bei den Erstdiagnosen. So ergab sich bei einer detaillierteren Analyse der Kombinationen von Erst- und Zweitdiagnose, dass bei einer Komorbidität von Depressiver Störung und Angststörung die Depressive Störung bevorzugt als Hauptdiagnose und die Angststörung als Nebendiagnose gewählt wird und nicht umgekehrt. Auch Persönlichkeitsstörungen und Somatisierungsstörungen werden eher an zweiter als an erster Stelle verschlüsselt.

Tabelle 53. Erste Diagnose (ICD-Buchstabe und erste Ziffer)

		IG		KG		Vergleich	
		N	%	N	%	d	p
F0	Organische psychische Störungen	0	0,0	0	0,0	-	-
F1	Psychische Störungen durch psychotrope Substanzen	3	0,3	5	0,8	-	-
F2	Schizophrenie und wahnhaftige Störungen	1	0,1	6	0,9	-	-
F3	Affektive Störungen	411	45,4	283	44,4	0,02	.695
F4	Neurotische, Belastungs- und Anpassungsstörungen	425	46,9	304	47,6	-0,01	.774
F5	Verhaltensauffälligkeiten mit körperlichen Störungen	30	3,3	24	3,8	-0,02	.635
F6	Persönlichkeits- und Verhaltensstörungen	33	3,6	11	1,7	0,11	.026
F7	Intelligenzstörungen	0	0,0	0	0,0	-	-
F8	Entwicklungsstörungen	0	0,0	1	0,2	-	-
F9	Störungen mit Beginn in Kindheit und Jugend	3	0,3	0	0,0	-	-
??	Somatische Diagnose (A-E, G-Z)	0	0,0	4	0,6	-	-
	Gesamt	906	100,0	638	100,0		

Anmerkung. T1 = 1.629. Für jede Diagnosegruppe wurde zunächst eine neue binäre Variable erstellt. In dieser wurde codiert, ob die betreffende Diagnosegruppe angegeben wurde (=1) oder nicht (=0). Auf diese Weise ist ein Vergleich der Häufigkeiten zwischen IG und KG für jede einzelne Diagnosegruppe möglich.

Tabelle 54. Zweite Diagnose (ICD-Buchstabe und erste Ziffer)

		IG		KG		Vergleich	
		N	%	N	%	d	p
F0	Organische psychische Störungen	0	0,0	0	0,0	-	-
F1	Psychische Störungen durch psychotrope Substanzen	18	4,0	5	2,2	0,09	.218
F2	Schizophrenie und wahnhaftige Störungen	2	0,4	1	0,4	-	-
F3	Affektive Störungen	113	25,2	55	24,1	0,02	.766
F4	Neurotische, Belastungs- und Anpassungsstörungen	236	52,6	105	46,1	0,12	.109
F5	Verhaltensauffälligkeiten mit körperlichen Störungen	25	5,6	8	3,5	0,09	.240
F6	Persönlichkeits- und Verhaltensstörungen	41	9,1	36	15,8	-0,20	.010
F7	Intelligenzstörungen	0	0,0	0	0,0	-	-
F8	Entwicklungsstörungen	1	0,2	0	0,0	-	-
F9	Störungen mit Beginn in Kindheit und Jugend	7	1,6	5	2,2	-0,05	.555
??	Somatische Diagnose (A-E, G-Z)	6	1,3	13	5,7	-0,25	.001
	Gesamt	449	100,0	228	100,0		

Anmerkung. T1 = 1.629. Für jede Diagnosegruppe wurde zunächst eine neue binäre Variable erstellt. In dieser wurde codiert, ob die betreffende Diagnosegruppe angegeben wurde (=1) oder nicht (=0). Auf diese Weise ist ein Vergleich der Häufigkeiten zwischen IG und KG für jede einzelne Diagnosegruppe möglich.

Tabelle 55. Dritte Diagnose (ICD-Buchstabe und erste Ziffer)

		IG		KG		Vergleich	
		N	%	N	%	d	p
F0	Organische psychische Störungen	0	0,0	0	0,0	-	-
F1	Psychische Störungen durch psychotrope Substanzen	11	7,0	8	10,7	-0,12	.342
F2	Schizophrenie und wahnhaftige Störungen	0	0,0	0	0,0	-	-
F3	Affektive Störungen	29	18,5	10	13,3	0,13	.328
F4	Neurotische, Belastungs- und Anpassungsstörungen	64	40,8	31	41,3	-0,01	.934
F5	Verhaltensauffälligkeiten mit körperlichen Störungen	12	7,6	11	14,7	-0,22	.094
F6	Persönlichkeits- und Verhaltensstörungen	29	18,5	7	9,3	0,23	.072
F7	Intelligenzstörungen	0	0,0	0	0,0		-
F8	Entwicklungsstörungen	0	0,0	0	0,0		-
F9	Störungen mit Beginn in Kindheit und Jugend	2	1,3	1	1,3		-
??	Somatische Diagnose (A-E, G-Z)	10	6,4	7	9,3	-0,11	.418
	Gesamt	157	100,0	75	100,0		

Anmerkung. T1 = 1.629. Für jede Diagnosegruppe wurde zunächst eine neue binäre Variable erstellt. In dieser wurde codiert, ob die betreffende Diagnosegruppe angegeben wurde (=1) oder nicht (=0). Auf diese Weise ist ein Vergleich der Häufigkeiten zwischen IG und KG für jede einzelne Diagnosegruppe möglich.

Tabelle 56. Vierte Diagnose (ICD-Buchstabe und erste Ziffer)

		IG		KG		Vergleich	
		N	%	N	%	d	p
F0	Organische psychische Störungen	0	0,0	0	0,0	-	-
F1	Psychische Störungen durch psychotrope Substanzen	3	4,8	2	16,7	-	-
F2	Schizophrenie und wahnhaftige Störungen	0	0,0	0	0,0	-	-
F3	Affektive Störungen	7	11,3	0	0,0	-	-
F4	Neurotische, Belastungs- und Anpassungsstörungen	26	41,9	2	16,7	-	-
F5	Verhaltensauffälligkeiten mit körperlichen Störungen	5	8,1	1	8,3	-	-
F6	Persönlichkeits- und Verhaltensstörungen	15	24,2	2	16,7	-	-
F7	Intelligenzstörungen	0	0,0	0	0,0	-	-
F8	Entwicklungsstörungen	0	0,0	0	0,0	-	-
F9	Störungen mit Beginn in Kindheit und Jugend	1	1,6	0	0,0	-	-
??	Somatische Diagnose (A-E, G-Z)	5	8,1	5	41,7	-0,67	.002
	Gesamt	62	100,0	12	100,0		

Anmerkung. T1 = 1.629. Für jede Diagnosegruppe wurde zunächst eine neue binäre Variable erstellt. In dieser wurde codiert, ob die betreffende Diagnosegruppe angegeben wurde (=1) oder nicht (=0). Auf diese Weise ist ein Vergleich der Häufigkeiten zwischen IG und KG für jede einzelne Diagnosegruppe möglich.

In Tabelle 57 ist die Verteilung der Erstdiagnosen in der nächsthöheren Auflösungsstufe wiedergegeben. Depressive Störungen, Angststörungen und Anpassungsstörungen sind bei weitem am häufigsten vertreten. Diese haben auch die größte Prävalenz in der Bevölkerung, weshalb diese Verteilung zu erwarten ist (Wittchen, et al., 1998); (Wittchen, et al., 1999). Beim Gruppenvergleich fällt auf, dass in der IG die Diagnose Depressive Episode (F32) mit 17,9% weniger häufig vergeben wurde als in der KG mit 24,1%.

Dagegen zeigt sich bei der rezidivierenden depressiven Störung (F33) eine Tendenz zu einer größeren Häufigkeit in der IG mit 18,4% gegenüber der KG mit 14,3%. Ein bedeutsamer Unterschied ergibt sich auch bei den anhaltenden affektiven Störungen (F34). Während diese in der IG 8,2% aller Hauptdiagnosen ausmachen, so liegt der Anteil in der KG lediglich bei 5,2% und Phobische Störungen werden in der IG mit 9,4% häufiger diagnostiziert als in der KG mit 5,8%.

In Tabelle 58 bis Tabelle 60 finden sich die entsprechenden Verteilungen für die Zweit-, Dritt- und Viertdiagnose. Es fällt auf, dass bei den Zweitdiagnosen die Phobische Störung (F40) in der IG mit 15,8% mehr als doppelt so häufig gestellt wird wie in der KG mit 6,1%. Bezüglich der spezifischen Persönlichkeitsstörungen (F60) ergibt sich kein eindeutiges Bild: Während bei der Erstdiagnose eine Tendenz zur häufigeren Diagnose in der IG besteht, welche sich auch bei Dritt- und Viertdiagnosen findet, werden unter den Zweitdiagnosen Persönlichkeitsstörungen häufiger in der KG vergeben.

Für die oft vertretenen Diagnosekapitel F32, F33, F34, F40, F41 und F43 werden darüber hinaus die Häufigkeiten der ersten Nachkommastelle der ICD-10 für die Erstdiagnosen in Tabelle 61 mitgeteilt. Es ist ersichtlich, dass Anpassungsstörungen (F43.2) die häufigste Erstdiagnose ausmachen, gefolgt von der mittelgradigen depressiven Episode (F32.1), der mittelgradigen rezidivierenden depressiven Störung (F33.1) und der Dysthymia (F34.1).

Ein Großteil der Patienten kommt also mit einem depressiven Störungsbild in die Behandlung, wenn man sich vergegenwärtigt, dass die Anpassungsstörung (F43.2) in der ICD-10 die frühere Diagnose Depressive Reaktion (309) aus der ICD-9 abgelöst hat. Nimmt man die Kapitel F32, F33, F34.1 und F43.2 zusammen, so entspricht dies einem Anteil von insgesamt 917 Patienten (59,4% von 1.544).

Tabelle 57. Erste Diagnose (ICD-Buchstabe und erste zwei Ziffern)

	IG		KG		Vergleich	
	N	%	N	%	d	p
F10 Psychische und Verhaltensstörungen durch Alkohol	3	0,3	4	0,6	-	-
F19 Psychische und Verhaltensstörungen andere Substanzen	0	0,0	1	0,2	-	-
F20 Schizophrenie	1	0,1	4	0,6	-	-
F23 Akute vorübergehende psychotische Störungen	0	0,0	1	0,2	-	-
F25 Schizoaffective Störungen	0	0,0	1	0,2	-	-
F30 Manische Episode	0	0,0	1	0,2	-	-
F31 Bipolare Affektive Störung	8	0,9	4	0,6	-	-
F32 Depressive Episode	162	17,9	154	24,1	-0,15	.003
F33 Rezidivierende depressive Störung	167	18,4	91	14,3	0,11	.031
F34 Anhaltende Affektive Störungen	74	8,2	33	5,2	0,12	.022
F40 Phobische Störungen	85	9,4	37	5,8	0,13	.010
F41 Andere Angststörungen	100	11,0	80	12,5	-0,05	.365
F42 Zwangsstörungen	16	1,8	12	1,9	-0,01	.868
F43 Belastungs- und Anpassungsstörungen	174	19,2	147	23,0	-0,09	.067
F44 Dissoziative Störungen	7	0,8	2	0,3	-	-
F45 Somatoforme Störungen	40	4,4	23	3,6	0,04	.428
F48 Andere neurotische Störungen	3	0,3	3	0,5	-	-
F50 Essstörungen	26	2,9	16	2,5	0,02	.667
F51 Nichtorganische Schlafstörungen	0	0,0	1	0,2	-	-
F52 Sexuelle Funktionsstörungen	1	0,1	2	0,3	-	-
F54 Psychische Faktoren bei anderen Krankheiten	3	0,3	5	0,8	-	-
F60 Spezifische Persönlichkeitsstörungen	29	3,2	9	1,4	0,11	.025
F62 Andauernde Persönlichkeitsänderungen	1	0,1	0	0,0	-	-
F63 Abnorme Gewohnheiten und Störungen der Impulskontrolle	0	0,0	2	0,3	-	-
F65 Störungen der Sexualpräferenz	2	0,2	0	0,0	-	-
F66 Verhaltensstörungen in der sexuellen Entwicklung	1	0,1	0	0,0	-	-
F80 Entwicklungsstörungen des Sprechens und der Sprache	0	0,0	1	0,2	-	-
F90 Hyperkinetische Störungen	2	0,0	0	0,0	-	-
F98 Andere Verhaltensstörungen mit Beginn in Kindheit und Jugend	1	0,0	0	0,0	-	-
??? Somatische Diagnose (A-E, G-Z)	0	0,0	4	0,6	-	-
Gesamt	906	100,0	638	100,0		

Anmerkung. T1 = 1.629. Für jede Diagnosegruppe wurde zunächst eine neue binäre Variable erstellt. In dieser wurde codiert, ob die betreffende Diagnosegruppe angegeben wurde (=1) oder nicht (=0). Auf diese Weise ist ein Vergleich der Häufigkeiten zwischen IG und KG für jede einzelne Diagnosegruppe möglich. Nur Diagnosen mit n > 0 in IG bzw. KG sind aufgeführt.

Tabelle 58. Zweite Diagnose (ICD-Buchstabe und erste zwei Ziffern)

	IG		KG		Vergleich	
	N	%	N	%	d	p
F10 Psychische und Verhaltensstörungen durch Alkohol	14	3,1	4	1,8	-	-
F12 Psychische und Verhaltensstörungen durch Cannabinoide	1	0,2	0	0,0	-	-
F13 Psychische und Verhaltensstörungen Sedativa u. Hypnotika	1	0,2	0	0,0	-	-
F17 Psychische und Verhaltensstörungen durch Tabak	2	0,4	1	0,4	-	-
F20 Schizophrenie	2	0,4	1	0,4	-	-
F31 Bipolare Affektive Störung	0	0,0	2	0,9	-	-
F32 Depressive Episode	48	10,7	27	11,8	-0,03	.652
F33 Rezidivierende depressive Störung	27	6,0	11	4,8	0,05	.525
F34 Anhaltende affektive Störungen	37	8,2	15	6,6	0,06	.443
F38 Andere affektive Störungen	1	0,2	0	0,0	-	-
F40 Phobische Störungen	71	15,8	14	6,1	0,27	<.001
F41 Andere Angststörungen	56	12,5	29	12,7	-0,01	.927
F42 Zwangsstörungen	10	2,2	7	3,1	-0,05	.508
F43 Belastungs- und Anpassungsstörungen	62	13,8	28	12,3	0,04	.580
F44 Dissoziative Störungen	2	0,4	1	0,4	-	-
F45 Somatoforme Störungen	33	7,3	26	11,4	-0,14	.077
F48 Andere neurotische Störungen	2	0,4	0	0,0	-	-
F50 Essstörungen	15	3,3	3	1,3	-	-
F51 Nichtorganische Schlafstörungen	2	0,4	4	1,8	-	-
F52 Sexuelle Funktionsstörungen	1	0,2	0	0,0	-	-
F54 Psychische Faktoren bei anderen Krankheiten	7	1,6	1	0,4	-	-
F60 Spezifische Persönlichkeitsstörungen	40	8,9	32	14,0	-0,16	.041
F61 Kombinierte und andere Persönlichkeitsstörungen	0	0,0	2	0,9	-	-
F63 Abnorme Gewohnheiten und Störungen der Impulskontrolle	1	0,2	2	0,9	-	-
F81 Entwicklungsstörungen schulischer Fertigkeiten	1	0,2	0	0,0	-	-
F90 Hyperkinetische Störungen	4	0,9	4	1,8	-	-
F91 Störungen des Sozialverhaltens	0	0,0	1	0,4	-	-
F93 Emotionale Störungen des Kindesalters	1	0,2	0	0,0	-	-
F98 Andere Verhaltensstörungen mit Beginn in Kindheit und Jugend	2	0,4	0	0,0	-	-
??? Somatische Diagnose (A-E, G-Z)	6	1,3	13	5,7	-0,25	.001
Gesamt	449	100,0	228	100,0		

Anmerkung. T1 = 1.629. Für jede Diagnosegruppe wurde zunächst eine neue binäre Variable erstellt. In dieser wurde codiert, ob die betreffende Diagnosegruppe angegeben wurde (=1) oder nicht (=0). Auf diese Weise ist ein Vergleich der Häufigkeiten zwischen IG und KG für jede einzelne Diagnosegruppe möglich. Nur Diagnosen mit n > 0 in IG bzw. KG sind aufgeführt.

Tabelle 59. Dritte Diagnose (ICD-Buchstabe und erste zwei Ziffern)

	IG		KG		Vergleich	
	N	%	N	%	d	p
F10 Psychische und Verhaltensstörungen durch Alkohol	5	3,2	1	1,3	-	-
F12 Psychische und Verhaltensstörungen durch Cannabinoide	4	2,5	0	0,0	-	-
F13 Psychische und Verhaltensstörungen Sedativa u. Hypnotika	1	0,6	2	2,7	-	-
F17 Psychische und Verhaltensstörungen durch Tabak	1	0,6	4	5,3	-	-
F19 Psychische und Verhaltensstörungen andere Substanzen	0	0,0	1	1,3	-	-
F30 Manische Episode	1	0,6	0	0,0	-	-
F32 Depressive Episode	11	7,0	4	5,3	-	-
F33 Rezidivierende depressive Störung	7	4,5	4	5,3	-	-
F34 Anhaltende affektive Störungen	10	6,4	2	2,7	-	-
F40 Phobische Störungen	26	16,6	7	9,3	0,19	.140
F41 Andere Angststörungen	7	4,5	5	6,7	-0,09	.478
F42 Zwangsstörungen	8	5,1	2	2,7	-	-
F43 Belastungs- und Anpassungsstörungen	6	3,8	5	6,7	-0,13	.340
F44 Dissoziative Störungen	1	0,6	2	2,7	-	-
F45 Somatoforme Störungen	16	10,2	8	10,7	-0,01	.911
F48 Andere neurotische Störungen	0	0,0	2	2,7	-	-
F50 Essstörungen	5	3,2	5	6,7	-0,16	.222
F51 Nichtorganische Schlafstörungen	2	1,3	4	5,3	-	-
F54 Psychische Faktoren bei anderen Krankheiten	5	3,2	2	2,7	-	-
F60 Spezifische Persönlichkeitsstörungen	28	17,8	4	5,3	-	-
F61 Kombinierte und andere Persönlichkeitsstörungen	0	0,0	2	2,7	-	-
F65 Störungen der Sexualpräferenz	0	0,0	1	1,3	-	-
F66 Verhaltensstörungen der sexuellen Entwicklung	1	0,6	0	0,0	-	-
F90 Hyperkinetische Störungen	1	0,6	0	0,0	-	-
F98 Andere Verhaltensstörungen mit Beginn in Kindheit und Jugend	1	0,6	1	1,3	-	-
??? Somatische Diagnose (A-E, G-Z)	10	6,4	7	9,3	-0,11	.418
Gesamt	157	100,0	75	100,0		

Anmerkung. T1 = 1.629. Für jede Diagnosegruppe wurde zunächst eine neue binäre Variable erstellt. In dieser wurde codiert, ob die betreffende Diagnosegruppe angegeben wurde (=1) oder nicht (=0). Auf diese Weise ist ein Vergleich der Häufigkeiten zwischen IG und KG für jede einzelne Diagnosegruppe möglich. Nur Diagnosen mit n > 0 in IG bzw. KG sind aufgeführt.

Tabelle 60. Vierte Diagnose (ICD-Buchstabe und erste zwei Ziffern)

	IG		KG		Vergleich	
	N	%	N	%	d	p
F10 Psychische und Verhaltensstörungen durch Alkohol	0	0,0	2	16,7	-	-
F12 Psychische und Verhaltensstörungen durch Cannabinoide	1	1,6	0	0,0	-	-
F17 Psychische und Verhaltensstörungen durch Tabak	1	1,6	0	0,0	-	-
F19 Psychische und Verhaltensstörungen andere Substanzen	1	1,6	0	0,0	-	-
F32 Depressive Episode	4	6,5	0	0,0	-	-
F33 Rezidivierende depressive Störung	2	3,2	0	0,0	-	-
F34 Anhaltende affektive Störungen	1	1,6	0	0,0	-	-
F40 Phobische Störungen	5	8,1	0	0,0	-	-
F41 Andere Angststörungen	4	6,5	1	8,3	-	-
F42 Zwangsstörungen	6	9,7	0	0,0	-	-
F43 Belastungs- und Anpassungsstörungen	5	8,1	0	0,0	-	-
F45 Somatoforme Störungen	6	9,7	1	8,3	-	-
F50 Essstörungen	2	3,2	1	8,3	-	-
F51 Nichtorganische Schlafstörungen	1	1,6	0	0,0	-	-
F54 Psychische Faktoren bei anderen Krankheiten	2	3,2	0	0,0	-	-
F60 Spezifische Persönlichkeitsstörungen	15	24,2	0	0,0	-	-
F61 Kombinierte und andere Persönlichkeitsstörungen	0	0,0	2	16,7	-	-
F90 Hyperkinetische Störungen	1	1,6	0	0,0	-	-
??? Somatische Diagnose (A-E, G-Z)	5	8,1	5	41,7	-0,67	.002
Gesamt	62	100,0	12	100,0		

Anmerkung. T1 = 1.629. Für jede Diagnosegruppe wurde zunächst eine neue binäre Variable erstellt. In dieser wurde codiert, ob die betreffende Diagnosegruppe angegeben wurde (=1) oder nicht (=0). Auf diese Weise ist ein Vergleich der Häufigkeiten zwischen IG und KG für jede einzelne Diagnosegruppe möglich. Nur Diagnosen mit n > 0 in IG bzw. KG sind aufgeführt.

Tabelle 61. Erste Diagnose. Differenzierte Darstellung der am häufigsten dokumentierten Untergruppen

	IG		KG		Vergleich	
	N	%	N	%	d	p
F32.0 Leichte depressive Episode	35	3,9	27	4,2	-0,02	.716
F32.1 Mittelgradige depressive Episode	105	11,6	108	16,9	-0,15	.003
F32.2 Schwere depressive Episode ohne psychotische Symptome	13	1,4	7	1,1	0,03	.563
F32.3 Schwere depressive Episode mit psychotischen Symptomen	1	0,1	2	0,3	-	-
F32.9 Depressive Episode nicht näher bezeichnet	8	0,9	10	1,6	-0,06	.217
F33.0 Leichte rezidivierende depressive Störung	36	4,0	25	3,9	0,00	.956
F33.1 Mittelgradige rezidivierende depressive Störung	94	10,4	56	8,8	0,05	.297
F33.2 Schwere rez. depressive Störung ohne psychotische Symptome	25	2,8	8	1,3	0,10	.044
F33.3 Schwere rez. depressive Störung mit psychotischen Symptomen	3	0,3	0	0,0	-	-
F33.9 Rezidivierende depressive Störung nicht näher bezeichnet	9	1,0	2	0,3	-	-
F34.0 Zylothymia	3	0,3	2	0,3	-	-
F34.1 Dysthymia	65	7,2	28	4,4	0,12	.023
F34.9 Anhaltende Affektive Störung nicht näher bezeichnet	6	0,7	3	0,5	-	-
F40,0 Agoraphobie	35	3,9	22	3,4	0,02	.670
F40,1 Soziale Phobie	25	2,8	6	0,9	0,13	.012
F40,2 Spezifische (isolierte) Phobien	20	2,2	9	1,4	0,06	.256
F40,9 Phobische Störung, nicht näher bezeichnet	5	0,6	0	0,0	-	-
F41.0 Panikstörung	39	4,3	27	4,4	0,00	.945
F41.1 Generalisierte Angststörung	39	4,3	15	2,4	0,10	.040
F41.2 Angst und Depression gemischt	15	1,7	34	5,3	-0,20	<.001
F41.9 Angststörung, n.n. bez.	7	0,8	4	0,6	-	-
F43.0 Akute Belastungsreaktion	8	0,9	14	2,2	-0,02	.738
F43.1 Posttraumatische Belastungsstörung	27	3,0	9	1,4	0,10	.044
F43.2 Anpassungsstörungen	131	14,5	116	18,2	-0,10	.049
F43.9 Reaktion auf schwere Belastung, nicht näher bezeichnet	8	0,9	8	1,3	-0,04	.479
Summe	762	84,5	542	85,1		
Gesamt	906	100,0	638	100,0		

Anmerkung. T1 = 1.629. Für jede Diagnosegruppe wurde zunächst eine neue binäre Variable erstellt. In dieser wurde codiert, ob die betreffende Diagnosegruppe angegeben wurde (=1) oder nicht (=0). Auf diese Weise ist ein Vergleich der Häufigkeiten zwischen IG und KG für jede einzelne Diagnosegruppe möglich. Nur die häufig vertretenen Kapitel F32, F33, F34, F40, F41 und F43 sind hier aus Platzgründen aufgeführt.

**Tabelle 62. Akute Suizidalität**

	IG		KG		Vergleich	
	N	%	N	%	d	p
Nein	834	92,3	600	94,2	-0,09	.222
Fraglich	66	7,3	33	5,2		
Ja	4	0,4	4	0,6		
Gesamt	904	100,0	637	100,0		

Anmerkung. T1 = 1.629.

**Tabelle 63. Erkrankungsdauer und Beginn der ersten professionellen Behandlung**

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Erstmanifestation der Erkrankung vor x Jahren	788	8,34	9,09	596	6,83	8,13	0,17	.001
Beginn der ersten professionellen Behandlung vor x Jahren	647	5,26	6,62	548	4,06	5,58	0,19	<.001
Aktuelle Manifestation der Erkrankung seit x Jahren	694	2,32	3,67	495	2,30	3,01	0,01	.913

Anmerkung. T1 = 1.629.

**Tabelle 64. Art der ersten professionellen Behandlung**

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Ambulante Psychotherapie	435	47,3	362	55,4	-0,16	.002
Stationäre Psychotherapie	50	5,4	45	6,9	-0,06	.234
Ambulant psychiatrisch (auch pharmakologisch)	136	14,8	89	13,6	0,03	.514
Stationär psychiatrisch	35	3,8	31	4,7	-0,05	.360
Andere ambulante Therapie	96	10,4	56	8,6	0,06	.216
Andere stationäre Therapie	24	2,6	20	3,1	-0,03	.593

Anmerkung. T1 = 1.629.

**Tabelle 65. Art der letzten professionellen Behandlung**

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Ambulante Psychotherapie	282	30,7	229	35,1	-0,09	.067
Stationäre Psychotherapie	59	6,4	42	6,4	0,00	.992
Ambulant psychiatrisch (auch pharmakologisch)	153	16,6	92	14,1	0,07	.168
Stationär psychiatrisch	26	2,8	25	3,8	-0,06	.270
Andere ambulante Therapie	97	10,6	62	9,5	0,03	.492
Andere stationäre Therapie	25	2,7	21	3,2	-0,03	.566

Anmerkung. T1 = 1.629.

**Tabelle 66. Therapiemotivation des Patienten**

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Therapiemotivation	907	1,86	0,63	643	1,91	0,67	-0,08	.103

Anmerkung. T1 = 1.629. Codierung: 1 = sehr hoch, 2 = hoch, 3 = mittel, 4 = gering.

Weitere Aspekte, die in der Therapieeingangsdagnostik abgefragt wurden, beziehen sich auf eine etwaige akut vorhandene Suizidalität, das Jahr der Erstmanifestation der Erkrankung, Beginn und Art der ersten professionellen Behandlung, die Art der letzten professionellen Behandlung, die Dauer der aktuellen Manifestation der Erkrankung sowie die Therapiemotivation. Keine statistisch signifikanten Unterschiede zwischen IG und KG bestehen hinsichtlich Suizidalität Dauer der aktuellen Krankheitsmanifestation, der Therapiemotivation und Art der ersten bzw. letzten professionellen Behandlung mit einer Ausnahme: Bei den Patienten der KG war die ambulante Psychotherapie als erste professionelle Behandlungsform mit 55,4% etwas häufiger als bei den Patienten der IG mit 47,3%. Hinsichtlich Erstmanifestation der Erkrankung und Beginn der ersten professionellen Behandlung wurden allerdings statistisch signifikante Mittelwertsunterschiede dokumentiert: In der IG liegen sowohl Erstmanifestation als auch Beginn der ersten Behandlung ca. 1,5 Jahre weiter zurück als in der KG.

#### 4.2.3 Erfassung der Therapieziele mit dem Berner Inventar für Therapieziele (BIT-CT)

Zu Beginn und zum Ende der Therapie wurden die Therapieziele mit der BIT-CT erfasst, einer eigens für die vorliegende Studie revidierten Form der von Grosse Holtforth und Grawe (2002) entwickelten Zielcheckliste des Berner Inventars für Therapieziele. Die Therapeuten waren aufgefordert, die Therapieziele des jeweiligen Patienten in der 64 Therapieziele umfassenden BIT-CT anzukreuzen. Nicht in der BIT-CT vorkommende Ziele konnten freitextlich ergänzt werden. Im Anschluss sollten die vier wichtigsten Ziele ausgewählt und ins Hogrefe Testsystem übertragen werden. Die 64 Ziele der BIT-CT gliedern sich in 28 Therapiezielkategorien und 5 Therapiezielarten. Da die Therapieziele ausschließlich in der Interventionsgruppe erhoben wurden, beziehen sich die nachfolgenden Ausführungen ausschließlich auf die IG.

In der Interventionsgruppe liegen für N = 859 Patienten (93,5% von 919 IG-Patienten der Stichprobe T1<sub>B</sub>) überhaupt Therapieziele vor. Für 855 Patienten liegen mindestens zwei, für 848 Patienten mindestens drei und für 786 Patienten liegen vier Therapieziele vor. Insgesamt wurden von den Therapeuten der Interventionsgruppe K = 3.348 Therapieziele angegeben. In Tabelle 67 ist dargestellt, für wie viele der 859 Patienten der Interventionsgruppe, für die Therapieziele vorliegen, Therapieziele aus den fünf Therapiezielarten des BIT-CT jeweils angegeben werden.

**Tabelle 67. Verfolgte Therapiezielarten.**

Therapiezielart nach Anteil der Patienten, die sie verfolgen	N	Anteil	
Bewältigung bestimmter Probleme und Symptome	755	87,9%	
Ziele im zwischenmenschlichen Bereich	576	67,1%	
Selbstbezogene Ziele	576	55,4%	
Orientierung im Leben	205	23,9%	
Verbesserung des Wohlbefindens	196	22,8%	
Sonstige, im BIT-CT nicht formulierte, Ziele	25	2,9%	

**Anmerkung.** N = 859 Patienten, für die Therapieziele vorliegen.

Es wird deutlich, dass die Bewältigung bestimmter Probleme und Symptome wie zu erwarten am häufigsten angestrebt wird. Auch Ziele im zwischenmenschlichen Bereich und selbstbezogene Ziele liegen häufig im Fokus. Die Möglichkeit, weitere, im BIT-CT nicht formulierte Ziele zu definieren, wurde nur für eine geringe Anzahl an Patienten (2.9%) genutzt.



In Tabelle 69 ist aufgeführt, welche Therapieziele aus dem Katalog am häufigsten genannt wurden. Die ausgewählten Themen zielen auf eine Reduktion der häufigsten Symptome (Depression und Angst) ab. Auch bewältigungsorientierte Ziele werden oft genannt, die sich um die Entwicklung von sozialer Kompetenz wie Beziehungs- und Abgrenzungsfähigkeit oder den Umgang mit konkreten Problemen im privaten und beruflichen Umfeld drehen.

**Tabelle 69. Häufig genannte Therapieziele aus dem BIT-CT**

Nr.	Therapieziel	N
2	Aus gedrückter Stimmung, Traurigkeit und innerer Leere herauskommen	231
1	Negative, kreisende Gedanken oder Schuldgefühle überwinden	187
54	Mehr Selbstvertrauen und Selbstsicherheit entwickeln	182
39	Sich anderen gegenüber besser durchsetzen und abgrenzen lernen	127
34	Die Beziehung zu seinen Eltern verändern (Ablösung, Abhängigkeit, Schuldgefühle)	113
4	Wieder mehr Antrieb und Energie bekommen	107
8	Angst- und Panikanfälle in den Griff zu bekommen	103
7	Eine konkrete Angst bewältigen oder besser mit ihr umgehen	99
10	Wieder Dinge tun, die man jetzt aus Angst vermeidet	92
30	Seine Erwartungen und Gefühle bezüglich Partner klären	82
3	Mit Stimmungsschwankungen besser umgehen	81
25	Konkrete Probleme mit seiner Arbeit oder Ausbildung bewältigen	79
40	Mit Reaktionen anderer (Kritik, Ablehnung, Lob) besser umgehen lernen	77
13	Traumatische Erlebnisse verarbeiten (Unfall, Gewaltverbrechen)	74
52	Neue Zukunftsperspektiven (private oder berufliche) entwickeln	73
61	Seine hohen Ansprüche an sich und andere herabsetzen lernen	70
27	Besser mit Stresssituationen umgehen lernen	68
50	Mit Teilen seiner Vergangenheit besser zurechtkommen lernen	68
64	Mit starken negativen Gefühlen (z.B. Ärger, Wutausbrüchen) umgehen lernen	66
51	Sich klarer werden, wer er ist, was er kann und was er will	66
42	Nähe zulassen und Vertrauen zu anderen Menschen aufbauen	64
28	Die Beziehung mit seinem Partner verbessern	64
48	Mehr Optimismus und Lebensfreude entwickeln	64
63	Gefühle zulassen und äußern lernen	63
36	Die Trennung von seinem Ex-Partner verarbeiten	62
56	Seine Bedürfnisse wahrnehmen und ausdrücken lernen	62
47	Lernen, Probleme und Herausforderungen gelassener anzugehen	58
57	Seine Grenzen besser erkennen und danach handeln lernen	57
20	Schlafprobleme (Ein- oder Durchschlafen, frühes Erwachen) bewältigen	52

Anmerkung. N = 859 Patienten; K = 3.348 Therapieziele (bis zu vier Nennungen pro Patient möglich). Mindestens 50x genannte Ziele sind hier aufgeführt, absteigend nach Häufigkeit sortiert.

#### 4.2.4 Psychopharmakologische Behandlung

Insgesamt liegt für T1<sub>c</sub> = 1.532 Patienten (IG: 898, KG: 634) die Therapeutendokumentation der psychopharmakologischen Behandlung bei Aufnahme vor. Hinsichtlich der Frage, ob überhaupt eine Medikation zu Therapiebeginn „mitgebracht“ wird, ergeben sich keine Unterschiede zwischen IG und KG (Tabelle 70).

In Tabelle 71 ist wiedergegeben, ob bestimmte Medikamentengruppen unter den bis zu sieben möglichen Nennungen vom Therapeuten erwähnt wurden. Antidepressiva führen die Liste klar an, die von jedem vierten Patient zum Zeitpunkt des Therapiebeginns eingenommen werden. Alle anderen Medikamente sind zahlenmäßig deutlich seltener vertreten. Zwischen IG und KG ergeben sich keine signifikanten Unterschiede in der Art der Medikation (einziger Ausreißer sind die antipsychotischen Neuroleptika, die in der IG mit 1,1% seltener eingenommen werden als in der KG mit 2,8%, die Fallzahlen sind hier aber zu gering, um praktisch bedeutsame Unterschiede konstituieren zu können). Zählt man zusammen, wie viele Medikamente pro Patient genannt wurden, so ergeben sich ebenfalls keine signifikanten Unterschiede zwischen IG und KG (Tabelle 72).

**Tabelle 70. Derzeitige psychopharmakologische Behandlung (bei Therapiebeginn)**

	IG		KG		Vergleich	
	N	%	N	%	d	p
Ja	314	35,0	235	37,1	-0,04	.399
Nein	584	65,0	399	62,9		
Gesamt	898	100,0	634	100,0		

Anmerkung. T1 = 1.629.

**Tabelle 71. Medikation bei Therapiebeginn**

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Antidepressivum	237	26,3	164	25,7	0,01	.816
Lithium bzw. Carbamazepin	6	0,7	8	1,3	0,06	.229
Neuroleptika antipsychotische Zielsetzung	10	1,1	18	2,8	-0,13	.013
Tranquilizer	25	2,8	18	2,8	0,00	.946
Neuroleptika Tranquilizer-Zielsetzung	16	1,8	12	1,9	-0,01	.873
Hypnotika	8	0,9	10	1,6	-0,06	.219
Antikonvulsivum	4	0,4	7	1,1	-	-
Hirnstoffwechselaktives Medikament	1	0,1	2	0,3	-	-
Parkinsonmittel bzw. Hyperkinetikum	0	0,0	1	0,2	-	-
Analgetika	16	1,8	8	1,3	0,04	.420
Antiphlogistika und Spasmolytika	1	0,1	0	0,0	-	-
Homöopathie, Phytotherapeutika	18	2,0	13	2,1	0,00	.947
Sonstige / Andere Psychopharmaka	17	1,9	8	1,1	0,05	.337

Anmerkung. T1 = 1.629.

**Tabelle 72. Anzahl der genannten Medikamente (bis zu sieben Nennungen möglich)**

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Anzahl der genannten Medikamente	898	0,39	0,63	634	0,42	0,67	-0,05	.432

Anmerkung. T1 = 1.629.

#### 4.2.5 Therapeuten-Ratings bei Therapiebeginn

Insgesamt sind  $T1_D = 1.574$  Fragebogen mit den Therapeuten-Ratings zum Therapiebeginn vorhanden (IG: 916, KG: 658). Tabelle 73 stellt die Ergebnisse für die IG und KG vergleichend dar.

Einzelne Unterschiede bestehen beim Problemrating Familiärer Bereich (IG besser als KG;  $d = 0,14$ ) sowie beim BSS sozialkommunikativ (IG schlechter als KG;  $d = -0,16$ ).

Auf dem Multiplen Statusindikator STA\_TH, der ein Aggregat aus 13 Aspekten darstellt (vgl. Abschnitt 5.3.2) zeigen sich allerdings keine statistisch signifikanten Unterschiede zwischen IG und KG. Dies bedeutet, dass sich beide Gruppen in ihrem Allgemeinbefinden (bzw. allgemeinen Schweregrad) aus Sicht der Therapeuten zu Therapiebeginn nicht voneinander unterscheiden. Die beiden statistisch signifikanten Einzelvergleiche sollten auf diesem Hintergrund daher nicht überinterpretiert werden.

Tabelle 73. Therapeuten-Ratings bei Therapiebeginn

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Problemrating Beruf, Ausbildung	907	4,94	3,33	654	4,91	3,30	-0,01	.844
Problemrating Partnerschaft	911	5,90	3,06	647	6,16	3,12	0,08	.099
Problemrating Familie	909	5,56	2,78	652	5,96	2,71	0,14	.005
Problemrating Freizeit	901	4,06	2,64	643	4,26	2,62	0,07	.147
Problemrating Sozial (Wohnen, Finanzen)	903	3,38	2,78	654	3,27	2,90	-0,04	.460
Problemrating Selbstverantwortung	908	4,27	2,75	651	4,19	2,66	-0,03	.561
Problemrating Gesundheitsverhalten	901	3,86	2,63	641	3,94	2,75	0,03	.552
BSS letzte 7 Tage körperlich	908	1,59	1,05	651	1,59	1,05	0,00	.954
BSS letzte 7 Tage psychisch	908	2,73	0,77	651	2,67	0,74	-0,08	.099
BSS letzte 7 Tage sozialkommunikativ	907	1,95	0,99	649	1,79	0,98	-0,16	.002
GAF Funktionsniveau aktuell	852	56,4	10,4	621	56,6	10,9	-0,02	.764
<b>Multipler Statusindikator STA_TH<sup>a</sup></b>	<b>900</b>	<b>55,3</b>	<b>13,4</b>	<b>646</b>	<b>55,0</b>	<b>12,7</b>	<b>0,02</b>	<b>.750</b>

Anmerkungen.  $T1 = 1.629$ . Codierungen: Problemratings 0 (problemlos) – 10 (sehr problematisch); Beeinträchtigung BSS: 0 (gar nicht) – 4 (extrem); Funktionsniveau GAF: 1 (sehr schlecht) – 100 (sehr gut); <sup>a</sup> Die Konstruktion der aus 13 Items bestehenden Skala STA\_TH, die wie der GAF einen möglichen Wertebereich von 0 (sehr schlecht) – 100 (sehr gut) Punkten umfasst, ist in Abschnitt 5.3.2 beschrieben.

#### 4.3 Patientenangaben bei Therapiebeginn (Stichprobe P1)

Insgesamt liegt für  $P1 = 1.598$  Patienten die Erstmessung auf den psychometrischen Fragebögen vor. Dies entspricht einem Anteil von 93,6% von 1.708. In der IG sind mit 1.011 (98,1% von 1.031) mehr Erstmessungen vorhanden als in der KG mit 587 (86,7% von 677). Der entsprechende Vierfelder-Chi<sup>2</sup>-Test ist statistisch signifikant ( $p < .001$ ,  $d = 0,45$ ).

Nachfolgend werden die Ergebnisse der Überprüfung der Randomisierung für die Selbstangaben der Patienten berichtet. Die beiden Variablen Alter und Geschlecht liegen alle 1.598 Patienten der Stichprobe P1 vor. Es zeigen sich keine Unterschiede zwischen IG und KG hinsichtlich des Geschlechts ( $p = .608$ ;  $d = 0,02$ ). Beim Alter zeigt sich wieder der leichte Trend dahingehend, dass die Patienten der IG mit 40,3 Jahren ( $SD = 11,4$ ) etwas jünger sind als die Patienten der KG mit 41,6 Jahren ( $SD = 11,2$ ). Der entsprechende t-Test ist mit  $p = .030$  statistisch signifikant ( $d = -0,11$ ).

Tabelle 74 zeigt den Vergleich der Erstmessung für die IG und KG. Es ist festzustellen, dass sich die Patienten der IG zum Therapiebeginn etwas stärker beeinträchtigt darstellen als die Patienten der KG. Die Effektgröße ist allerdings klein mit einem Wert von  $-0,17$ .

Tabelle 74. Patientenangaben bei Therapiebeginn (A-Messung, Rohwerte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	1000	1,26	0,67	577	1,17	0,66	-0,14	.042
IIP-D Gesamtwert	1006	1,58	0,55	582	1,51	0,55	-0,13	.078
SF12 Körperliche Befinden	1002	46,32	10,87	581	47,35	10,50	-0,10	.084
SF12 Psychisches Befinden	1002	32,42	10,55	581	33,21	10,66	-0,07	.214
BDI Gesamtwert	711	21,26	9,09	444	19,55	9,37	-0,19	.008
AKV Gesamtwert	226	2,21	0,62	84	2,23	0,56	-0,03	.752
EDI Gesamtwert	34	338,91	54,52	13	348,40	42,61	0,18	.577
HZK Gesamtwert	30	25,21	11,40	13	19,13	8,66	-0,57	.095
SOM Gesamtwert	67	36,78	24,75	20	33,98	20,84	-0,12	.690
<b>Multipler Statusindikator STA_PA<sup>a</sup></b>	<b>1009</b>	<b>54,4</b>	<b>9,9</b>	<b>586</b>	<b>56,1</b>	<b>10,0</b>	<b>-0,17</b>	<b>.001</b>

Anmerkung. P1 = 1.598. Codierung: BSI und IIP-D: 0 (keine Belastung) – 4 (starke Belastung). SF-12-Skalen: 0 (schlechtes Befinden) – 100 (gutes Befinden). BDI: 0 (keine Depressivität) – 63 (starke Depressivität). AKV: 1 (keine Belastung) – 5 (starke Belastung). EDI: 91 (keine Belastung) – 546 (starke Belastung). HZK: 0 (keine Belastung) – 72 (starke Belastung). SOM 0 (keine Belastung) – 208 (starke Belastung). Die p-Werte wurden unter Berücksichtigung der Intra-Class-Korrelationen berechnet.

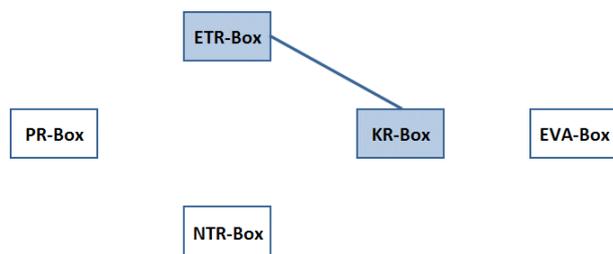
<sup>a</sup> Der Multiple Statusindikator STA\_PA mit einem Wertebereich von 0 (schlechtes Befinden) – 100 (gutes Befinden) wurde aus den lineartransformierten Skalen BSI, IIP-D, SF12k, SF12p sowie der lineartransformierten jeweils störungsspezifischen Skala BDI, AKV, EDI, HZK oder SOM gebildet (vgl. Abschnitt 5.4.2).

#### 4.4 Zusammenfassende Bewertung

Insgesamt finden sich keine gravierenden Verletzungen der internen Validität durch systematische Unterschiede zwischen IG und KG. Dies bedeutet, dass die Randomisierung erfolgreich war und beide Gruppen hinsichtlich wichtiger Merkmale zum Therapiebeginn miteinander vergleichbar sind. Bei dem verwendeten Versuchsplan handelt es sich ja um ein clusterrandomisiertes Design („Place Randomized Trial“). Die Einheit der Randomisierung war der Therapeut und nicht der Versicherte. Die Randomisierung hat sich aber von einigen kleinen Ausnahmen abgesehen auch auf die Patientenebene übertragen. Die nachfolgend aufgeführten kleineren Unterschiede wurden zwischen IG und KG beobachtet:

- Die Patienten der IG sind im Durchschnitt um ein Jahr jünger als in der KG, was sich zum Teil auch in der privaten und beruflichen Situation der Studienteilnehmer widerspiegelt.
- In der IG werden mehr Diagnosen vergeben als in der KG, was möglicherweise auf die intensivere Diagnostik mit der IDCL und Therapiezielliste durch das Modellvorhaben in der IG bedingt ist und somit auch ein Prozessmerkmal der NTR-Box darstellt. Dazu passt, dass in der IG im Durchschnitt eine halbe probatorische Sitzung mehr durchgeführt wird als in der KG und die Therapeuten bei der IG im Durchschnitt ein um anderthalb Jahre früheres Erstauftreten der Erkrankung in der Anamnese feststellen.
- Bei den Patienten der IG wurden im Durchschnitt mehr abgeschlossene Behandlungen pro Therapeut dokumentiert als in der KG (der entsprechende Vergleich zwischen IG und KG findet sich bei der Evaluation der Mikroebene in Abschnitt 7.1.4).
- Depressive Episoden werden in der IG im Vergleich zur KG etwas seltener als Hauptdiagnose festgestellt, dafür werden rezidivierende depressive Störungen, Dysthymia und phobische Störungen etwas häufiger diagnostiziert. Ob dies auf tatsächlich bereits vor Studienbeginn vorhandene Unterschiede oder aber die modifizierte Vorgehensweise bei der IG zurückzuführen ist, lässt sich an dieser Stelle nicht beantworten.
- Der Allgemeinbefinden bei Therapiebeginn auf dem multiplen Statusindikator STA\_PA ist aus Sicht der Patienten in der IG etwas schlechter ( $d = -0,17$ ) als in der KG. Aus Sicht der Therapeuten bestehen diesbezüglich auf dem multiplen Statusindikator STA\_TH keine statistisch signifikanten Unterschiede zwischen der IG und KG.

## 5 Wirksamkeit (Makroebene)



### 5.1 Abschlussdokumentation der Therapeuten (Stichprobe T2)

Insgesamt liegt für 934 Patienten (Stichprobe T2; IG: 564, KG: 370 Patienten) die Abschlussdokumentation der Therapeuten vor. Nachfolgend werden die Verteilungen für die IG und KG vergleichend beschrieben.

#### 5.1.1 Behandlungsverlauf und Situation bei Therapieende

Hinsichtlich Behandlungsunterbrechungen, etwaiger stationärer Behandlungstage und begleitender medizinischer Behandlungen und Komplikationen ergeben sich keine auffälligen Unterschiede zwischen IG und KG im Therapieverlauf, mit Ausnahme etwas häufigerer gynäkologischer Behandlungen ( $d = 0,15$ ) sowie körperlicher Erkrankungen ( $d = 0,14$ ) in der IG. Auch hinsichtlich möglicher Veränderungen der beruflichen Situation bei Therapieende sowie den empfohlenen Nachfolmaßnahmen bestehen keine Unterschiede zwischen IG und KG.

Tabelle 75. Behandlungsunterbrechung

	IG		KG		Vergleich	
	N	%	N	%	d	p
Ja	58	11,7	51	14,9	-0,09	.175
Nein	439	88,3	292	85,1		
Gesamt	497	100,0	343	100,0		

Anmerkung. T2 = 934.

Tabelle 76. Falls Behandlungsunterbrechung durch stationären / teilstationären Aufenthalt: Behandlungstage

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Behandlungstage stationär / teilstationär	561	3,03	11,4	365	3,63	12,5	-0,05	.449

Anmerkung. T2 = 934.

Tabelle 77. Begleitende medizinische Behandlung

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Psychiatrisch	108	19,1	61	16,5	0,07	.301
Neurologisch	37	6,6	18	4,9	0,07	.282
Orthopädisch	51	9,0	24	6,5	0,09	.160
Hausärztlich	250	44,3	169	45,7	-0,03	.685
Internistisch	40	7,1	27	7,3	-0,01	.905
Gynäkologisch	34	6,0	10	2,7	0,15	.019
Dermatologisch	11	2,0	4	1,1	-	-

Anmerkung. T2 = 934.

Tabelle 78. Komplikationen während der Behandlung

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Keine	274	48,6	199	53,8	-0,10	.120
Akute Suizidalität	9	1,6	6	1,6	0,00	.975
Suizidversuche	3	0,5	2	0,5	-	-
Selbstschädigungen	7	1,2	4	1,1	-	-
Beruf	81	14,4	58	15,7	-0,04	.581
Alkohol, Drogen, Medikamente	20	3,5	16	4,3	-0,04	.546
Straftat	2	0,4	1	0,3	-	-
Trennung, Scheidung	56	9,9	35	9,5	0,02	.813
Finanzen	36	6,4	20	5,4	0,04	.538
Körperliche Erkrankung	69	12,2	29	7,8	0,14	.032

Anmerkung. T2 = 934.

Tabelle 79. Berufliche Situation bei Therapieende

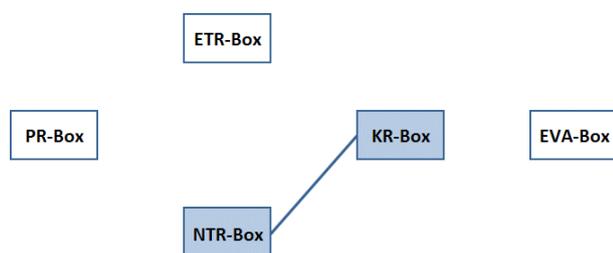
	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Erwerbstätigkeit unverändert	363	64,4	241	65,1	-0,02	.809
Arbeitsplatzwechsel	60	10,6	36	9,7	0,03	.655
Neu / wieder erwerbstätig	42	7,4	30	8,1	-0,02	.711
Erhöhte Arbeitszeit	31	5,5	13	3,5	0,09	.162
Verminderte Arbeitszeit	8	1,4	8	2,2	-0,06	.392
Nicht mehr erwerbstätig	23	4,1	11	3,0	0,06	.778
Neu in Ausbildung / Fortbildung	18	3,2	16	4,3	-0,06	.366
Neu geschützter Arbeitsplatz	8	1,4	2	0,5	-	-
Neu Altersrente / Pension	8	1,4	1	0,3	-	-
Neu arbeitslos	10	1,8	4	1,1	-	-

Anmerkung. T2 = 934.

Tabelle 80. Nachfolmaßnahmen / Überweisung

	IG		KG		Vergleich	
	N	%	N	%	d	p
<i>Mehrere Angaben sind möglich</i>						
Nachfolmaßnahmen keine	425	74,4	283	76,5	-0,02	.693
Ambulante Psychotherapie	30	5,3	18	4,9	0,02	.758
Stationäre Psychotherapie	3	0,5	2	0,5	-	-
Ambulante psychiatrische Behandlung	30	5,3	20	5,4	0,00	.954
Stationäre psychiatrische Behandlung	0	0,0	1	0,3	-	-
Rehabilitationsmaßnahme	4	0,7	4	1,1	-	-
Selbsthilfegruppe	15	2,7	8	2,2	0,03	.631
Psychosozialer Dienst	2	0,4	3	0,8	-	-

Anmerkung. T2 = 934.



### 5.1.2 Therapieziele zum Ende der Therapie

Zum Therapieende wurden nochmals mit der BIT-CT die vier wichtigsten Therapieziele der Patienten erhoben. Hier liegen für insgesamt für 542 Patienten der IG Therapieziele vor. Für 317 Patienten mit zu Beginn formulierten Therapiezielen liegen für diesen Zeitpunkt keine Ziele (mehr) vor. Für 466 Patienten werden zum Therapieende exakt die gleichen Ziele wie zu Beginn angegeben, für insgesamt 76 Patienten wurde eine Änderung vorgenommen.

Zugleich mit der Erfassung der Therapieziele zu Therapieende wurde beim Therapeuten für jedes Therapieziel auch der Grad der Zielerreichung auf einer vierstufigen Skala erfragt. Insgesamt haben die Therapeuten von 544 Patienten (96,5% von 564) aus der Interventionsgruppe 2.092 Therapieziele bzgl. des Grades ihrer Erreichung bewertet. Tabelle 81 gibt den Zielerreichungsgrad für die bis zu vier genannten Ziele sowie insgesamt wieder. Bildet man explorativ aus den vier Zielerreichungsskalierungen eine Gesamtskala, so hat diese einen Mittelwert von 2,51 (SD = 0,58; Cronbachs Alpha = .77). Dies bedeutet, dass aus Sicht der Therapeuten die Ziele ihrer Patienten im Durchschnitt 'teilweise erreicht' bis 'erreicht' wurden.

Tabelle 81. Therapiezielerreichung im BIT-CT (Therapeuteneinschätzungen) in der Interventionsgruppe

Zielerreichungsgrad	Ziel 1	Ziel 2	Ziel 3	Ziel 4	Gesamt	
<b>1 = Therapieziel nicht erreicht</b>	42 7,8%	46 8,6%	50 9,5%	48 9,9%	186 8,9%	
<b>2 = Therapieziel teilweise erreicht</b>	177 32,8%	202 37,7%	189 35,7%	201 41,3%	769 36,8%	
<b>3 = Therapieziel erreicht</b>	288 53,3%	259 48,3%	257 48,6%	212 43,5%	1016 48,6%	
<b>4 = Therapieziel mehr als erreicht</b>	33 6,1%	29 5,4%	33 6,2%	26 5,3%	121 5,8%	
	540 100,0%	536 100,0%	529 100,0%	487 100,0%	2092 100,0%	

## 5.2 Angaben der Therapeuten zur OPD (optional, nur IG)

Insgesamt in 228 Fällen wurde von den Therapeuten der Interventionsgruppe die Zusatzoption genutzt, zum Aufnahmezeitpunkt die operationalisierte psychodynamische Diagnostik (OPD) zu bearbeiten. 68,4% der Patienten sind weiblich. Der Altersdurchschnitt liegt bei 40,3 Jahren (SD = 11,2).

### 5.2.1 Zu Therapiebeginn ausgewählte Konflikt- und Strukturbereiche

Die Therapeuten konnten bis zu vier Fokusbereiche bei Therapiebeginn angeben. Tabelle 82 zeigt die Ergebnisse. Unter 750 Einzelnennungen wurden am häufigsten Konflikte im Hinblick auf Abhängigkeit / Autonomie genannt (K1), gefolgt von Selbstwertkonflikten (K4) und Konflikten, die sich um Versorgung / Autarkie (K3) drehen. Auch in der Strukturliste ging es häufig um Selbstwertregulation (St22) und Loslösung (St62).

Tabelle 82. Zu Therapiebeginn in der OPD ausgewählte Konflikt- und Strukturbereiche

Fokusliste Konflikte / Struktur	Fokus 1	Fokus 2	Fokus 3	Fokus 4	Gesamt
K1 Abhängigkeit/Autonomie	69 30,4%	7 3,1%	2 1,1%	3 2,8%	81 10,8%
K2 Unterwerfung/Kontrolle	20 8,8%	12 5,3%	1 0,5%	2 1,8%	35 4,7%
K3 Versorgung/Autarkie	45 19,8%	14 6,2%	3 1,6%	2 1,8%	64 8,5%
K4 Selbstwertkonflikte	42 18,5%	29 12,8%	5 2,7%	1 0,9%	77 10,3%
K5 Über-Ich und Schuldkonflikte	14 6,2%	26 11,5%	2 1,1%	2 1,8%	44 5,9%
K6 Ödipal-sexuelle Konflikte	2 0,9%	5 2,2%	1 0,5%	0 0,0%	8 1,1%
K7 Identitätskonflikte	7 3,1%	4 1,8%	0 0,0%	0 0,0%	11 1,5%
K8 Fehlende Konflikt- und Gefühlswahrnehmung	3 1,3%	6 2,6%	2 1,1%	0 0,0%	11 1,5%
K9 Konflikthafte äußere Lebensbelastung	5 2,2%	12 5,3%	4 2,1%	0 0,0%	21 2,8%

(Fortsetzung der Tabelle auf der nächsten Seite)

Tabelle (Fortsetzung). Zu Therapiebeginn in der OPD ausgewählte Konflikt- und Strukturbereiche

Fokusliste Konflikte / Struktur	Fokus 1	Fokus 2	Fokus 3	Fokus 4	Gesamt
St11 Selbstreflexion	2 0,9%	6 2,6%	6 3,2%	0 0,0%	14 1,9%
St12 Selbstbild	0 0,0%	10 4,4%	8 4,3%	5 4,6%	23 3,1%
St13 Identität	1 0,4%	3 1,3%	1 0,5%	1 0,9%	6 0,8%
St14 Affektdifferenzierung	1 0,4%	17 7,5%	12 6,4%	1 0,9%	31 4,1%
St21 Affekttoleranz	3 1,3%	9 4,0%	11 5,9%	5 4,6%	28 3,7%
St22 Selbstwertregulation	5 2,2%	21 9,3%	27 14,4%	3 2,8%	56 7,5%
St23 Impulssteuerung	5 2,2%	8 3,5%	14 7,5%	3 2,8%	30 4,0%
St24 Antizipation	0 0,0%	1 0,4%	2 1,1%	0 0,0%	3 0,4%
St31 Intrapsychische Abwehr	0 0,0%	3 1,3%	14 7,5%	10 9,2%	27 3,6%
St32 Flexibilität der Abwehr	0 0,0%	4 1,8%	8 4,3%	4 3,7%	16 2,1%
St41 Selbst-Objekt-Differenzierung	1 0,4%	2 0,9%	10 5,3%	7 6,4%	20 2,7%
St42 Empathie	0 0,0%	1 0,4%	1 0,5%	1 0,9%	3 0,4%
St43 Ganzheitliche Objektwahrnehmung	0 0,0%	2 0,9%	13 7,0%	3 2,8%	18 2,4%
St44 Objektbezogene Affekte	0 0,0%	1 0,4%	3 1,6%	2 1,8%	6 0,8%
St51 Kontakt	0 0,0%	1 0,4%	5 2,7%	7 6,4%	13 1,7%
St52 Verstehen fremder Affekte	0 0,0%	1 0,4%	8 4,3%	0 0,0%	9 1,2%
St53 Mitteilen eigener Affekte	0 0,0%	6 2,6%	7 3,7%	7 6,4%	20 2,7%
St54 Reziprozität	0 0,0%	0 0,0%	2 1,1%	3 2,8%	5 0,7%
St61 Internatlisierung	0 0,0%	4 1,8%	4 2,1%	4 3,7%	12 1,6%
St62 Loslösung	2 0,9%	10 4,4%	10 5,3%	24 22,0%	46 6,1%
St63 Variabilität der Bindung	0 0,0%	2 0,9%	1 0,5%	9 8,3%	12 1,6%
Gesamt	227 100,0%	227 100,0%	187 100,0%	109 100,0%	750 100,0%

## 5.2.2 Konstruktion der Fokusskala

Für jeden der bis zu vier Fokusbereiche konnten die Therapeuten zu Therapiebeginn und bei Therapieende eine Lokalisierung des Patienten auf der insgesamt 19 Feinabstufungen umfassenden Fokusskala vornehmen. Diese reicht von einer Abwehr des Fokus bis hin zu dessen Auflösung und lässt sich im Sinne des psychotherapeutischen Prozessgeschehens interpretieren, so dass sich im Prä/Post-Vergleich entsprechende Effektgrößen berechnen lassen. Hierzu wurden die Ankreuzkategorien zunächst einer Recodierung unterzogen, um die statistischen Datenanalysen durchführen zu können.

Die auf einen Wertebereich von 1-19 rekodierten (bis zu vier Items) wurden aufsummiert und durch die Anzahl der beantworteten Items geteilt, woraus der Gesamtskalenwert resultiert, der damit ebenfalls einen Wertebereich von 1-19 Punkten aufweist. Die interne Konsistenz der aus den bis zu vier Items berechneten Skala (Cronbachs Alpha) beträgt .81 und ist somit zufriedenstellend.

Bei immerhin 95 Patienten (41,7% von 228) liegt neben der Messung zu Therapiebeginn auch eine entsprechende Einschätzung zum Therapieende vor. Hier sind entsprechende Prä/Post-Vergleiche auf der Fokusskala möglich. Tabelle 83 veranschaulicht die Vorgehensweise und die resultierende Häufigkeitsverteilung für die Prä- und Post-Messung.

Tabelle 83. Fokusskala OPD vor und nach der Behandlung.

Fokusbezeichnung	Codierung Original	Codierung Neu	Häufigkeit Prä	Häufigkeit Post
Abwehr des Fokus	1	1	0	0
Abwehr des Fokus	1+	2	2	0
Ungewollte Beschäftigung	2-	3	9	2
Ungewollte Beschäftigung	2	4	19	6
Ungewollte Beschäftigung	2+	5	15	5
Vage Fokuswahrnehmung	3-	6	22	4
Vage Fokuswahrnehmung	3	7	13	11
Vage Fokuswahrnehmung	3+	8	9	10
Anerkennung und Erkundung	4-	9	5	11
Anerkennung und Erkundung	4	10	0	12
Anerkennung und Erkundung	4+	11	1	14
Auflösung alter Strukturen	5-	12	0	4
Auflösung alter Strukturen	5	13	0	7
Auflösung alter Strukturen	5+	14	0	2
Neuordnung des Fokusbereichs	6-	15	0	6
Neuordnung des Fokusbereichs	6	16	0	1
Neuordnung des Fokusbereichs	6+	17	0	0
Auflösung des Fokus	7-	18	0	0
Auflösung des Fokus	7	19	0	0

Anmerkung. N = 95 vollständige Prä-/Post-Messungen mit der OPD. Diese wurde nur von einem Teil der Therapeuten der Interventionsgruppe optional genutzt, woraus sich die vergleichsweise geringe Stichprobengröße erklärt.

### 5.2.3 Prä-Post-Vergleich auf der Fokusskala

Abbildung 13 veranschaulicht die Entwicklung im Therapieverlauf. Während sich zu Beginn der Therapie die meisten Patienten mit einem Mittelwert von 5,61 noch mehrheitlich im Bereich „Ungewollte Beschäftigung“ bis „Vage Fokuswahrnehmung“ befinden, hat sich die Verteilung zum Therapieende deutlich nach rechts verschoben und der durchschnittliche Patient befindet sich im Bereich „Anerkennung und Erkundung“ ( $M = 9,28$ ). Einige Patienten haben sich sogar bis in den Bereich „Auflösung von Strukturen“ bzw. „Neuordnung des Fokusbereichs“ entwickelt.

Die Patienten haben hier aus Sicht ihrer Therapeuten einerseits somit deutliche Schritte in Richtung mehr Selbsterkenntnis und Weiterentwicklung ihrer psychodynamischen Struktur unternommen, sind andererseits zum Therapieende mehrheitlich aber auch noch „unterwegs“ und noch nicht am Ende dieses innerhalb der Therapie initiierten Prozesses angekommen.

Berechnet man die Prä-Post-Effektgröße für die Fokusskala, so ergibt sich ein beeindruckend hoher Wert von  $d = (9,28 - 5,61) / 1,81 = 2,03$ . Die Grenze für einen großen Effekt nach Cohen liegt bereits bei  $d > 0,80$ . Die hier explorativ gebildete Fokusskala weist somit eine hohe Änderungssensitivität auf.

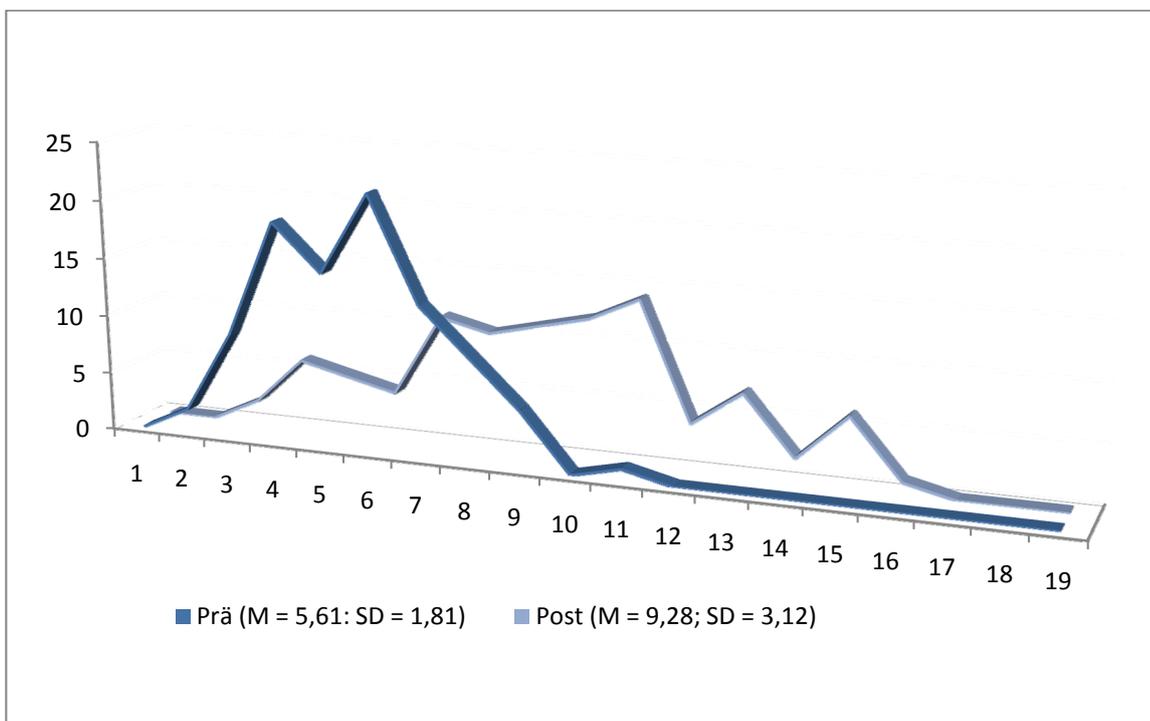


Abbildung 13. Lokalisierung der Patienten auf der Fokusskala durch die Therapeuten zu Beginn (Prä) und am Ende (Post) der Therapie.  $N = 95$  Patienten, für die beide Messungen vorliegen. Codierung: 1 = Abwehr des Fokus... 19 = Auflösung des Fokus.

Therapeuten mit ( $N = 95$ ) und ohne ( $N = 121$ ) Abschlussmessung unterscheiden sich dabei hinsichtlich ihrer Prä-Werte auf der Fokusskala: Therapeuten ohne Abschlussmessung haben einen etwas besseren Anfangswert ( $M = 6,30$ ) dokumentiert als Therapeuten mit Abschlussmessung ( $M = 5,61$ ), der Unterschied um 0,71 Skaleneinheiten fällt allerdings nicht sehr groß aus.

### 5.2.4 Validierung der Fokusskala anhand der Patientenmessungen

Setzt man die Therapeutenangaben auf der Fokusskala mit den Patientenangaben zum jeweiligen Messzeitpunkt in Beziehung, so ergibt sich folgender Befund (Tabelle 84): Zum Zeitpunkt des Therapiebeginns zeigen sich noch keine signifikanten Zusammenhänge zwischen der Fokusskala und der Symptombelastung auf den psychometrischen Skalen. Auch die beiden Skalen des Helping Alliance Questionnaire weisen zu diesen frühen Zeitpunkt noch keine statistisch signifikante Korrelation mit der Fokusskala auf.

Zum Zeitpunkt der Therapiebeendigung zeigen sich hingegen zwei statistisch signifikante Korrelationen mit den Skalen des HAQ. Mit dem BSI, IIP-D, BDI und SF-12 ergeben sich hingegen auch bei Therapieende keine signifikanten Korrelationen.

Tabelle 84. Zusammenhänge der Fokusskala mit den psychometrischen Skalen der Patientenmessung

Produkt-Moment-Korrelationen mit der OPD-Fokusskala zum jeweiligen Messzeitpunkt	Messung bei Therapiebeginn PRÄ	Messung bei Therapieende POST
BSI Psychische Belastung	.07	.17
IIP-D Interpersonale Probleme	-.03	.07
BDI Depressivität	.10	.09
SFK Körperliches Befinden	.07	.07
SFP Psychisches Befinden	.03	.13
HAQ-B Beziehungszufriedenheit	.11	.27*
HAQ-E Erfolgswzufriedenheit	.04	.31**

Anmerkungen. Alle Korrelationskoeffizienten sind so gepolt, dass eine gutes Befinden auf der jeweiligen Skala mit einem höheren Wert auf der Fokusskala gleichsinnig assoziiert ist. Stichprobengrößen Prämessung: N = 139 (BDI) bzw. 200 (alle anderen Skalen) und Postmessung: N = 53 (BDI) bzw. 83 (alle anderen Skalen). Korrelation statistisch signifikant: \* $p < .05$  und \*\* $p < .01$ .

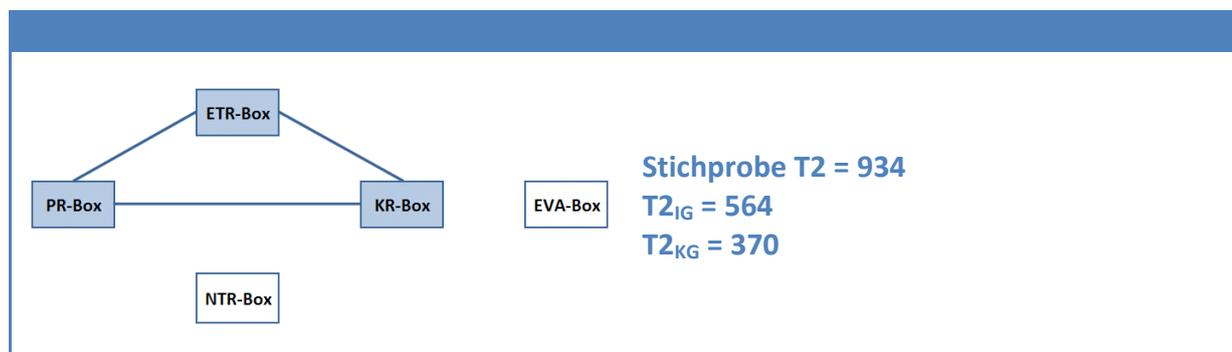
Dieser Befund lässt sich wie folgt interpretieren: Die OPD ist ein Instrument zur Erfassung der therapeutischen Prozessqualität aus Sicht der Therapeuten, was durch die Zusammenhänge mit den HAQ-Skalen illustriert wird. Die OPD steht hingegen weniger im Zusammenhang mit der aktuell empfundenen Symptombelastung durch die Patienten. Es ist aber zu erwarten, dass sich die durch die OPD hier gemessenen eindrucksvollen strukturellen Veränderungen bzw. Zuwächse an Selbsteinsicht bei den Patienten längerfristig auch in einer reduzierten Symptombelastung niederschlagen werden, was durch die bei den Patienten erhobenen katamnestischen Daten ja auch bestätigt wird.

### 5.2.5 Zusammenfassende Bewertung

Mit der OPD fand für eine Teilstichprobe ein für die therapeutische Prozessevaluation sehr vielversprechendes und inhaltlich interessantes Instrument im Rahmen der Studie Anwendung. Leider liegen nur vergleichsweise wenige Datensätze zur OPD vor, weshalb von weiter differenzierenden Auswertungen für bestimmte Subgruppen Abstand genommen werden musste.

Auf der hier explorativ zur statistischen Datenauswertung berechneten Fokusskala zeigten sich sehr hohe Effektgrößen. Damit leistet die OPD zur Evaluation der Prozessqualität einen bedeutsamen Beitrag, wie durch entsprechende Zusammenhangsanalysen mit den beiden Skalen zur therapeutischen Beziehungs- und Erfolgsqualität im Helping Alliance Questionnaire gezeigt werden konnte.

## 5.3 Therapeutenratings



### 5.3.1 Drop-Out-Analyse für die Therapeutenangaben

Angesichts der hohen Ausfallquote von 42,7% fehlenden Therapieabschlussdokumentation (die Rücklaufquote beträgt  $T2 / T1 = 934 / 1.629 = 57,3\%$ ) wird eine Analyse zur Beantwortung folgender Fragen durchgeführt:

1. Unterscheidet sich die Drop-Out-Rate zwischen IG und KG generell?
2. Gibt es systematische Selektionseffekte im Drop-Out zwischen IG und KG in Abhängigkeit von bestimmten Ausgangsmerkmalen, welche die Ergebnisinterpretation verzerren könnten?

Tabelle 85. Kreuztabelle zur Drop-Out-Rate bei den Therapieabschlussdokumentationen der Therapeuten

	IG	KG	Gesamt
TBE nicht vorhanden	404 41,7%	291 44,0%	695 42,7%
TBE vorhanden	564 58,3%	370 56,0%	934 57,3%
Gesamt	968 100,0%	661 100,0%	1629 100,0%

Anmerkung. T1 = 1.629.  $\chi^2 = 0,750$ ; df = 1; p = .386.

Aus Tabelle 85 geht hervor, dass in der IG die Rücklaufquote der Therapeuten-Abschlussdokumentationen 58,3% beträgt und in der KG 56,0%. Dieser Unterschied ist statistisch nicht signifikant. Weiter wird untersucht, ob systematische Zusammenhänge zwischen der Drop-Out-Rate und anderen Patientenmerkmalen bestehen. Tabelle 86 zeigt die Ergebnisse. Patienten, bei denen keine Abschlussdokumentation der Therapeuten vorliegt, haben eine höhere Ausgangsbelastung auf dem multiplen Statusindikator STA\_TH ( $d = 0,20$ ) und sind etwas jünger ( $d = 0,15$ ) als Patienten, bei denen das Therapieende von den Therapeuten dokumentiert wurde. Weitere statistisch signifikante Unterschiede bestehen nicht.

Tabelle 86. Drop-Out-Analyse für die Therapeutenangaben.

Abschlussmessung TBE liegt vor?	Ja N = 934	Nein N = 634	d	Sig. p
STA_TH Messung bei Therapiebeginn <sup>a</sup>	56,28	53,63	0,20	<.001
Alter in Jahren bei Therapiebeginn	41,35	39,61	0,15	.002
Geschlecht weiblich (0=nein, 1=ja)	0,669	0,708	-0,08	.094
Schulabschluss (0=bis HS, 1=RS, 2=(Fach-)Abitur)	1,392	1,439	-0,07	.197
Patient mindestens 15h / Woche erwerbstätig	0,591	0,575	0,03	.538
Wohnsituation allein (0=nein, 1=ja)	0,243	0,237	0,01	.781
Therapiemotivation (0=gering... 3=sehr hoch)	2,12	2,124	-0,01	.900
Anzahl der ICD-10-Diagnosen	1,626	1,581	0,05	.310
Zurückliegen Erstmanifestation Krankheit in Jahren	8,025	7,206	0,09	.086

Anmerkung. T1 = 1.629. <sup>a</sup> Die Konstruktion der Skala STA\_TH ist in Abschnitt 5.3.2 beschrieben

### 5.3.2 Status- und Ergebnisindikatoren der Therapeutenangaben

Bei den Therapeutenangaben wurden folgende Merkmale sowohl zu Behandlungsbeginn (Prä) als auch bei Behandlungsende (Post) erhoben und eignen sich daher zur Bildung eines multiplen Ergebniskriteriums (vgl. Abschnitt 3.3.2), das auf Prä-Post-Differenzwerten beruht:

- Aktuelle Problemausprägung in problemrelevanten Lebensbereichen (sieben Items, Wertebereich 0-10). Eine erfolgreiche Veränderung in diesem Merkmalsbereich steht für den Erwerb neuer Bewältigungsstrategien im realen Alltag, die auch für die Zeit nach der Therapie eine höhere soziale Kompetenz, Eigenverantwortlichkeit und Belastbarkeit im täglichen privaten und beruflichen Leben erwarten lassen.
- Aktueller Beeinträchtigungsschwere-Score BSS (drei Items, Wertebereich 0-4). Eine Verbesserung in diesem Merkmalsbereich steht für einen Rückgang der körperlichen, psychischen und sozialen Beeinträchtigung des Patienten und damit einen Rückgang der Symptomatik.
- Aktuelles Funktionsniveau GAF (ein Item). Wertebereich: 1-100, Eine erfolgreiche Veränderung des GAF steht für eine Verbesserung des allgemeinen Funktionsniveaus des Patienten.
- Aktuelle Arbeitsfähigkeit (ein Item). Wertebereich: 0-1. Eine Verbesserung der Arbeitsfähigkeit schützt den Patienten vor dem Verlust seines Arbeitsplatzes bzw. ermöglicht ihm die Wiederaufnahme einer Erwerbstätigkeit.
- Aktuelle Medikation (ein Item). Wertebereich: 0-1. Eine Verringerung der Medikation mit Psychopharmaka ist zum einen für die TK als Kostenträger relevant, steht aber auch für eine Verbesserung des psychischen Allgemeinbefindens, die es dem Patienten ermöglicht, ein Leben ohne die Abhängigkeit von Medikamenten mit ihren Gefahren von Abhängigkeit oder beeinträchtigenden Nebenwirkungen zu führen.

Die 13 Einzelaspekte lassen sich als singuläre Statusindikatoren auffassen und zu einem multiplen Statusindikator aggregieren. Hierzu erfolgt zunächst eine Lineartransformation aller Einzelaspekte in eine einheitliche Metrik. Der aus den 13 Einzelaspekten neu gebildete multiple Statusindikator STA\_TH steht für das Gesamtbefinden des Patienten aus Sicht des Therapeuten bei Therapieanfang bzw. Therapieende einbezogen werden.

Für die Therapeutenratings wird folgende Transformationsvorschrift verwendet, wobei x für den Originalwert und y für den neuen, lineartransformierten Wert steht:

- 7 Items zum Problemrating:  $y = 100 - 10 * x$
- 3 Items zum BSS:  $y = 100 - 25 * x$
- 1 Item zum GAF:  $y = x$  (der GAF hat bereits einen Range von 0-100 Punkten)
- 1 Item zur Arbeitsfähigkeit:  $y = 100 * x$
- 1 Item zur Medikation:  $y = 100 - 100 * x$

Es ist uns bewusst, dass mit der Arbeitsfähigkeit und mit der Medikation auch zwei dichotome Items in die neu gebildete Skala einfließen. Das Allgemeine Lineare Modell ist gegenüber solchen kleineren Voraussetzungsverletzungen allerdings robust (Cohen et al., 2003). Arbeitsfähigkeit und Medikation bilden die Interessen wichtiger Stakeholdergruppen (z.B. Arbeitgeber und Krankenversicherung) ab, weshalb diese beiden Items hier berücksichtigt und in die Skala mit aufgenommen werden.

### 5.3.3 Vergleich zwischen Interventions- und Kontrollgruppe

In Tabelle 87 und Tabelle 88 sind zunächst die Rohwerte der Items für die Prä- und Post-Messung wiedergegeben. Hierbei wurden nur vollständige Messungen der Stichprobe T2 berücksichtigt (d.h. sowohl die Prä- als auch die Post-Messung für das betreffende Item liegt vor).

Tabelle 89 und Tabelle 90 enthalten die auf einen Wertebereich von 0 – 100 lineartransformierten Werte der Einzelaspekte. Vergleicht man diese mit den Rohwerten, so wird deutlich, dass die Lineartransformation keinen Einfluss auf die Effektgrößen und inferenzstatistischen Tests hat, dafür jetzt aber alle Einzelaspekte direkt miteinander vergleichbar sind.

Besonders deutliche Beeinträchtigungen aus Therapeutesicht bestehen zu Therapiebeginn beim psychischen Befinden sowie im familiären und partnerschaftlichen Bereich, gefolgt von beruflichen Problemen. Der neu gebildete multiple Statusindikator STA\_TH lässt sich als Globalindikator für das Allgemeinbefinden der Patienten aus Therapeutesicht interpretieren. Weder bei Therapiebeginn, noch bei Therapieende finden sich auf STA\_TH statistisch bedeutsame Unterschiede zwischen IG und KG, beide Gruppen verbessern sich im Prä-Post-Vergleich im Durchschnitt sehr deutlich.

Einzelne statistisch signifikante Mittelwertsunterschiede zwischen den beiden Gruppen sollten auf dem Hintergrund einer möglichen Alpha-Fehler-Inflation daher nicht überinterpretiert werden.

Tabelle 87. Therapeutenratings bei Therapiebeginn (Prämessung, Rohwerte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Problemrating Beruf, Ausbildung	563	4,72	3,39	359	4,62	3,29	-0,03	.665
Problemrating Partnerschaft	565	5,90	3,09	356	6,13	3,07	0,07	.276
Problemrating Familie	564	5,57	2,85	358	5,73	2,79	0,06	.419
Problemrating Freizeit	562	3,81	2,66	346	4,01	2,61	0,08	.270
Problemrating Sozial (Wohnen, Finanzen)	562	3,31	2,82	360	3,03	2,88	-0,10	.145
Problemrating Selbstverantwortung	568	4,14	2,77	359	4,09	2,67	-0,02	.764
Problemrating Gesundheitsverhalten	555	3,77	2,63	345	3,72	2,72	-0,02	.797
BSS letzte 7 Tage körperlich	551	1,60	1,03	355	1,53	1,02	-0,06	.354
BSS letzte 7 Tage psychisch	548	2,73	0,76	353	2,63	0,73	-0,13	.050
BSS letzte 7 Tage sozialkommunikativ	548	1,91	1,02	353	1,76	1,04	-0,16	.025
GAF Funktionsniveau aktuell	504	56,2	10,0	332	56,8	10,9	-0,06	.405
Arbeitsfähigkeit	552	0,76	0,43	360	0,78	0,41	-0,06	.382
Medikation	529	0,37	0,48	352	0,35	0,48	-0,04	.621

Anmerkungen. T2 = 934. Codierung: Problemratings 0 (problemlos) – 10 (sehr problematisch); Beeinträchtigung BSS: 0 (gar nicht) – 4 (extrem); Funktionsniveau GAF: 1 (sehr schlecht) – 100 (sehr gut); Arbeitsfähigkeit/Medikation: 0 (nein) – 1 (ja).

Tabelle 88. Therapeutenratings bei Therapieende (Postmessung, Rohwerte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Problemrating Beruf, Ausbildung	563	2,75	2,70	359	3,05	2,79	0,11	.097
Problemrating Partnerschaft	565	3,47	2,80	356	4,35	3,09	0,30	<.001
Problemrating Familie	564	3,24	2,51	358	3,78	2,77	0,21	.003
Problemrating Freizeit	562	2,07	2,05	346	2,64	2,31	0,26	<.001
Problemrating Sozial (Wohnen, Finanzen)	562	2,12	2,35	360	2,14	2,45	0,09	.925
Problemrating Selbstverantwortung	568	2,37	2,33	359	2,77	2,52	0,17	.015
Problemrating Gesundheitsverhalten	555	2,25	2,22	345	2,49	2,34	0,11	.115
BSS letzte 7 Tage körperlich	551	0,89	0,94	355	0,74	0,85	-0,17	.013
BSS letzte 7 Tage psychisch	548	1,18	0,97	353	0,95	0,79	-0,26	<.001
BSS letzte 7 Tage sozialkommunikativ	548	0,85	0,88	353	0,68	0,79	-0,21	.002
GAF Funktionsniveau aktuell	504	73,5	14,1	332	76,1	13,8	-0,19	.010
Arbeitsfähigkeit	552	0,78	0,41	360	0,83	0,38	-0,11	.098
Medikation	529	0,20	0,40	352	0,20	0,40	-0,01	.902

Anmerkungen. T2 = 934. Codierung: Problemratings 0 (problemlos) – 10 (sehr problematisch); Beeinträchtigung BSS: 0 (gar nicht) – 4 (extrem); Funktionsniveau GAF: 1 (sehr schlecht) – 100 (sehr gut); Arbeitsfähigkeit/Medikation: 0 (nein) – 1 (ja).

Tabelle 89. Therapeutenratings bei Therapiebeginn (Prämessung, Lineartransformierte Werte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Problemrating Beruf, Ausbildung	563	52,8	33,9	359	53,8	32,9	-0,03	.665
Problemrating Partnerschaft	565	41,0	30,9	356	38,7	30,7	0,07	.276
Problemrating Familie	564	44,3	28,5	358	42,7	27,9	0,06	.419
Problemrating Freizeit	562	61,9	26,6	346	59,9	26,1	0,08	.270
Problemrating Sozial (Wohnen, Finanzen)	562	66,9	28,1	360	69,7	28,8	-0,10	.145
Problemrating Selbstverantwortung	568	58,6	27,7	359	59,1	26,7	-0,02	.764
Problemrating Gesundheitsverhalten	555	62,3	26,3	345	62,8	27,2	-0,02	.797
BSS letzte 7 Tage körperlich	551	60,1	25,7	355	61,7	25,5	-0,06	.354
BSS letzte 7 Tage psychisch	548	31,7	19,0	353	34,2	18,2	-0,13	.050
BSS letzte 7 Tage sozialkommunikativ	548	52,1	25,5	353	56,1	25,9	-0,16	.025
GAF Funktionsniveau aktuell	504	56,2	10,0	332	56,8	10,9	-0,06	.405
Arbeitsfähigkeit	552	75,5	43,0	360	78,1	41,4	-0,06	.382
Medikation	529	63,1	48,3	352	64,8	47,8	-0,04	.621
<b>Multipler Statusindikator STA_TH</b>	<b>539</b>	<b>55,8</b>	<b>13,5</b>	<b>352</b>	<b>56,8</b>	<b>12,7</b>	<b>-0,08</b>	<b>.267</b>

Anmerkung. T2 = 934. Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden). Die Skala STA\_TH wurde berechnet, indem für jeden Patienten die Summe der beantworteten Items durch die Anzahl der beantworteten Items geteilt wurde, sofern mindestens 11 Items gültig beantwortet wurden.

Tabelle 90. Therapeutenratings bei Therapieende (Postmessung, Lineartransformierte Werte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Problemrating Beruf, Ausbildung	563	72,5	27,0	359	69,5	27,9	0,11	.097
Problemrating Partnerschaft	565	65,3	28,0	356	56,5	30,9	0,30	<.001
Problemrating Familie	564	67,6	25,1	358	62,2	27,7	0,21	.003
Problemrating Freizeit	562	79,3	20,5	346	73,6	23,1	0,26	<.001
Problemrating Sozial (Wohnen, Finanzen)	562	78,8	23,5	360	76,6	24,5	0,09	.925
Problemrating Selbstverantwortung	568	76,3	23,3	359	72,3	25,2	0,17	.015
Problemrating Gesundheitsverhalten	555	77,5	22,2	345	75,1	23,4	0,11	.115
BSS letzte 7 Tage körperlich	551	77,8	23,4	355	81,6	21,1	-0,17	.013
BSS letzte 7 Tage psychisch	548	70,5	24,1	353	76,3	19,8	-0,26	<.001
BSS letzte 7 Tage sozialkommunikativ	548	78,6	22,1	353	83,1	19,6	-0,21	.002
GAF Funktionsniveau aktuell	504	73,5	14,1	332	76,1	13,8	-0,19	.010
Arbeitsfähigkeit	552	78,1	41,4	360	82,5	38,0	-0,11	.098
Medikation	529	79,8	40,2	352	80,1	40,0	-0,01	.902
<b>Multipler Statusindikator STA_TH</b>	<b>539</b>	<b>75,5</b>	<b>15,1</b>	<b>352</b>	<b>75,0</b>	<b>14,1</b>	<b>0,03</b>	<b>.631</b>

Anmerkung. T2 = 934. Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden). Die Skala STA\_TH wurde berechnet, indem für jeden Patienten die Summe der beantworteten Items durch die Anzahl der beantworteten Items geteilt wurde, sofern mindestens 11 Items gültig beantwortet wurden.

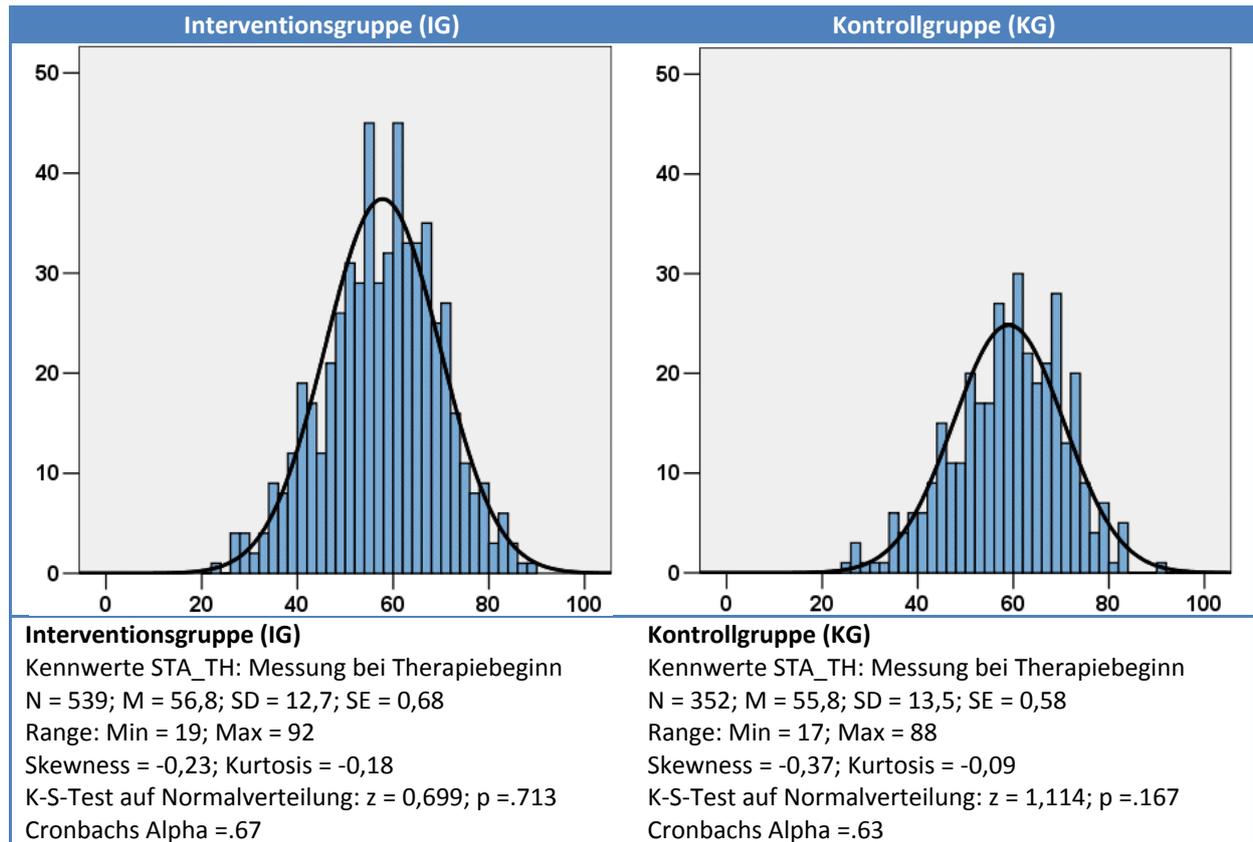


Abbildung 14. T2 = 934. Verteilungskennwerte des multiplen Statusindikators STA\_TH bei Therapiebeginn

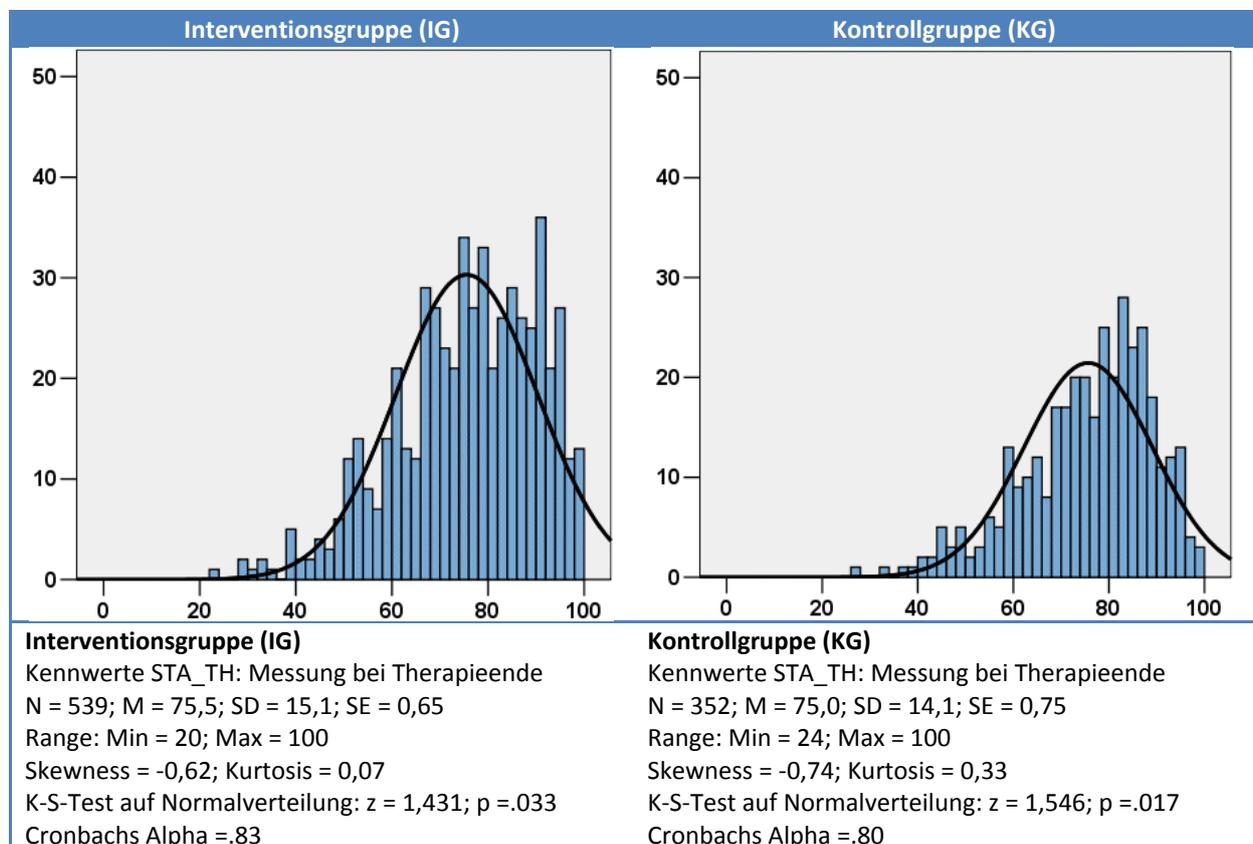


Abbildung 15. T2 = 934. Verteilungskennwerte des multiplen Statusindikators STA\_TH bei Therapieende

Neben dem Status bei Aufnahme und Entlassung lassen sich für alle 13 lineartransformierten Items auch entsprechende Prä-Post-Differenzwerte berechnen. Aus diesen 13 singulären Ergebniskriterien wurde das multiple Ergebniskriterium MEK\_TH berechnet. Tabelle 91 zeigt die Ergebnisse für die 13 Items und die Skala MEK\_TH für die IG und KG im Vergleich.

Zwischen IG und KG zeigen sich dabei keine statistisch signifikanten Unterschiede auf dem multiplen Ergebniskriterium MEK\_TH, d.h. die Patienten der IG verbessern sich im Verlauf der Therapie (M = +19,7 Punkte) ebenso stark wie die Patienten der KG (M = +18,2 Punkte).

**Tabelle 91. Therapeutenratings lineartransformierte PRÄ-POST-Differenzwerte**

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
Problemrating Beruf, Ausbildung	563	19,7	32,0	359	15,7	31,1	0,12	.058
Problemrating Partnerschaft	565	24,3	29,5	356	17,8	30,3	0,21	.001
Problemrating Familie	564	23,3	27,2	358	19,5	30,1	0,13	.047
Problemrating Freizeit	562	17,4	26,1	346	13,7	27,6	0,14	.040
Problemrating Sozial (Wohnen, Finanzen)	562	11,9	24,1	360	6,9	25,4	0,18	.076
Problemrating Selbstverantwortung	568	17,7	26,7	359	13,2	28,5	0,16	.014
Problemrating Gesundheitsverhalten	555	15,2	26,6	345	12,3	26,9	0,11	.112
BSS letzte 7 Tage körperlich	551	17,7	26,9	355	19,9	28,4	-0,09	.242
BSS letzte 7 Tage psychisch	548	38,8	26,5	353	42,1	25,9	-0,18	.067
BSS letzte 7 Tage sozialkommunikativ	548	26,5	28,7	353	27,0	27,7	-0,02	.805
GAF Funktionsniveau aktuell	504	17,3	15,4	332	19,3	16,6	-0,19	.087
Arbeitsfähigkeit	552	2,6	34,5	360	4,4	34,7	-0,04	.416
Medikation	529	16,7	45,9	352	15,3	46,4	0,03	.683
<b>Multiples Ergebniskriterium MEK_TH</b>	<b>539</b>	<b>19,7</b>	<b>15,1</b>	<b>352</b>	<b>18,2</b>	<b>13,9</b>	<b>0,10</b>	<b>.138</b>

Anmerkung. T2 = 934. Codierung: -100 (maximal mögliche Verschlechterung) – 100 (maximal mögliche Verbesserung). Die Skala MEK\_TH wurde berechnet, indem für jeden Patienten die Summe der beantworteten Items durch die Anzahl der beantworteten Items geteilt wurde, sofern mindestens 10 Items gültig beantwortet wurden.

Abbildung 14 bis Abbildung 17 geben die Verteilungsmerkmale der Skalen für die IG und KG mitsamt Scatter-Plot zwischen Prä- und Post-Messung wieder. Insgesamt machen die Verteilungen einen plausiblen Eindruck, so ist sowohl der multiple Statusindikator STA\_TH in den beiden Gruppen zum A- und E-Zeitpunkt als auch der aus den Prä-Post-Differenzwerten gebildete multiple Ergebnisindikator MEK\_TH bis auf kleinere Abweichungen normalverteilt. Neben der überwiegenden Mehrzahl von Patienten, bei denen eine deutliche Verbesserung des Befindens festgestellt wurde (Punkte oberhalb der Diagonalen in den Scatterplots), gibt es auch einige Patienten, bei denen sich das Befinden nicht verbessert bzw. sogar verschlechtert hat (Punkte auf bzw. unterhalb der Diagonalen).

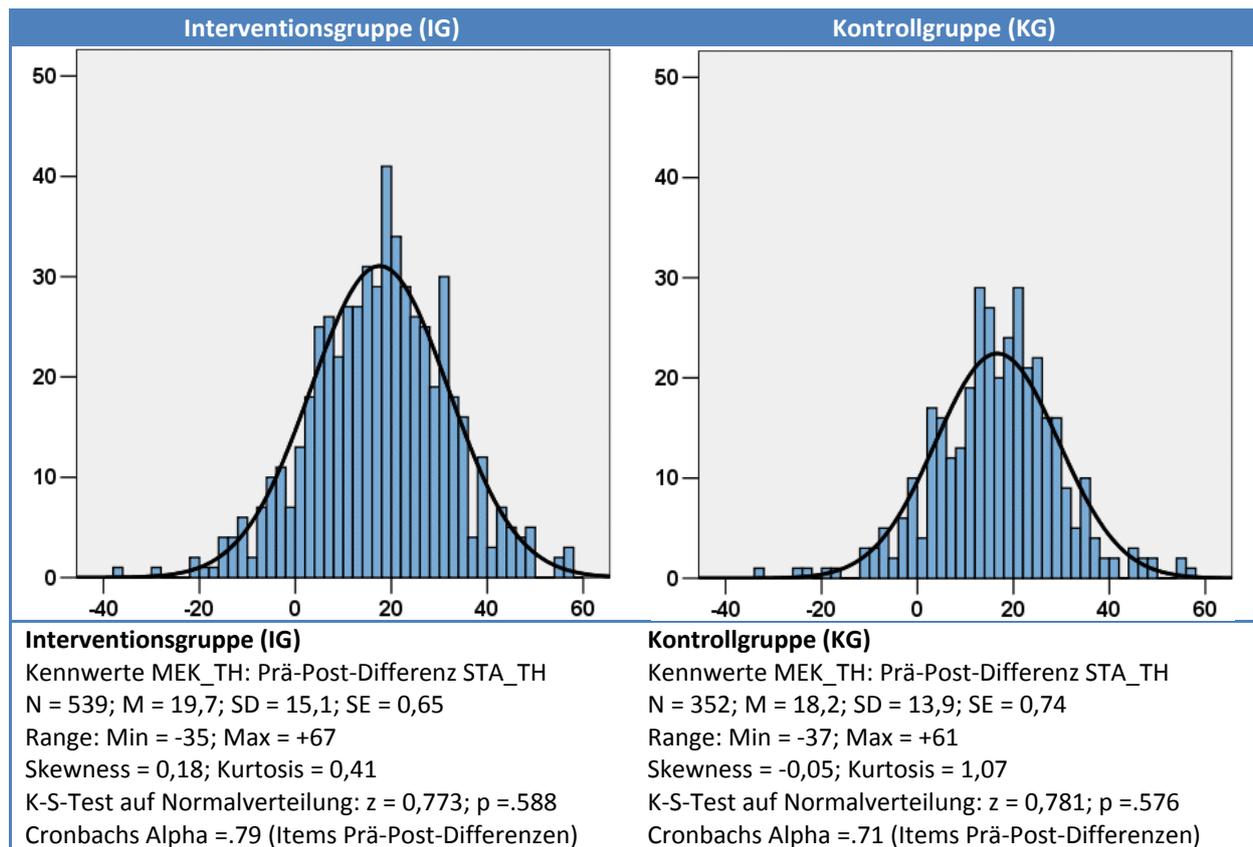


Abbildung 16. T2 = 934. Verteilungskennwerte des multiplen Ergebnisindikators MEK\_TH (Prä-Post-Differenz STA\_TH)

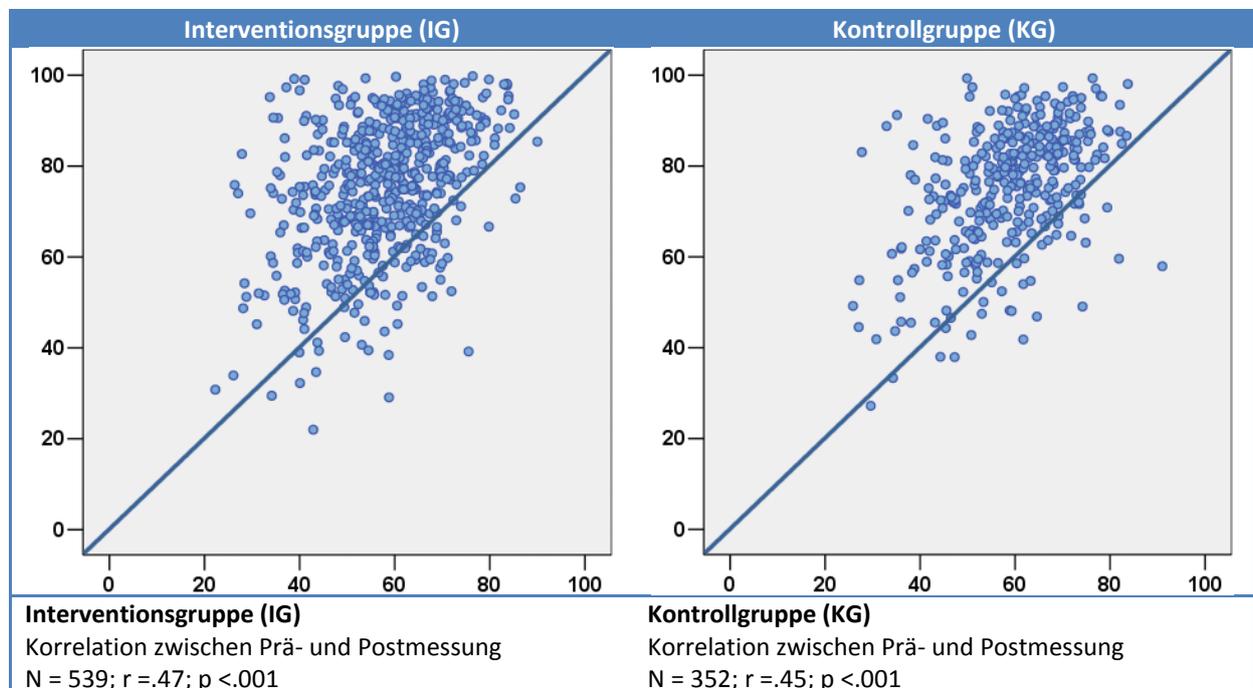


Abbildung 17. T2 = 934. Scatterplot des multiplen Statusindikators STA\_TH (A/E-Vergleich). x-Achse: A-Messung bei Therapiebeginn. y-Achse: E-Messung bei Therapieende.

### 5.3.4 Vergleich zwischen Therapiebeginn und Therapieende

Auch die Prä-Post-Effektgrößen wurden für die IG und KG berechnet und sollen nachfolgend deskriptiv dargestellt werden (Tabelle 92 und Tabelle 93).

Tabelle 92. Therapeutenratings für die IG (PRÄ-POST-Vergleich, Lineartransformierte Werte)

Interventionsgruppe (IG)	Prämessung			Postmessung			Vergleich	
	N	M	SD	N	M	SD	d	p
Problemrating Beruf, Ausbildung	563	52,8	33,9	563	72,5	27,0	0,59	<.001
Problemrating Partnerschaft	565	41,0	30,9	565	65,3	28,0	0,79	<.001
Problemrating Familie	564	44,3	28,5	564	67,6	25,1	0,82	<.001
Problemrating Freizeit	562	61,9	26,6	562	79,3	20,5	0,66	<.001
Problemrating Sozial (Wohnen, Finanzen)	562	66,9	28,1	562	78,8	23,5	0,42	<.001
Problemrating Selbstverantwortung	568	58,6	27,7	568	76,3	23,3	0,65	<.001
Problemrating Gesundheitsverhalten	555	62,3	26,3	555	77,5	22,2	0,57	<.001
BSS letzte 7 Tage körperlich	551	60,1	25,7	551	77,8	23,4	0,69	<.001
BSS letzte 7 Tage psychisch	548	31,7	19,0	548	70,5	24,1	2,08	<.001
BSS letzte 7 Tage sozialkommunikativ	548	52,1	25,5	548	78,6	22,1	1,03	<.001
GAF Funktionsniveau aktuell	504	56,2	10,0	504	73,5	14,1	1,67	<.001
Arbeitsfähigkeit	552	75,5	43,0	552	78,1	41,4	0,06	.016
Medikation	529	63,1	48,3	529	79,8	40,2	0,35	<.001
<b>Multipler Statusindikator STA_TH</b>	<b>539</b>	<b>55,8</b>	<b>13,5</b>	<b>539</b>	<b>75,5</b>	<b>15,1</b>	<b>1,49</b>	<b>&lt;.001</b>

Anmerkung. T2 = 934. Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden). Die Skala STA\_TH wurde berechnet, indem für jeden Patienten die Summe der beantworteten Items durch die Anzahl der beantworteten Items geteilt wurde, sofern mindestens 10 Items gültig beantwortet wurden. Zur Berechnung der Effektgröße d wurde aus Gründen der Vergleichbarkeit hier die gepoolte Standardabweichung der Prä-Messung der IG und KG verwendet.

Tabelle 93. Therapeutenratings für die KG (PRÄ-POST-Vergleich, Lineartransformierte Werte)

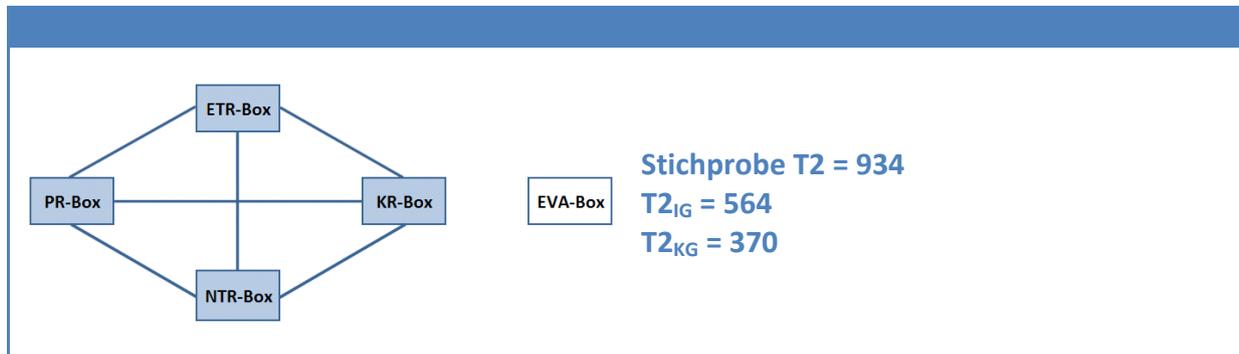
Kontrollgruppe (KG)	Prämessung			Postmessung			Vergleich	
	N	M	SD	N	M	SD	d	p
Problemrating Beruf, Ausbildung	359	53,8	32,9	359	69,5	27,9	0,47	<.001
Problemrating Partnerschaft	356	38,7	30,7	356	56,5	30,9	0,58	<.001
Problemrating Familie	358	42,7	27,9	358	62,2	27,7	0,69	<.001
Problemrating Freizeit	346	59,9	26,1	346	73,6	23,1	0,52	<.001
Problemrating Sozial (Wohnen, Finanzen)	360	69,7	28,8	360	76,6	24,5	0,24	<.001
Problemrating Selbstverantwortung	359	59,1	26,7	359	72,3	25,2	0,48	<.001
Problemrating Gesundheitsverhalten	345	62,8	27,2	345	75,1	23,4	0,46	<.001
BSS letzte 7 Tage körperlich	355	61,7	25,5	355	81,6	21,1	0,78	<.001
BSS letzte 7 Tage psychisch	353	34,2	18,2	353	76,3	19,8	2,25	<.001
BSS letzte 7 Tage sozialkommunikativ	353	56,1	25,9	353	83,1	19,6	1,05	<.001
GAF Funktionsniveau aktuell	332	56,8	10,9	332	76,1	13,8	1,86	<.001
Arbeitsfähigkeit	360	78,1	41,4	360	82,5	38,0	0,10	.085
Medikation	352	64,8	47,8	352	80,1	40,0	0,32	<.001
<b>Multipler Statusindikator STA_TH</b>	<b>352</b>	<b>56,8</b>	<b>12,7</b>	<b>352</b>	<b>75,0</b>	<b>14,1</b>	<b>1,38</b>	<b>&lt;.001</b>

Anmerkung. T2 = 934. Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden). Die Skala STA\_TH wurde berechnet, indem für jeden Patienten die Summe der beantworteten Items durch die Anzahl der beantworteten Items geteilt wurde, sofern mindestens 10 Items gültig beantwortet wurden. Zur Berechnung der Effektgröße d wurde aus Gründen der Vergleichbarkeit hier die gepoolte Standardabweichung der Prä-Messung der IG und KG verwendet.

Auf der Gesamtskala STA\_TH zeigt sich eine Verbesserung um 19,7 Punkte bei der IG (d = 1,49) bzw. 18,2 Punkte (d = 1,38) bei der KG. Aus Sicht der Therapeuten hat sich das Allgemeinbefinden der Patienten

also deutlich verbessert, nach Cohen entsprechen Effektgrößen ab  $d = 0,80$  bereits einem großen Effekt. Bei den Einzelvergleichen hat sich die psychische Beeinträchtigung aus Sicht der Therapeuten mit einer Effektgröße von über 2,00 am deutlichsten verbessert. Für alle Diagnosegruppen ergeben sich dabei auf dem multiplen Statusindikator sowohl für die IG als auch für die KG große Prä-Post-Effektgrößen.

### 5.3.5 Statistische Kontrolle möglicherweise vorhandener Confounder



Um sicherzugehen, dass keine Verdeckung von Unterschieden zwischen IG und KG durch konfundierende Störvariablen (Confounder) stattgefunden hat, wurde eine multiple Regression berechnet. Als abhängige Variable wurde dabei die Messung auf dem multiplen Statusindikator STA\_TH bei Therapieende verwendet. Tabelle 94 gibt die Ergebnisse wieder. Die Regressionsgleichung ist insgesamt statistisch signifikant ( $N = 891$ ;  $p < .001$ ;  $R = .491$ ;  $R^2_{adj} = .232$ ), d.h. mindestens einer der einbezogenen Prädiktoren steht in Beziehung zum multiplen Statusindikator STA\_TH bei Therapieende.

Tabelle 94 gibt die Ergebnisse für alle Prädiktoren wieder. Zunächst ist festzustellen, dass die ausgewählten Prädiktoren hohe Toleranzwerte und somit kaum Multikollinearität aufweisen, d.h. das Regressionsmodell zur Abschätzung möglicher Konfundierungseffekte ist stabil.

Am deutlichsten ist das ebenfalls als Prädiktor einbezogene Niveau der Ausgangsmessung auf der Skala STA\_TH mit dem erreichten Niveau bei Therapieende assoziiert ( $Beta = .41$ ). Dieser Befund ist vergleichsweise trivial, so steht das Befinden bei Therapiebeginn erfahrungsgemäß fast immer in deutlicher Beziehung mit dem Befinden bei Therapieende. Die Unvollständigkeit des Zusammenhangs impliziert dabei, dass stark beeinträchtigte Patienten bereits rein rechnerisch eine größere Chance haben, große Prä-Post-Veränderungen zu erzielen als moderat beeinträchtigte Patienten. Dieses Phänomen ist auch als Regression zur Mitte bekannt und wird hier durch die Einbeziehung der Prä-Messung statistisch kontrolliert. Weiter wird durch die statistische Kontrolle der Prä-Messung erreicht, dass sich die übrigen Prädiktoren im Hinblick auf ihren Einfluss auf die tatsächlich im Therapieverlauf erzielten Prä-Post-Differenzen interpretieren lassen.

Ein negativer Prädiktor für das Therapieergebnis ist, wenn der Patient alleine lebt. Das Vorliegen dieses Merkmals lässt im Durchschnitt ein um  $B = 4,513$  Punkte schlechteres Ergebnis auf dem multiplen Statusindikator bei Entlassung erwarten. Hier wird die Rolle der sozialen Unterstützung für den Therapieerfolg bzw. des Einübens neuer Verhaltensmuster mit Interaktionspartnern im privaten Bereich bereits während des Therapieverlaufes deutlich (auch wenn der Partialzusammenhang mit  $Beta = -.13$  klein ist). Möglicherweise leben aber auch Patienten mit einem schlechteren Allgemeinbefinden und mehr Problemen von vorneherein mit einer größeren Wahrscheinlichkeit alleine.

Weiter stellt eine lange Krankheitsgeschichte (Zurückliegen des Erstauftretens der Erkrankung) einen schwach negativen, ein höheres Bildungsniveau hingegen einen schwach positiven Prädiktor für das Therapieergebnis dar. Darüber hinaus schätzen Therapeuten, die mehr Patienten im Modellvorhaben behandelt haben, die Behandlungsergebnisse etwas optimistischer ein. Alters- und Geschlechtseffekte bestehen nicht. Eben so wenig ist die Anzahl der Diagnosen von Belang für den Status bei Therapieende.

Auch die vom Therapeuten bei Behandlungsbeginn beim Patienten wahrgenommene Therapiemotivation hat hier keinen Einfluss auf das Behandlungsergebnis.

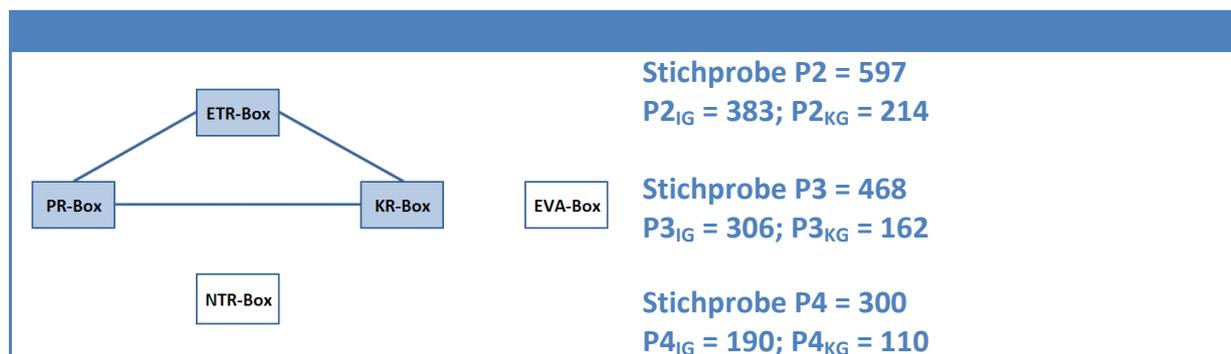
Kommen wir nun abschließend zur wichtigsten Fragestellung: Zeigen sich nach statistischer Kontrolle all dieser Prädiktorvariablen Unterschiede zwischen Interventions- und Kontrollgruppe auf dem multiplen Statusindikator STA\_TH? Bei einem Beta-Gewicht von .05 ( $p = .114$ ) kann diese Frage verneint werden, d.h. das Therapieergebnis der IG- und KG-Patienten aus Sicht der Therapeuten unterscheidet sich auch bei Berücksichtigung der Confounder nicht statistisch signifikant.

**Tabelle 94. Regressionsgleichung zur statistischen Kontrolle möglicher Confoundervariablen bei Bestimmung möglicherweise vorhandener Unterschiede zwischen IG und KG auf dem multiplen Statusindikator STA\_TH bei Therapieende.**

Kriterium: STA_TH Messung bei Therapieende	B	SE	Beta	t	Sig. p	Toleranz
<i>Regressionskonstante a</i>	45,599	3,610		12,632	<.000	
STA_TH Messung bei Therapiebeginn	0,461	0,034	.41	13,407	<.000	.91
Alter in Jahren bei Therapiebeginn	0,007	0,042	.01	0,158	.874	.86
Geschlecht weiblich (0=nein, 1=ja)	0,614	0,949	.02	0,647	.518	.95
Schulabschluss (0=bis HS, 1=RS, 2=(Fach-)Abitur)	1,271	0,630	.06	2,017	.044	.91
Patient mindestens 15h / Woche erwerbstätig	1,993	0,936	.07	2,130	.033	.89
Wohnsituation allein (0=nein, 1=ja)	-4,513	1,022	-.13	-4,415	<.000	.97
Therapiemotivation (0=gering... 3=sehr hoch)	0,956	0,730	.04	1,309	.191	.94
Anzahl der ICD-10-Diagnosen	-0,222	0,512	-.01	-0,433	.665	.91
Zurückliegen Erstmanifestation Krankheit in Jahren	-0,159	0,053	-.09	-3,013	.003	.92
Gruppenzugehörigkeit (0=KG, 1=IG)	1,425	0,901	.05	1,580	.114	.96

Anmerkung. T2 = 934.

## 5.4 Patientenratings



Für die Ergebnisevaluation der Studie aus Patientensicht wurden für die IG und KG zu drei Messzeitpunkten Daten auf den psychometrischen Skalen erhoben:

- A: Aufnahme-Messung bei Therapiebeginn (Stichprobe P1)
- E: Entlass-Messung bei Therapieende (Stichprobe P2; relevant für alle A/E-Vergleiche)
- K: Katamnesemessung ein Jahr nach E (Stichprobe P3; relevant für alle A/K-Vergleiche)

Für Verlaufsbetrachtungen über alle drei Messzeitpunkte hinweg ist darüber hinaus die Stichprobe P4 relevant. Diese enthält alle Patienten, für die eine vollständige Dreipunktmessung A, E und K vorliegt. Auch die Einzelbewertungen der Patienten werden zu multiplen Status- und Ergebnisindikatoren integriert, um eine zusammenfassende Evaluation für die IG und KG durchführen zu können.

### 5.4.1 Drop-Out-Analyse für die Patientenangaben

Auch hier werden angesichts der hohen Drop-Out-Raten entsprechende statistische Analysen für die Teilstichproben P2, P3 und P4 durchgeführt (vgl. Abschnitt 5.3.1).

Tabelle 95. Kreuztabelle zur Drop-Out-Rate bei der Stichprobe P2

	IG	KG	Gesamt
P2 nicht vorhanden	628 62,1%	373 63,5%	1001 62,6%
P2 vorhanden	383 37,9%	214 36,5%	597 37,4%
Gesamt	1011 100,0%	587 100,0%	1598 100,0%

Anmerkung.  $P1 = 1.598$ .  $\text{Chi}^2 = 0,265$ ;  $\text{df} = 1$ ;  $p = .607$ .

Tabelle 96. Kreuztabelle zur Drop-Out-Rate bei der Stichprobe P3

	IG	KG	Gesamt
P3 nicht vorhanden	705 69,7%	425 72,4%	1130 70,7%
P3 vorhanden	306 30,3%	162 27,6%	468 29,3%
Gesamt	1011 100,0%	587 100,0%	1598 100,0%

Anmerkung.  $P1 = 1.598$ .  $\text{Chi}^2 = 1,152$ ;  $\text{df} = 1$ ;  $p = .283$ .

Tabelle 97. Kreuztabelle zur Drop-Out-Rate bei der Stichprobe P4

	IG	KG	Gesamt
P4 nicht vorhanden	821 81,2%	477 81,3%	1298 81,2%
P4 vorhanden	190 18,8%	110 18,7%	300 18,8%
Gesamt	1011 100,0%	587 100,0%	1598 100,0%

Anmerkung. P1 = 1.598.  $\chi^2 = 0,001$ ; df = 1; p = .979.

Aus Tabelle 95 bis Tabelle 97 geht hervor, dass sich die Drop-Out-Raten in den drei Teilstichproben P2, P3 und P4 zwischen IG und KG nicht statistisch signifikant unterscheiden. Dabei ist festzuhalten, dass die Rücklaufquoten mit 37,4% (P2), 29,3% (P3) bzw. 18,8% (P4), bezogen auf P1 mit 1.598 Ausgangsmessungen, deutlich hinter den Erwartungen zurückgeblieben sind, weshalb eine sorgfältige statistische Berücksichtigung möglicher Confounder-Variablen beim Vergleich der Ergebnisqualität für die IG und KG notwendig ist. Aus Tabelle 98 bis Tabelle 100 geht hervor, dass vor allem hinsichtlich der Ausgangsbelastung ein Zusammenhang mit der Drop-Out-Rate in allen drei Teilstichproben P2, P3 und P4 besteht. So sind Patienten mit fehlenden Abschluss- bzw. Katamnesemessungen zu Therapiebeginn um zwei bis drei Punkte stärker belastet.

Tabelle 98. Drop-Out-Analyse für die Patientenangaben, Teilstichprobe P2.

Abschlussmessung (P2) liegt vor?	Ja N = 597	Nein N = 1001	d	Sig. p
STA_PA Messung bei Therapiebeginn <sup>a</sup>	56,64	54,26	0,18	.001
Alter in Jahren bei Therapiebeginn	41,37	40,36	0,09	.089
Geschlecht weiblich (0=nein, 1=ja)	67,3%	68,7%	-0,03	.563
Schulabschluss (0=bis HS, 1=RS, 2=(Fach-)Abitur)	1,40	1,41	-0,02	.674
Patient mindestens 15h / Woche erwerbstätig	57,8%	58,9%	-0,02	.699
Wohnsituation allein (0=nein, 1=ja)	22,9%	24,8%	-0,04	.396
Therapiemotivation (0=gering... 3=sehr hoch)	2,09	2,15	-0,09	.075
Anzahl der ICD-10-Diagnosen	1,60	1,62	-0,02	.638
Zurückliegen Erstmanifestation Krankheit in Jahren	8,37	7,35	0,12	.041

Anmerkung. P1 = 1.598. <sup>a</sup> Die Konstruktion der Skala STA\_PA ist in Abschnitt 5.4.2 beschrieben

Tabelle 99. Drop-Out-Analyse für die Patientenangaben, Teilstichprobe P3.

Katamnesemessung (P3) liegt vor?	Ja N = 468	Nein N = 1130	d	Sig. p
STA_PA Messung bei Therapiebeginn <sup>a</sup>	57,14	54,37	0,21	<.001
Alter in Jahren bei Therapiebeginn	41,61	40,38	0,11	.050
Geschlecht weiblich (0=nein, 1=ja)	67,1%	68,6%	-0,03	.538
Schulabschluss (0=bis HS, 1=RS, 2=(Fach-)Abitur)	1,44	1,39	0,06	.303
Patient mindestens 15h / Woche erwerbstätig	58,0%	58,7%	-0,01	.826
Wohnsituation allein (0=nein, 1=ja)	25,1%	23,7%	0,03	.564
Therapiemotivation (0=gering... 3=sehr hoch)	2,12	2,13	0,00	.932
Anzahl der ICD-10-Diagnosen	1,57	1,64	-0,08	.153
Zurückliegen Erstmanifestation Krankheit in Jahren	7,95	7,69	0,03	.614

Anmerkung. P1 = 1.598. <sup>a</sup> Die Konstruktion der Skala STA\_PA ist in Abschnitt 5.4.2 beschrieben

Tabelle 100. Drop-Out-Analyse für die Patientenangaben, Teilstichprobe P4.

Abschluss und Katamnese (P4) liegen vor?	Ja N = 300	Nein N = 1298	d	Sig. p
STA_PA Messung bei Therapiebeginn <sup>a</sup>	57,80	54,56	0,25	<.001
Alter in Jahren bei Therapiebeginn	41,09	40,66	0,04	.550
Geschlecht weiblich (0=nein, 1=ja)	68,7%	68,1%	0,01	.851
Schulabschluss (0=bis HS, 1=RS, 2=(Fach-)Abitur)	1,40	1,41	-0,01	.887
Patient mindestens 15h / Woche erwerbstätig	54,6%	59,4%	-0,10	.131
Wohnsituation allein (0=nein, 1=ja)	22,1%	24,6%	-0,06	.361
Therapiemotivation (0=gering... 3=sehr hoch)	2,08	2,14	-0,08	.194
Anzahl der ICD-10-Diagnosen	1,58	1,63	-0,05	.410
Zurückliegen Erstmanifestation Krankheit in Jahren	8,44	7,59	0,10	.154

Anmerkung. P1 = 1.598. <sup>a</sup> Die Konstruktion der Skala STA\_PA ist in Abschnitt 5.4.2 beschrieben

#### 5.4.2 Status- und Ergebnisindikatoren der Patientenangaben

Analog zur Vorgehensweise in Abschnitt 5.3.2 werden auch für die Patientensicht multiple Status- bzw. Ergebnisindikatoren gebildet, indem die psychometrischen Skalen zunächst durch Lineartransformation in eine einheitliche Metrik mit einem Wertebereich von 0-100 überführt werden.

Die lineartransformierten Skalen werden zu dem multiplen Statusindikator STA\_PA aggregiert. Auch die Differenzwerte der lineartransformierten Items E-A bzw. K-A werden zu entsprechenden multiplen Ergebnisindikatoren (MEK\_PA) für den Vergleich zwischen Aufnahme und Entlassung bzw. Aufnahme und Katamnese aggregiert. Mindestens vier gültige Skalenwerte bzw. Differenzwerte müssen vorliegen.

Neben der allgemeinen psychischen Symptombelastung im BSI und störungsspezifischen Symptombelastung im jeweiligen Instrument (BDI, AKV, EDI, HZI oder SOMS) wird die soziale Kompetenz über den IIP-D sowie die Funktionsfähigkeit im Alltag über den SF-12 erfasst. Folgende Skalen gehen somit in den multiplen Status- bzw. Ergebnisindikator ein:

- Globale psychische Symptombelastung (Gesamtskala BSI, 53 Items).
- Globales Ausmaß an interpersonalen Problemen (Gesamtskala IIP-D, 64 Items).
- Leistungsfähigkeit im Alltag aufgrund körperlichem Befinden (Skala SF12k mit 6 Items).
- Leistungsfähigkeit im Alltag aufgrund psychischem Befinden (Skala SF12p mit 6 Items).
- Jeweilige störungsspezifische Skala für die Diagnosegruppen Depression (BDI), Angst (AKV), Essstörungen (EDI), Zwangsstörungen (HZI) oder aber Somatoforme Störungen (SOMS).

Die 5 Skalen (4 generische Skalen + 1 störungsspezifische Skala) lassen sich als singuläre Statusindikatoren auffassen und zu einem multiplen Statusindikator aggregieren. Hierzu erfolgt zunächst eine Lineartransformation aller Einzelaspekte in eine einheitliche Metrik. Der aus den 5 Einzelaspekten neu gebildete multiple Statusindikator STA\_PA steht für das Gesamtbefinden des Patienten aus Sicht des Therapeuten bei Therapieanfang bzw. Therapieende.

Für die psychometrischen Skalen wird folgende Transformationsvorschrift verwendet, wobei x für den Originalwert und y für den neuen, lineartransformierten Wert steht:

- BSI Gesamtwert:  $y = 100 - 25 * x$
- IIP-D Gesamtwert:  $y = 100 - 25 * x$
- SF12 Körperliches Befinden:  $y = x$  (hat bereits einen Range von 0-100 Punkten)
- SF12 Psychisches Befinden:  $y = x$  (hat bereits einen Range von 0-100 Punkten)
- BDI Gesamtwert:  $y = 100 * (63 - x)/63$
- AKV Gesamtwert:  $y = 20 * (5 - x)$
- EDI Gesamtwert:  $y = 100 * (1 - (x - 91)/455)$
- HZI Gesamtwert:  $y = 100 * (1 - x/72)$
- SOM Gesamtwert:  $y = 100 * (1 - x/208)$

### 5.4.3 Vergleich zwischen Interventions- und Kontrollgruppe (Stichprobe P2)

In Tabelle 101 und Tabelle 102 sind zunächst die Rohwerte der Patientenskalen für die A- und E-Messung der Stichprobe P2 vergleichend für die IG und KG wiedergegeben.

**Tabelle 101. Patientenangaben bei Therapiebeginn (A-Messung, Rohwerte)**

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	375	1,22	0,65	212	1,14	0,64	-0,12	.137
IIP-D Gesamtwert	381	1,55	0,52	213	1,51	0,55	-0,08	.459
SF12 Körperliches Befinden	376	46,80	10,42	211	49,00	10,00	-0,21	.013
SF12 Psychisches Befinden	376	32,05	9,97	211	33,00	10,28	-0,09	.279
BDI Gesamtwert	261	20,80	9,00	167	18,75	8,64	-0,23	.019
AKV Gesamtwert	84	2,18	0,57	25	2,30	0,62	0,21	.381
EDI Gesamtwert	8	340,9	29,5	4	366,5	50,9	0,69	.293
HZK Gesamtwert	9	29,44	10,60	5	22,20	11,12	-0,67	.252
SOM Gesamtwert	17	37,82	19,75	6	36,03	24,57	-0,09	.885

Anmerkung. P2 = 597. Codierung: BSI und IIP-D: 0 (keine Belastung) – 4 (starke Belastung). SF-12-Skalen: 0 (schlechtes Befinden) – 100 (gutes Befinden). BDI: 0 (keine Depressivität) – 63 (starke Depressivität). AKV: 1 (keine Belastung) – 5 (starke Belastung). EDI: 91 (keine Belastung) – 546 (starke Belastung). HZK: 0 (keine Belastung) – 72 (starke Belastung). SOM 0 (keine Belastung) – 208 (starke Belastung). Die Intra-Class-Korrelationen wurden bei Berechnung der p-Werte berücksichtigt.

**Tabelle 102. Patientenangaben bei Therapieende (E-Messung, Rohwerte)**

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	375	0,61	0,54	212	0,52	0,45	0,18	.038
IIP-D Gesamtwert	381	1,09	0,59	213	1,19	0,55	-0,17	.156
SF12 Körperliches Befinden	376	49,00	10,10	211	50,02	8,45	-0,11	.193
SF12 Psychisches Befinden	376	45,57	11,78	211	47,41	10,49	-0,16	.074
BDI Gesamtwert	261	8,77	8,65	167	8,34	6,36	0,05	.496
AKV Gesamtwert	84	1,71	0,52	25	1,65	0,44	0,12	.712
EDI Gesamtwert	8	281,2	51,3	4	315,5	16,3	-0,78	.242
HZK Gesamtwert	9	23,33	10,97	5	15,88	12,47	0,65	.394
SOM Gesamtwert	17	22,43	21,67	6	40,97	34,02	-0,74	.253

Anmerkung. P2 = 597. Codierung: BSI und IIP-D: 0 (keine Belastung) – 4 (starke Belastung). SF-12-Skalen: 0 (schlechtes Befinden) – 100 (gutes Befinden). BDI: 0 (keine Depressivität) – 63 (starke Depressivität). AKV: 1 (keine Belastung) – 5 (starke Belastung). EDI: 91 (keine Belastung) – 546 (starke Belastung). HZK: 0 (keine Belastung) – 72 (starke Belastung). SOM 0 (keine Belastung) – 208 (starke Belastung). Die Intra-Class-Korrelationen wurden bei Berechnung der p-Werte berücksichtigt.

Die entsprechend lineartransformierten Werte für die Statusmessungen sowie für die Differenzwerte finden sich in Tabelle 103 bis Tabelle 105. Während sich die Patienten der IG zu Therapiebeginn etwas stärker beeinträchtigt auf dem multiplen Statusindikator STA\_PA darstellen als die Patienten der KG, besteht dieser Unterschied bei Therapieende nicht mehr. Während sich die Patienten der IG im Durchschnitt um 11,6 Punkte im Therapieverlauf verbessern, sind es bei der KG 10,9 Punkte. Dieser Unterschied um 0,7 Punkte ( $d = 0,06$ ) ist statistisch nicht signifikant.

Bei den Einzelvergleichen weist die IG zu Therapiebeginn auf den Skalen körperlich und BDI etwas schlechtere Werte als die KG auf, bei Therapieende ist dieser Unterschied nicht mehr vorhanden. Allerdings stellt sich die IG bei Therapieende auf den Skalen BSI und IIP-D etwas besser dar als die KG und bezogen auf den Zuwachs zwischen A und E ist die Prä-Post-Differenz der IG auf der Skala IIP-D etwas größer. Auf dem Hintergrund der statistisch nicht signifikanten Gruppenvergleiche für den multiplen

Status- und Ergebnisindikator sollten diese Befunde allerdings nicht überinterpretiert werden (Gefahr der Alpha-Fehler-Inflation bei Durchführung mehrerer Einzeltests).

Tabelle 103. Patientenangaben bei Therapiebeginn (A-Messung, lineartransformierte Werte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	375	69,5	16,2	212	71,6	15,9	-0,13	.137
IIP-D Gesamtwert	381	61,3	13,0	213	62,2	13,8	-0,07	.459
SF12 Körperliches Befinden	376	46,8	10,4	211	49,0	10,0	-0,21	.013
SF12 Psychisches Befinden	376	32,1	10,0	211	33,0	10,3	-0,09	.279
BDI Gesamtwert	261	67,0	14,3	167	70,2	13,7	-0,23	.019
AKV Gesamtwert	84	56,4	11,4	25	54,0	12,4	0,21	.381
EDI Gesamtwert	8	45,0	6,4	4	39,5	11,1	0,68	.293
HZK Gesamtwert	9	59,1	14,8	5	69,2	15,5	-0,67	.252
SOM Gesamtwert	17	81,7	14,4	6	82,7	11,9	-0,07	.885
<b>Multipler Statusindikator STA_PA</b>	<b>382</b>	<b>54,9</b>	<b>9,3</b>	<b>214</b>	<b>56,7</b>	<b>9,7</b>	<b>-0,19</b>	<b>.028</b>

Anmerkung. P2 = 597. Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden). Die Intra-Class-Korrelationen wurden bei Berechnung der p-Werte berücksichtigt.

Tabelle 104. Patientenangaben bei Therapieende (E-Messung, lineartransformierte Werte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	375	84,7	13,5	212	86,9	11,2	-0,14	.038
IIP-D Gesamtwert	381	72,8	14,8	213	70,2	13,8	0,20	.156
SF12 Körperliches Befinden	376	49,0	10,1	211	50,0	8,4	-0,10	.193
SF12 Psychisches Befinden	376	45,6	11,8	211	47,4	10,5	-0,18	.074
BDI Gesamtwert	261	86,0	13,7	167	86,7	10,1	-0,05	.496
AKV Gesamtwert	84	65,7	10,3	25	67,0	8,9	-0,11	.712
EDI Gesamtwert	8	58,3	11,2	4	50,8	3,6	0,93	.242
HZK Gesamtwert	9	67,6	15,4	5	78,0	17,5	-0,69	.394
SOM Gesamtwert	17	89,2	10,4	6	80,3	16,3	0,64	.253
<b>Multipler Statusindikator STA_PA</b>	<b>382</b>	<b>66,5</b>	<b>10,2</b>	<b>214</b>	<b>67,6</b>	<b>8,3</b>	<b>-0,12</b>	<b>.140</b>

Anmerkung. P2 = 597. Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden). Die Intra-Class-Korrelationen wurden bei Berechnung der p-Werte berücksichtigt.

Tabelle 105. Prä-Post-Differenzen (Vergleich zwischen A- und E-Messung, lineartransformierte Werte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	375	15,2	14,6	212	15,3	16,5	-0,01	.908
IIP-D Gesamtwert	381	11,5	13,6	213	8,0	13,2	0,26	.003
SF12 Körperliches Befinden	376	2,2	10,1	211	1,0	10,2	0,12	.174
SF12 Psychisches Befinden	376	13,5	13,2	211	14,4	13,1	-0,09	.425
BDI Gesamtwert	261	19,0	15,9	167	16,5	14,8	0,18	.094
AKV Gesamtwert	84	9,3	10,9	25	13,0	14,4	-0,32	.175
EDI Gesamtwert	8	13,3	11,4	4	11,3	8,7	0,25	.766
HZK Gesamtwert	9	8,5	10,5	5	8,8	7,7	-0,02	.949
SOM Gesamtwert	17	7,5	8,6	6	-2,4	6,4	0,72	.019
<b>Multipler Ergebnisindikator MEK_PA</b>	<b>382</b>	<b>11,6</b>	<b>10,2</b>	<b>214</b>	<b>10,9</b>	<b>10,1</b>	<b>0,06</b>	<b>.464</b>

Anmerkung. P2 = 597. Codierung: -100 (maximal mögliche Verschlechterung) bis +100 (maximal mögliche Verbesserung). Die Intra-Class-Korrelationen wurden bei Berechnung der p-Werte berücksichtigt.

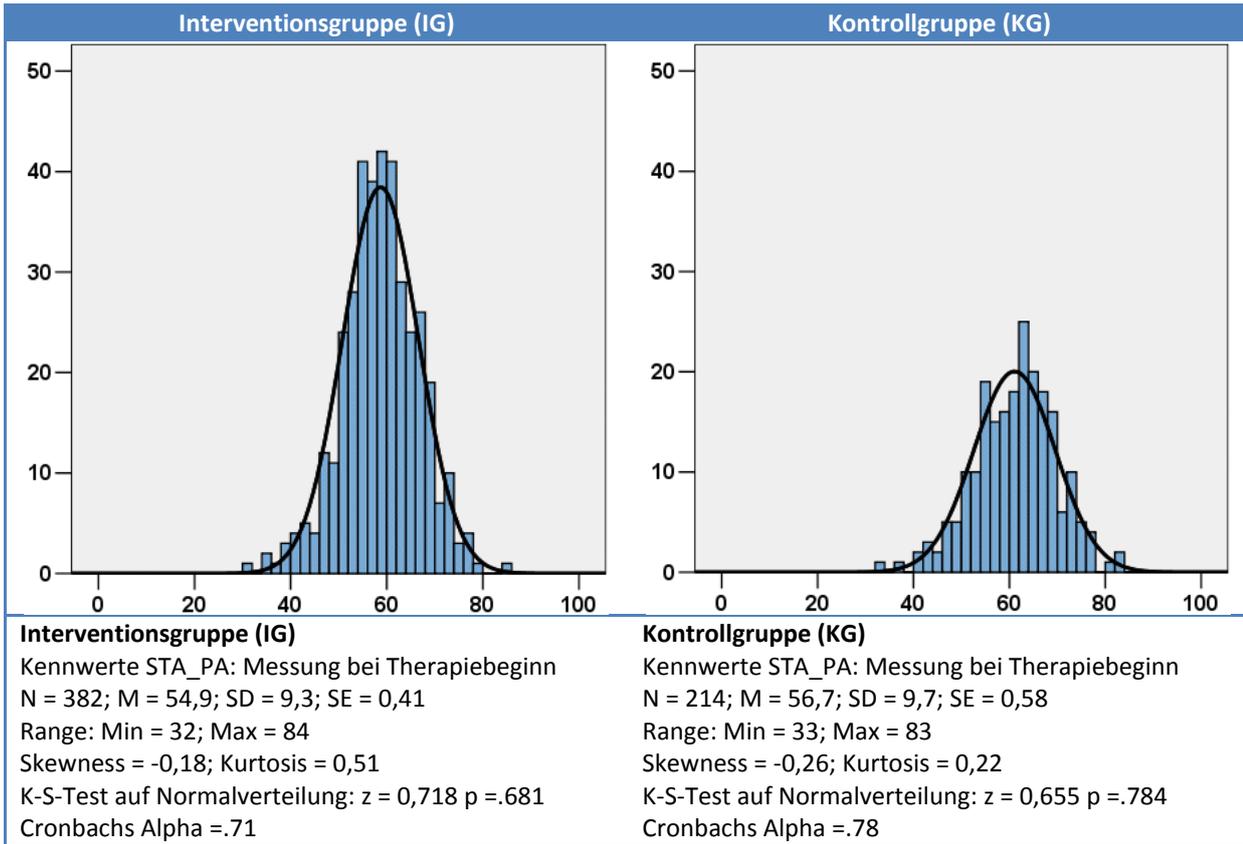


Abbildung 18. P2 = 597. Verteilungskennwerte des multiplen Statusindikators STA\_PA bei Therapiebeginn (A-Messung)

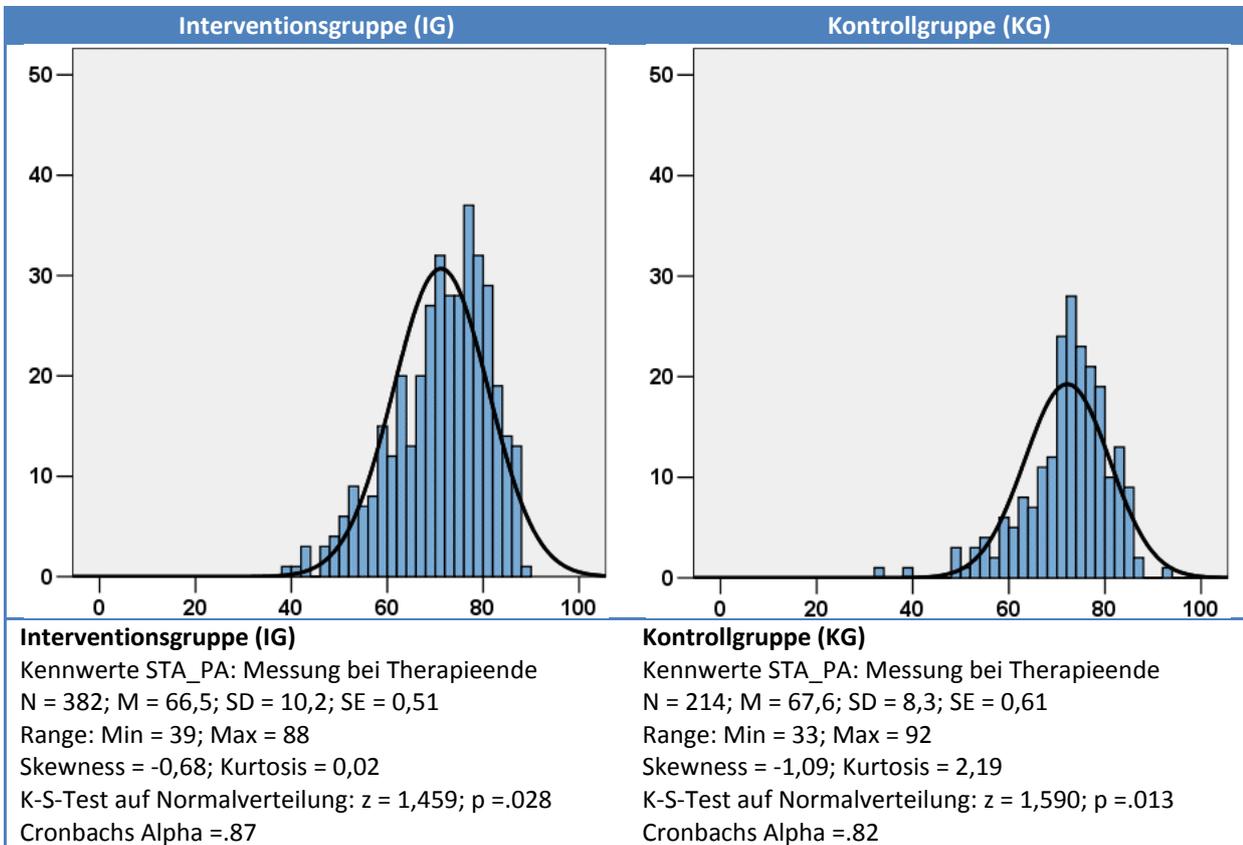


Abbildung 19. P2 = 597. Verteilungskennwerte des multiplen Statusindikators STA\_PA bei Therapieende (E-Messung)

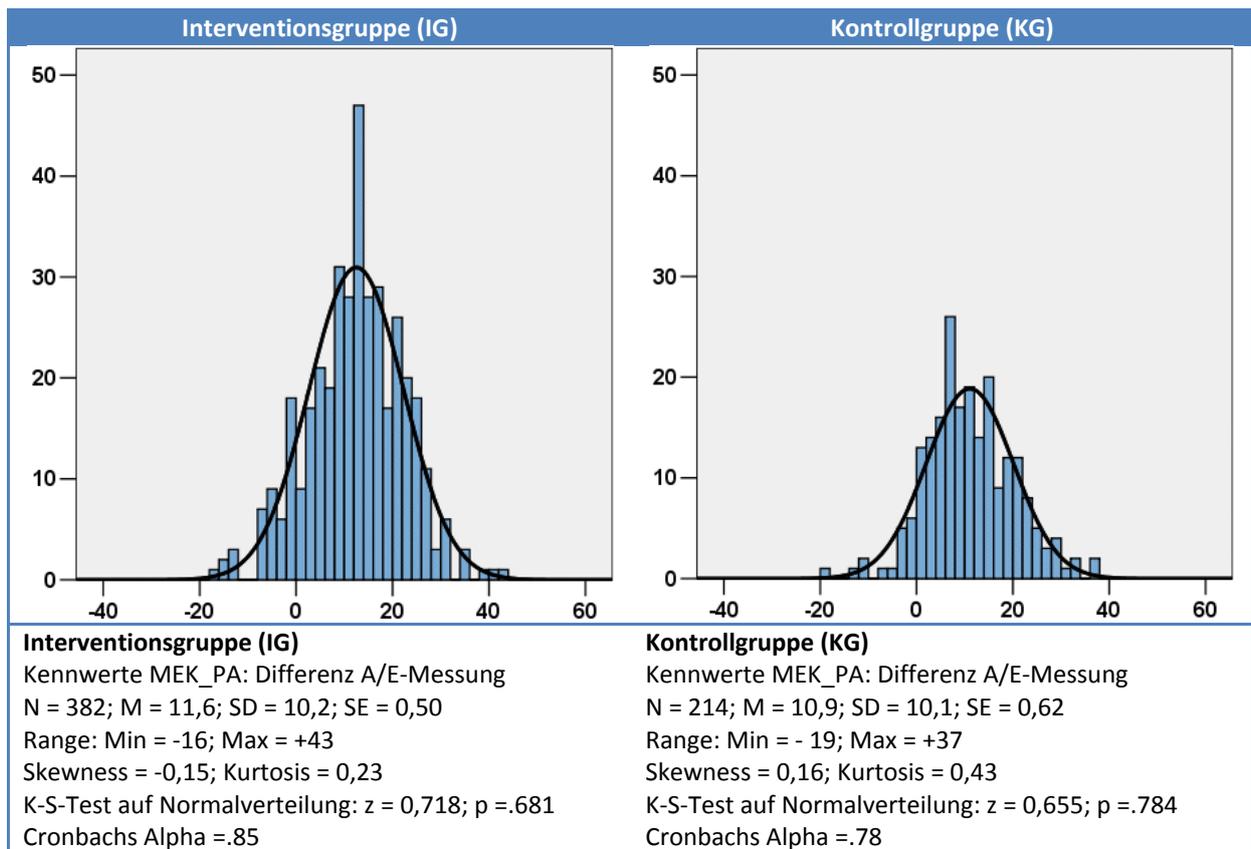


Abbildung 20. P2 = 597. Verteilungskennwerte des multiplen Ergebnisindicators MEK\_PA (A/E-Differenz)

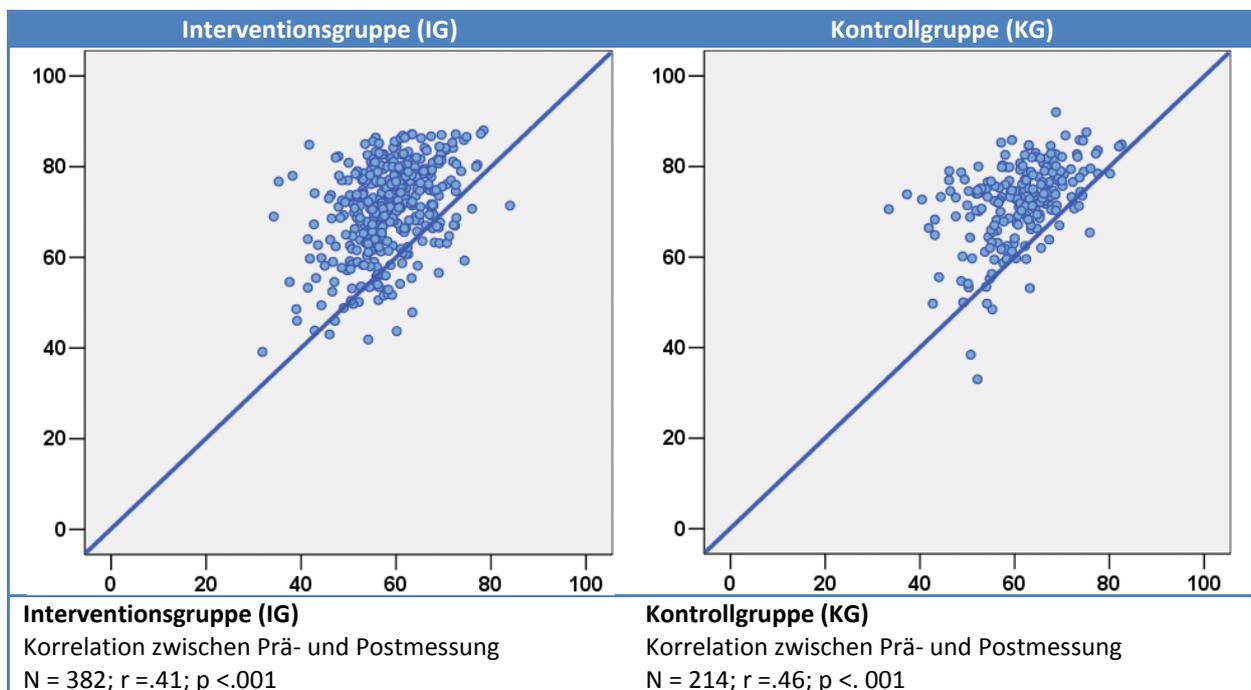


Abbildung 21. P2 = 597. Scatterplot des multiplen Statusindicators STA\_PA (A/E-Vergleich). x-Achse: A-Messung bei Therapiebeginn. y-Achse: E-Messung bei Therapieende. P2

#### 5.4.4 Statistische Kontrolle möglicherweise vorhandener Confounder (Stichprobe P2)

Auch für die Patientenangaben wurden entsprechende multiple Regressionsgleichungen für die Stichproben P2 und P3 berechnet, um mögliche Störvariablen statistisch zu kontrollieren. Die Gesamtgleichung für die Stichprobe P2 wird statistisch signifikant ( $p < .001$ ;  $R = .4868$ ;  $R^2 \text{ adj} = .222$ ). Neben dem Ausgangsniveau zeigt sich wie bereits bei den Therapeutenangaben (vgl. Abschnitt 5.3.5) ein schwacher Partialzusammenhang zwischen der Wohnsituation und dem Therapieergebnis. Aufgenommen wurde auch die Angabe, ob ein Jahr später der Katamnesefragebogen des Patienten vorlag oder nicht (Teilstichprobe P3). Patienten mit Katamnese schneiden bereits bei der Entlassung um  $B = 2,895$  Punkte besser ab, so dass durch Einbeziehung dieses Prädiktors Selektionseffekte insoweit in der Regression berücksichtigt werden.

Die Gruppenzugehörigkeit (IG/KG) hat auch bei statistischer Kontrolle aller einbezogenen Confounder keinen statistisch signifikanten Einfluss auf die mit dem multiplen Ergebnisindikator gemessenen Prä-Post-Veränderungen der Patienten ( $Beta = .00$ ;  $p = .934$ ), d.h. die IG und KG im Modellvorhaben schneiden bei der kurzfristigen Ergebnisqualität zum Entlasszeitpunkt gleich gut ab.

**Tabelle 106. Regressionsgleichung zur statistischen Kontrolle möglicher Confoundervariablen bei Bestimmung möglicherweise vorhandener Unterschiede zwischen IG und KG auf dem multiplen Statusindikator STA\_PA bei Therapieende.**

Kriterium: STA_PA Messung bei Therapieende	B	SE	Beta	t	Sig. p	Toleranz
<i>Regressionskonstante a</i>	40,071	3,301		12,140	<.001	
STA_PA Messung bei Therapiebeginn	0,399	0,038	.39	10,537	<.001	.94
Alter in Jahren bei Therapiebeginn	0,037	0,034	.04	1,089	.277	.82
Geschlecht weiblich (0=nein, 1=ja)	1,212	0,762	.06	1,590	.112	.93
Schulabschluss (0=bis HS, 1=RS, 2=(Fach-)Abitur)	0,720	0,515	.05	1,396	.163	.88
Patient mindestens 15h / Woche erwerbstätig	1,000	0,752	.05	1,330	.184	.88
Wohnsituation allein (0=nein, 1=ja)	-2,498	0,845	-.11	-2,955	.003	.95
Therapiemotivation (0=gering... 3=sehr hoch)	0,720	0,585	.05	1,230	.219	.93
Anzahl der ICD-10-Diagnosen	-0,561	0,403	-.05	-1,392	.164	.91
Zurückliegen Erstmanifestation Krankheit in Jahren	-0,090	0,041	-.08	-2,192	.029	.91
Katamnese-messung (P3) vorhanden (0=nein, 1=ja)	2,895	0,694	.15	4,173	<.001	.99
Gruppenzugehörigkeit (0=KG, 1=IG)	0,061	0,737	.00	0,083	.934	.95

Anmerkung. P2 = 597.

### 5.4.5 Vergleich zwischen Therapiebeginn und Therapieende (Stichprobe P2)

In Tabelle 107 und Tabelle 108 werden deskriptiv die Prä-Post-Effektgrößen für beide Studiengruppen mitgeteilt. Wie bei den Therapeutenratings zeigen sich sehr gute Behandlungsergebnisse und große Effekte ( $d > 0,80$ ) nach Cohen.

Tabelle 107. Patientenangaben der Interventionsgruppe (A/E-Vergleich, lineartransformierte Werte)

Interventionsgruppe (IG)	Therapiebeginn (A)			Therapieende (E)			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	375	69,5	16,2	375	84,7	13,5	0,94	<.001
IIP-D Gesamtwert	381	61,3	13,0	381	72,8	14,8	0,87	<.001
SF12 Körperliches Befinden	376	46,8	10,4	376	49,0	10,1	0,21	<.001
SF12 Psychisches Befinden	376	32,1	10,0	376	45,6	11,8	1,34	<.001
BDI Gesamtwert	261	67,0	14,3	261	86,0	13,7	1,35	<.001
AKV Gesamtwert	84	56,4	11,4	84	65,7	10,3	0,80	<.001
EDI Gesamtwert	8	45,0	6,4	8	58,3	11,2	1,64	.555
HZK Gesamtwert	9	59,1	14,8	9	67,6	15,4	0,57	.018
SOM Gesamtwert	17	81,7	14,4	17	89,2	10,4	0,54	<.001
<b>Multipler Statusindikator STA_PA</b>	<b>381</b>	<b>54,9</b>	<b>9,3</b>	<b>381</b>	<b>66,5</b>	<b>10,2</b>	<b>1,23</b>	<b>&lt;.001</b>

Anmerkung. P2 = 597. Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden).

Tabelle 108. Patientenangaben der Kontrollgruppe (A/E-Vergleich, lineartransformierte Werte)

Kontrollgruppe (KG)	Therapiebeginn (A)			Therapieende (E)			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	212	71,6	15,9	212	86,9	11,2	0,95	<.001
IIP-D Gesamtwert	213	62,2	13,8	213	70,2	13,8	0,60	<.001
SF12 Körperliches Befinden	211	49,0	10,0	211	50,0	8,4	0,10	<.001
SF12 Psychisches Befinden	211	33,0	10,3	211	47,4	10,5	1,42	.003
BDI Gesamtwert	167	70,2	13,7	167	86,7	10,1	1,17	<.001
AKV Gesamtwert	25	54,0	12,4	25	67,0	8,9	1,12	.591
EDI Gesamtwert	4	39,5	11,1	4	50,8	3,6	1,39	.235
HZK Gesamtwert	5	69,2	15,5	5	78,0	17,5	0,59	.038
SOM Gesamtwert	6	82,7	11,9	6	80,3	16,3	0,17	.005
<b>Multipler Statusindikator STA_PA</b>	<b>213</b>	<b>56,7</b>	<b>9,7</b>	<b>213</b>	<b>67,7</b>	<b>8,2</b>	<b>1,16</b>	<b>&lt;.001</b>

Anmerkung. P2 = 597. Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden).

### 5.4.6 Bias-Korrektur der Prä-Post-Effekte (Stichprobe P2)

In Abschnitt 5.4.4 wurde ausgeführt, dass die Prä-Post-Effektgrößen nicht unabhängig von den Drop-Out-Raten sind. Nachfolgend wird daher eine entsprechende Analyse präsentiert (vgl. Tabelle 109). In Fällen, wo neben der E-Messung zusätzlich auch die K-Messung vorliegt, verbessern sich die Patienten der IG um 2,3 und die KG um 2,5 Qualitätspunkte stärker, als wenn die K-Messung nicht vorliegt.

Bei einer Streuung von etwa 10 Punkten entspricht dies einer Effektgröße von  $d = 0,22$  (IG) bzw.  $0,24$  (KG). Generalisiert man die Ergebnisse nun auf die Gesamtstichprobe P1 mit  $N = 1.598$  Patienten (bei denen ja in  $P1 - P2 = 1.598 - 597 = 1.001$  Fällen neben der fehlenden E-Messung auch keine K-Messung vorliegt, so kann von einer entsprechenden Überschätzung der Prä-Post-Effekte ausgegangen werden.

Korrigiert man die in Tabelle 107 für die IG bzw. Tabelle 108 berichteten Effekte konservativ um die Beträge  $0,22$  (IG) bzw.  $0,24$  (KG) entsprechend nach unten, so verbleibt eine Prä-Post-Effektgröße von  $1,23 - 0,22 = 1,01$  für die IG und  $1,16 - 0,24 = 0,92$  für die KG, was noch immer deutlich im Bereich großer Effekte ( $> 0,80$ ) nach Cohen liegt.

Tabelle 109. Prä-Post-Differenzen (Vergleich zwischen A- und E-Messung, lineartransformierte Werte)

	K-Messung (P4) liegt vor			K-Messung (P4) liegt nicht vor			Vergleich	
	N	M	SD	N	M	SD	d	p
<b>Multipler Ergebnisindikator MEK_PA IG</b>	190	12,72	9,80	192	10,46	10,51	0,22	.030
<b>Multipler Ergebnisindikator MEK_PA KG</b>	110	12,14	10,17	104	9,68	10,00	0,24	.076

Anmerkung. P2 = 597. Codierung: -100 (maximal mögliche Verschlechterung) bis +100 (maximal mögliche Verbesserung).

### 5.4.7 Vergleich zwischen Interventions- und Kontrollgruppe (Stichprobe P3)

In diesem Abschnitt werden die Messungen bei Aufnahme mit den Messungen zum Zeitpunkt der 1-Jahres-Katamnese verglichen. In Tabelle 110 und Tabelle 111 sind die Rohwerte der Patientenskalen für die A- und K-Messung der Stichprobe P3 vergleichend für die IG und KG wiedergegeben.

Tabelle 110. Patientenangaben bei Therapiebeginn (A-Messung, Rohwerte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	302	1,19	0,64	160	1,13	0,65		.352
IIP-D Gesamtwert	305	1,50	0,53	158	1,49	0,57		.836
SF12 Körperliches Befinden	302	46,30	10,41	161	47,96	9,67		.095
SF12 Psychisches Befinden	302	33,01	10,68	161	33,19	9,83		.859
BDI Gesamtwert	205	20,72	9,20	124	19,26	8,96		.160
AKV Gesamtwert	71	2,22	0,52	24	2,32	0,60		.432
EDI Gesamtwert	7	342,3	35,2	1	-	-	-	-
HZK Gesamtwert	7	31,86	11,43	2	-	-	-	-
SOM Gesamtwert	11	42,03	33,77	5	28,85	13,94		.421

Anmerkung. P3 = 468. Codierung: BSI und IIP-D: 0 (keine Belastung) – 4 (starke Belastung). SF-12-Skalen: 0 (schlechtes Befinden) – 100 (gutes Befinden). BDI: 0 (keine Depressivität) – 63 (starke Depressivität). AKV: 1 (keine Belastung) – 5 (starke Belastung). EDI: 91 (keine Belastung) – 546 (starke Belastung). HZK: 0 (keine Belastung) – 72 (starke Belastung). SOM 0 (keine Belastung) – 208 (starke Belastung). Die Intra-Class-Korrelationen wurden bei Berechnung der p-Werte berücksichtigt.

Tabelle 111. Patientenangaben bei Therapieende (K-Messung, Rohwerte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	302	0,57	0,56	160	0,55	0,54		.698
IIP-D Gesamtwert	305	1,00	0,61	158	1,08	0,61		.280
SF12 Körperliches Befinden	302	49,09	9,94	161	49,55	8,79		.633
SF12 Psychisches Befinden	302	47,05	11,65	161	46,16	11,79		.539
BDI Gesamtwert	205	8,17	9,17	124	9,62	8,79		.166
AKV Gesamtwert	71	1,59	0,45	24	1,62	0,48		.821
EDI Gesamtwert	7	280,7	45,9	1	-	-	-	-
HZK Gesamtwert	7	19,14	9,65	2	-	-	-	-
SOM Gesamtwert	11	30,56	28,81	5	11,18	11,01		.107

Anmerkung. P3 = 468. Codierung: BSI und IIP-D: 0 (keine Belastung) – 4 (starke Belastung). SF-12-Skalen: 0 (schlechtes Befinden) – 100 (gutes Befinden). BDI: 0 (keine Depressivität) – 63 (starke Depressivität). AKV: 1 (keine Belastung) – 5 (starke Belastung). EDI: 91 (keine Belastung) – 546 (starke Belastung). HZK: 0 (keine Belastung) – 72 (starke Belastung). SOM 0 (keine Belastung) – 208 (starke Belastung). Die Intra-Class-Korrelationen wurden bei Berechnung der p-Werte berücksichtigt.

Die entsprechend lineartransformierten Werte für die Statusmessungen sowie für die Differenzwerte finden sich in Tabelle 112 bis Tabelle 114. Während sich die Patienten der IG zu Therapiebeginn etwas stärker beeinträchtigt auf dem multiplen Statusindikator STA\_PA darstellen als die Patienten der KG, besteht dieser Unterschied bei Therapieende nicht mehr.

Während sich die Patienten der IG beim Vergleich zwischen Aufnahme- und Katamnese-messung im Durchschnitt um 12,4 Punkte im Therapieverlauf verbessern, sind es bei der KG 10,8 Punkte. Dieser Unterschied um 1,6 Punkte ( $d = 0,15$ ) ist statistisch nicht signifikant ( $p = .127$ ). Auch bei den Einzelvergleichen zeigen sich mit Ausnahme der BDI Prä-Post-Differenz-Werte (die IG verbessert sich etwas stärker als die KG) keine signifikanten Unterschiede zwischen IG und KG.

Tabelle 112. Patientenangaben bei Therapiebeginn (A-Messung, lineartransformierte Werte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	302	70,2	16,0	160	71,7	16,3	0,09	.352
IIP-D Gesamtwert	305	62,6	13,1	158	62,9	14,2	0,02	.836
SF12 Körperliches Befinden	302	46,3	10,4	161	48,0	9,7	0,17	.095
SF12 Psychisches Befinden	302	33,0	10,7	161	33,2	9,8	0,02	.859
BDI Gesamtwert	205	67,1	14,6	124	69,4	14,2	0,16	.160
AKV Gesamtwert	71	55,7	10,3	24	53,7	12,0	0,19	.432
EDI Gesamtwert	7	44,7	7,6	1	-	-	-	-
HZK Gesamtwert	7	55,7	15,9	2	-	-	-	-
SOM Gesamtwert	11	79,6	16,3	5	86,2	6,8	0,46	.421
<b>Multipler Statusindikator STA_PA</b>	<b>306</b>	<b>55,3</b>	<b>9,4</b>	<b>162</b>	<b>56,6</b>	<b>9,9</b>	<b>-0,14</b>	<b>.159</b>

Anmerkung. P3 = 468. Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden). Die Intra-Class-Korrelationen wurden bei Berechnung der p-Werte berücksichtigt.

Tabelle 113. Patientenangaben bei 1-Jahres-Katamnese (K-Messung, lineartransformierte Werte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	302	85,8	14,0	160	86,3	13,4	-0,03	.698
IIP-D Gesamtwert	305	75,0	15,2	158	73,0	15,3	0,15	.280
SF12 Körperliches Befinden	302	49,1	9,9	161	49,6	8,8	-0,05	.633
SF12 Psychisches Befinden	302	47,1	11,7	161	46,2	11,8	0,09	.539
BDI Gesamtwert	205	87,0	14,5	124	84,7	13,9	0,16	.166
AKV Gesamtwert	71	68,1	9,0	24	67,6	9,5	0,05	.821
EDI Gesamtwert	7	58,3	10,0	1	-	-	-	-
HZK Gesamtwert	7	73,6	13,6	2	-	-	-	-
SOM Gesamtwert	11	82,2	13,8	5	94,8	5,3	-0,88	.107
<b>Multipler Statusindikator STA_PA</b>	<b>306</b>	<b>67,7</b>	<b>10,7</b>	<b>162</b>	<b>67,4</b>	<b>10,0</b>	<b>0,03</b>	<b>.764</b>

Anmerkung. P3 = 468. Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden). Die Intra-Class-Korrelationen wurden bei Berechnung der p-Werte berücksichtigt.

Tabelle 114. Prä-Post-Differenzen (Vergleich zwischen A- und K-Messung, lineartransformierte Werte)

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	302	15,6	16,0	160	14,6	17,2	0,06	.523
IIP-D Gesamtwert	305	12,4	14,1	158	10,1	13,8	0,17	.080
SF12 Körperliches Befinden	302	2,8	9,7	161	1,6	9,9	0,12	.209
SF12 Psychisches Befinden	302	14,1	14,1	161	13,0	13,7	0,11	.433
BDI Gesamtwert	205	19,9	16,7	124	15,3	17,3	0,32	.016
AKV Gesamtwert	71	12,4	10,8	24	13,9	13,1	-0,14	.578
EDI Gesamtwert	7	13,6	10,2	1	-	-	-	-
HZK Gesamtwert	7	17,9	8,8	2	-	-	-	-
SOM Gesamtwert	11	2,6	7,0	5	8,6	3,5	-0,42	.378
<b>Multipler Ergebnisindikator MEK_PA</b>	<b>306</b>	<b>12,4</b>	<b>10,9</b>	<b>162</b>	<b>10,8</b>	<b>10,9</b>	<b>0,15</b>	<b>.127</b>

Anmerkung. P3 = 468. Codierung: -100 (maximal mögliche Verschlechterung) bis +100 (maximal mögliche Verbesserung). Die Intra-Class-Korrelationen wurden bei Berechnung der p-Werte berücksichtigt.

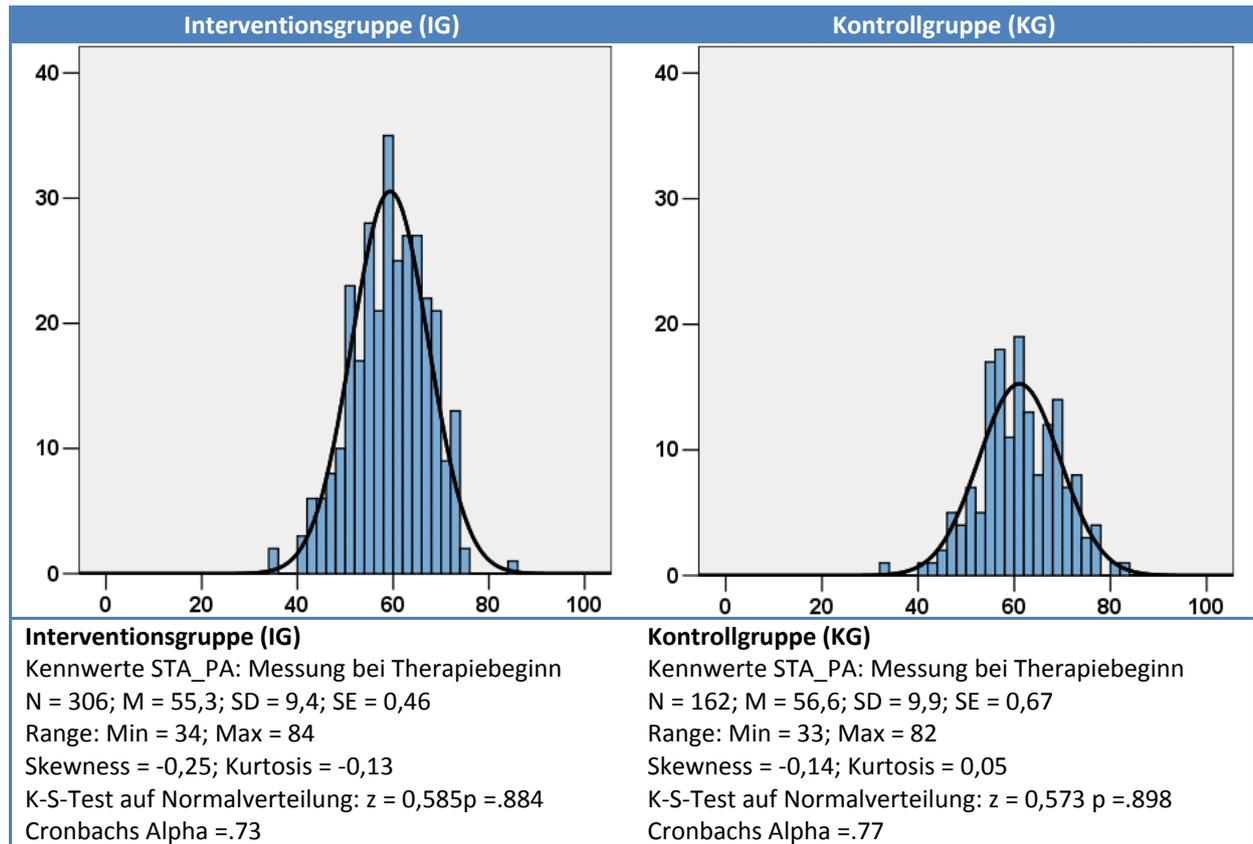


Abbildung 22. Verteilungskennwerte des multiplen Statusindicators STA\_PA bei Therapiebeginn (A-Messung). P3 = 468.

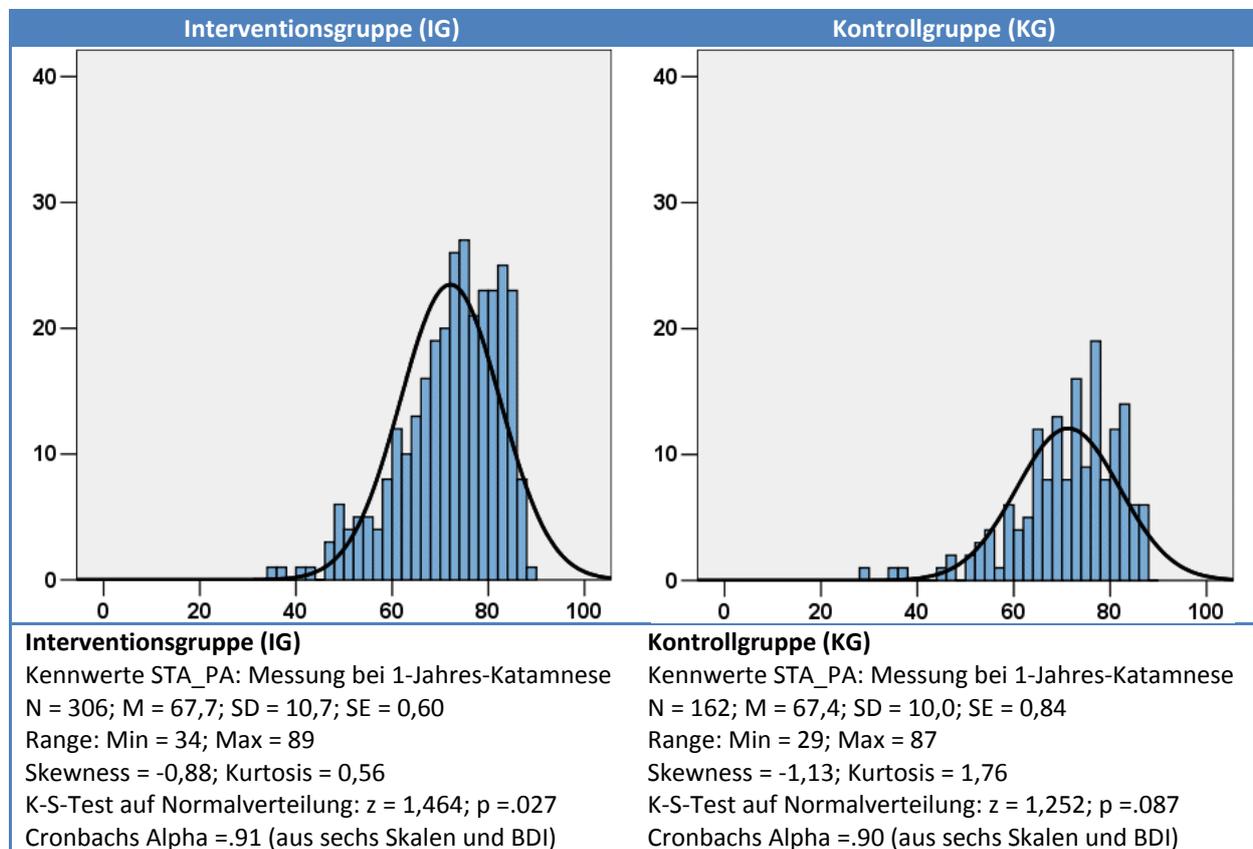


Abbildung 23. Verteilungskennwerte des multiplen Statusindicators STA\_PA bei 1-Jahres-Katamnese (K-Messung). P3 = 468.

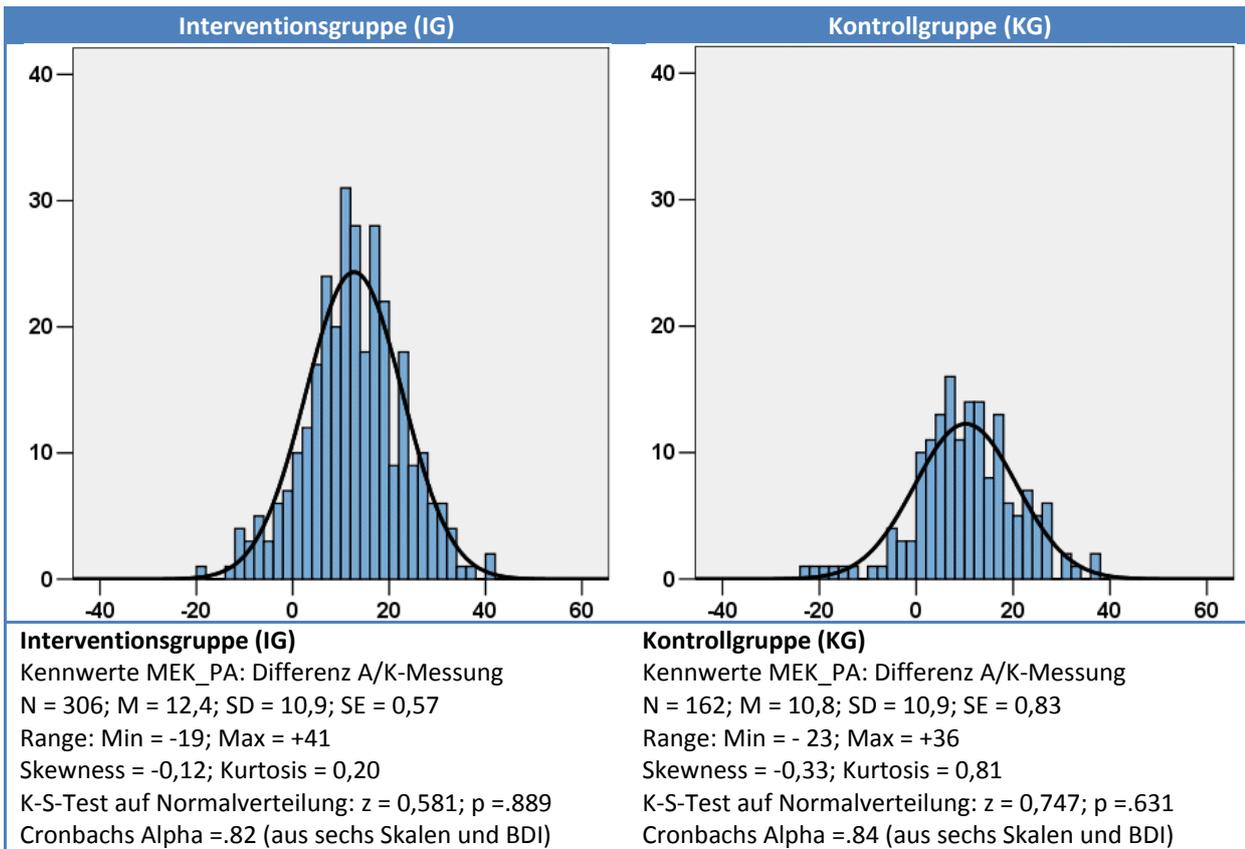


Abbildung 24. Verteilungskennwerte des multiplen Ergebnisindicators MEK\_PA (A/K-Differenz)

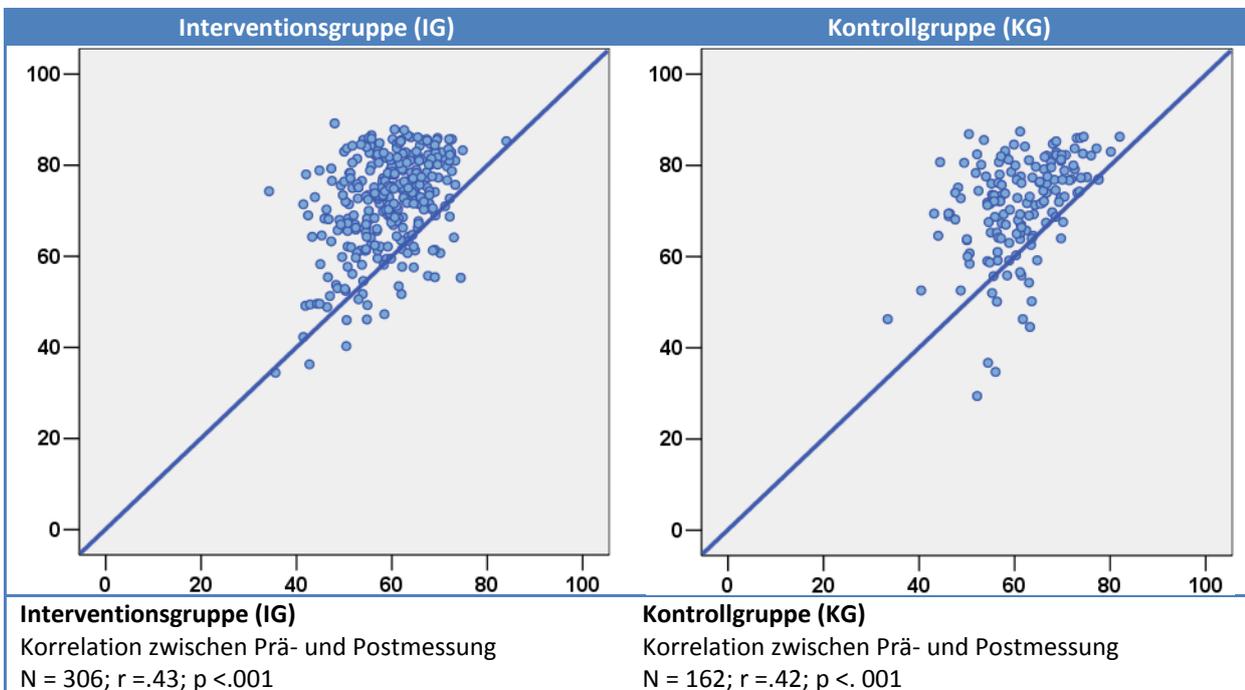


Abbildung 25. Scatterplot des multiplen Statusindicators STA\_PA (A/K-Vergleich). x-Achse: A-Messung bei Therapiebeginn. y-Achse: K-Messung bei Therapieende. P3 = 468.

#### 5.4.8 Statistische Kontrolle möglicherweise vorhandener Confounder (Stichprobe P3)

Berechnet man auch für den Vergleich zwischen Aufnahme- und Katamnese eine Regressionsgleichung ( $p < .001$ ;  $R = .450$ ;  $R^2 \text{ adj} = .184$ ), so ergibt sich folgender Befund: Neben der deutlichen Abhängigkeit vom Ausgangsniveau ( $\text{Beta} = .39$ ) zeigen sich zwei weitere schwache Partialzusammenhänge: So ist auch zum Katamnesezeitpunkt das Merkmal „Wohnsituation allein“ negativ ( $\text{Beta} = .08$ ), das Vorliegen der Entlassmessung ein Jahr vorher hingegen positiv ( $\text{Beta} = .09$ ) assoziiert.

Die Gruppenzugehörigkeit (IG/KG) hat bei statistischer Kontrolle der einbezogenen Confounder auch in der 1-Jahres-Katamnese keinen signifikanten Einfluss auf die mit dem multiplen Ergebnisindikator gemessenen Prä-Post-Veränderungen der Patienten ( $\text{Beta} = .05$ ;  $p = .279$ ).

Tabelle 115. Regressionsgleichung zur statistischen Kontrolle möglicher Confoundervariablen beim Vergleich zwischen IG und KG auf dem multiplen Statusindikator STA\_PA 1-Jahres-Katamnese.

Kriterium: STA_PA Messung bei Therapieende	B	SE	Beta	t	Sig. p	Toleranz
<i>Regressionskonstante a</i>	36,561	4,301		8,501	<.001	
STA_PA Messung bei Therapiebeginn	0,427	0,047	.39	9,096	<.001	.93
Alter in Jahren bei Therapiebeginn	0,043	0,042	.05	1,024	.306	.87
Geschlecht weiblich (0=nein, 1=ja)	0,401	0,949	.02	0,423	.673	.95
Schulabschluss (0=bis HS, 1=RS, 2=(Fach-)Abitur)	0,730	0,671	.05	1,088	.277	.89
Patient mindestens 15h / Woche erwerbstätig	1,632	0,931	.08	1,754	.080	.92
Wohnsituation allein (0=nein, 1=ja)	-2,397	1,020	-.10	-2,351	.019	.97
Therapiemotivation (0=gering... 3=sehr hoch)	1,283	0,755	.07	1,699	.090	.90
Anzahl der ICD-10-Diagnosen	-0,243	0,525	-.02	-0,463	.644	.92
Zurückliegen Erstmanifestation Krankheit in Jahren	-0,058	0,056	-.04	-1,043	.298	.95
Entlassmessung (P2) vorhanden (0=nein, 1=ja)	1,942	0,929	.09	2,090	.037	.96
Gruppenzugehörigkeit (0=KG, 1=IG)	1,025	0,946	.05	1,084	.279	.94

Anmerkung. P3 = 468.

### 5.4.9 Vergleich zwischen Therapiebeginn und 1-Jahres-Katamnese (Stichprobe P3)

Auch für die Patientensichtweise werden deskriptiv die Prä-Post-Effektgrößen mitgeteilt. Wie bei den Therapeutenratings zeigen sich sehr gute Behandlungsergebnisse und große Effekte ( $d > 0,80$ ) nach Cohen.

Tabelle 116. Patientenangaben der Interventionsgruppe (A/K-Vergleich, lineartransformierte Werte)

Interventionsgruppe (IG)	Therapiebeginn (A)			Katamnese (K)			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	302	70,2	16,0	302	85,8	14,0	0,97	<.001
IIP-D Gesamtwert	305	62,6	13,1	305	75,0	15,2	0,92	<.001
SF12 Körperliches Befinden	302	46,3	10,4	302	49,1	9,9	0,28	<.001
SF12 Psychisches Befinden	302	33,0	10,7	302	47,1	11,7	1,36	<.001
BDI Gesamtwert	205	67,1	14,6	205	87,0	14,5	1,38	<.001
AKV Gesamtwert	71	55,7	10,3	71	68,1	9,0	1,15	<.001
EDI Gesamtwert	7	44,7	7,6	7	58,3	10,0	1,79	<.001
HZK Gesamtwert	7	55,7	15,9	7	73,6	13,6	1,13	<.001
SOM Gesamtwert	11	79,6	16,3	11	82,2	13,8	0,18	.026
<b>Multipler Statusindikator STA_PA</b>	<b>306</b>	<b>55,3</b>	<b>9,4</b>	<b>306</b>	<b>67,7</b>	<b>10,7</b>	<b>1,29</b>	<b>&lt;.001</b>

Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden).

Tabelle 117. Patientenangaben der Kontrollgruppe (A/K-Vergleich, lineartransformierte Werte)

Kontrollgruppe (KG)	Therapiebeginn (A)			Katamnese (K)			Vergleich	
	N	M	SD	N	M	SD	d	p
BSI Gesamtwert	160	71,7	16,3	160	86,3	13,4	0,91	<.001
IIP-D Gesamtwert	158	62,9	14,2	158	73,0	15,3	0,75	<.001
SF12 Körperliches Befinden	161	48,0	9,7	161	49,6	8,8	0,16	<.001
SF12 Psychisches Befinden	161	33,2	9,8	161	46,2	11,8	1,25	.003
BDI Gesamtwert	124	69,4	14,2	124	84,7	13,9	1,06	<.001
AKV Gesamtwert	24	53,7	12,0	24	67,6	9,5	1,29	.591
EDI Gesamtwert	1	-	-	1	-	-	-	-
HZK Gesamtwert	2	-	-	2	-	-	-	-
SOM Gesamtwert	5	86,2	6,8	5	94,8	5,3	0,60	.005
<b>Multipler Statusindikator STA_PA</b>	<b>162</b>	<b>56,6</b>	<b>9,9</b>	<b>162</b>	<b>67,4</b>	<b>10,0</b>	<b>1,13</b>	<b>&lt;.001</b>

Codierung: 0 (schlechtes Befinden) – 100 (gutes Befinden).

### 5.4.10 Bias-Korrektur der Prä-Post-Effekte (Stichprobe P3)

Auch für die Stichprobe P3 wurde eine Analyse zur Abschätzung eines möglichen Positiv-Bias berechnet (vgl. Abschnitt 5.4.6). Tabelle 118 zeigt die Ergebnisse. In Fällen, wo neben der K-Messung zusätzlich auch die E-Messung vorliegt, verbessern sich die Patienten der IG um 2,3 und die KG um 2,5 Qualitätspunkte stärker, als wenn die E-Messung nicht vorliegt.

Bei einer Streuung von etwa 10 Punkten entspricht dies einer Effektgröße von  $d = 0,22$  (IG) bzw.  $0,24$  (KG). Generalisiert man die Ergebnisse nun auf die Gesamtstichprobe P1 mit  $N = 1.598$  Patienten (bei denen ja in  $P1 - P3 = 1.598 - 300 = 1.298$  Fällen neben der fehlenden E-Messung auch keine K-Messung vorliegt, so kann von einer entsprechenden Überschätzung der Prä-Post-Effekte ausgegangen werden.

Korrigiert man die in Tabelle 116 für die IG bzw. Tabelle 117 berichteten Effekte konservativ um die Beträge  $0,22$  (IG) bzw.  $0,24$  (KG) entsprechend nach unten, so verbleibt eine Prä-Post-Effektgröße von  $1,29 - 0,22 = 1,07$  für die IG und  $1,13 - 0,20 = 0,93$  für die KG, was auch für K-Messung noch immer deutlich im Bereich großer Effekte ( $> 0,80$ ) nach Cohen liegt.

Tabelle 118. Prä-Post-Differenzen (Vergleich zwischen A- und E-Messung, lineartransformierte Werte)

	K-Messung (P3) liegt vor			K-Messung (P3) liegt nicht vor			Vergleich	
	N	M	SD	N	M	SD	d	p
<b>Multipler Ergebnisindikator MEK_PA IG</b>	190	12,76	10,86	116	11,90	10,99	0,08	.505
<b>Multipler Ergebnisindikator MEK_PA KG</b>	110	11,50	10,61	52	9,37	11,38	0,20	.245

Anmerkung. P3 = 468. Codierung: -100 (maximal mögliche Verschlechterung) bis +100 (maximal mögliche Verbesserung).

#### 5.4.11 Verlaufsmessung auf dem multiplen Ergebnisindikator (Stichprobe P4)

Abbildung 26 veranschaulicht die Entwicklung für die IG und KG über die drei Messzeitpunkte A, E und K grafisch. Hinsichtlich der Verläufe auf dem Multiplen Ergebnisindikator zeigen sich keine statistisch signifikante Unterschiede zwischen IG und KG in der Stichprobe P4.

Die Effekte sind zwischen E- und K-Zeitpunkt zudem in beiden Gruppen stabil: Nach der deutlichen statistisch signifikanten Verbesserung zwischen A- und E-Zeitpunkt ist die Reduktion zwischen E- und K-Zeitpunkt um 0,2 Punkte (IG) bzw. 0,5 Punkte (KG) statistisch nicht signifikant.

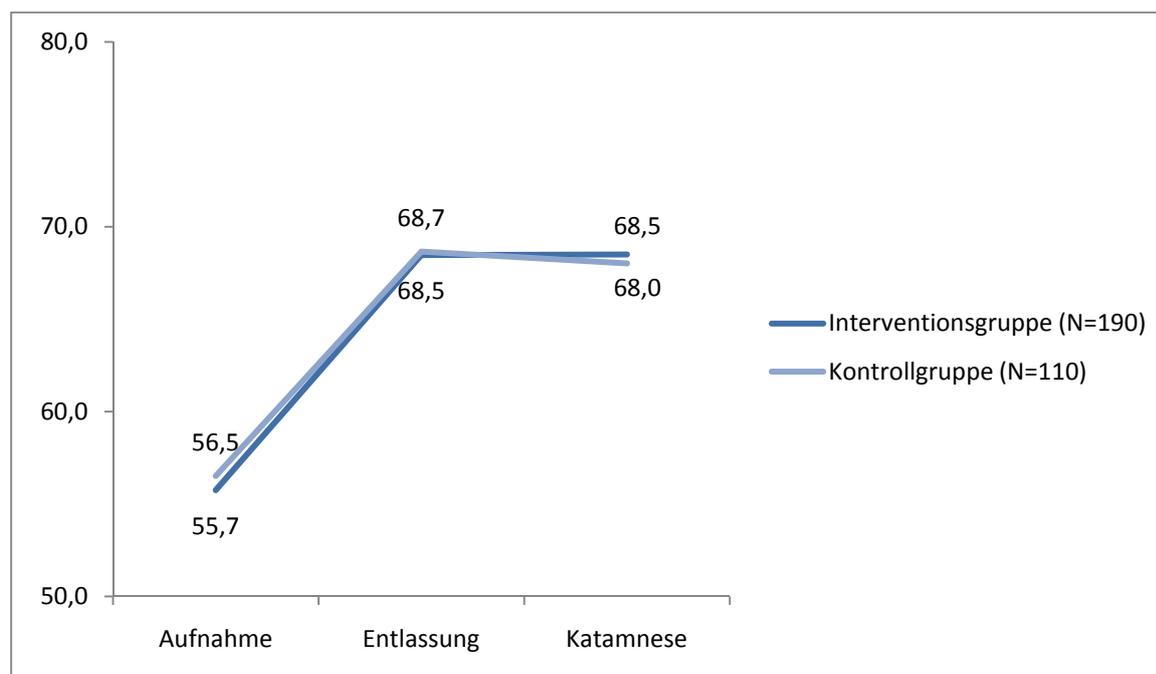


Abbildung 26. Dreipunkt-Messung auf dem multiplen Statusindikator STA\_PA (P4 = 300 Patienten).

#### 5.4.12 Explorative Abschätzung der Kosten-Nutzen-Relation anhand der Effektgrößen

Geht man davon aus, dass der hier beobachtete Therapieeffekt in einer Größenordnung von 1,0 mindestens ein Jahr lang anhält, so ergibt sich pro Patient folgende gesamtgesellschaftliche Kosten-Nutzen-Schätzung (Schmidt, Hunter & Pearlman, 1982; vgl. Abschnitt 2.7) bereits für das erste Jahr nach Therapieende, wenn man von durchschnittlich 32 bewilligten Therapiesitzungen (vgl. Abschnitt 6.3.3) mit 120 Euro direkten und indirekten Kosten (Verwaltung etc. mit berechnet) davon ausgeht, dass der Zusammenhang der Verbesserung des Allgemeinbefindens mit der Produktivität (Arbeitsfähigkeit)  $a = .50$  beträgt - so geht das Statistische Bundesamt für das Jahr 2005 von einer Bruttowertschöpfung der Wirtschaft in Höhe von 2.022,5 Mrd. Euro aus. Umgerechnet auf 38,8 Millionen Erwerbstätige im gleichen Jahr entspricht dies einem Betrag von durchschnittlich 52.126 Euro pro Jahr und Patient. 40% davon

(konservative Schätzung, der resultierende Break-Even-Point beträgt 0,31) entsprechen einer Standardabweichung der Produktivität von 20.850 Euro pro Patient:

$$U = N \cdot T \cdot d \cdot a \cdot SD_{prod} - N \cdot K$$

$$U = 1 \cdot 1 \cdot 1,0 \cdot 0,5 \cdot 20.850 - 1 \cdot 3.200 = 7.225 \text{ Euro pro Patient}$$

Dies entspricht einem Kosten-Nutzen-Verhältnis von  $10.425 / 3.200 = 3,25 : 1$ . Jeder in die ambulante Psychotherapie investierte Euro zahlt sich gesamtgesellschaftlich gesehen also dreifach aus unter der Prämisse, dass der Therapieeffekt ein Jahr lang nach Ende der Behandlung anhält.

## 6 Entscheidungsregeln im Modellvorhaben (Mesoebene)

### 6.1 Entscheidungsregeln zu Beginn sowie im Verlauf

Nach jedem Messzeitpunkt wurde eine Rückmeldung an den Therapeuten generiert, die dieser mit dem Patienten besprechen konnte. Die Rückmeldung bestand aus einer graphischen Darstellung der Symptomentwicklung sowie einer Handlungsempfehlung in Textform. Die Rückmeldungen wurden für jeden Patienten individuell nach bestimmten Regeln generiert. Diese Regeln werden im folgenden Abschnitt dargestellt. Dabei wird zunächst auf die Rückmeldung zu Therapiebeginn eingegangen und im Anschluss auf die Rückmeldungen im Therapieverlauf.

Die Rückmeldung zu Therapiebeginn für die IG enthielt die Entscheidungsregeln bezüglich der Aufnahme einer Psychotherapie. Dabei wurde eine Handlungsempfehlung abgegeben, die entweder eine Langzeittherapie indizierte oder aber die Aufnahme eines modifizierten Gutachterverfahrens verlangte (Abbildung 27).

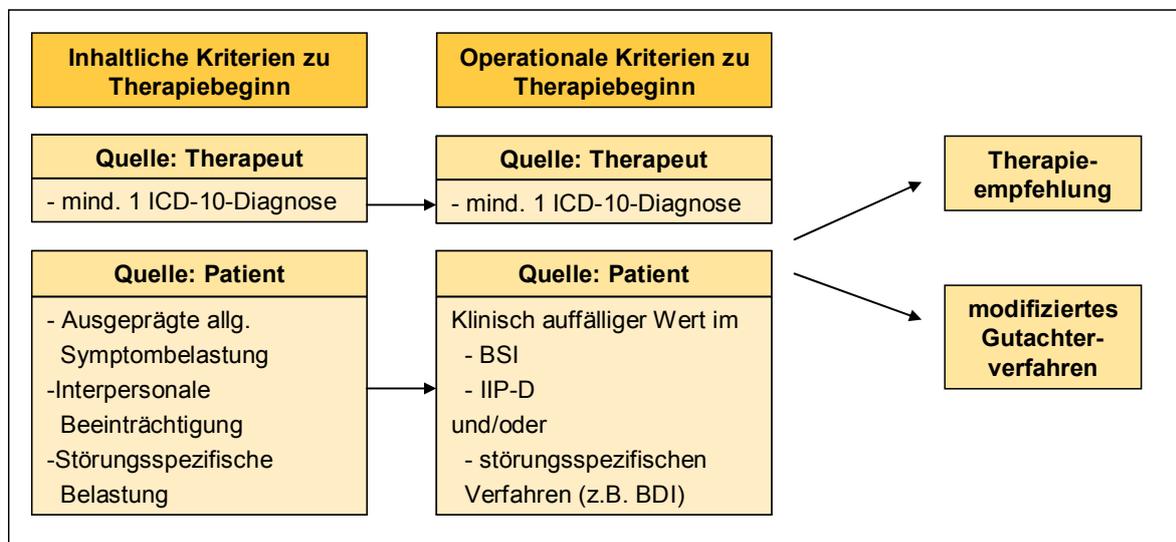


Abbildung 27. Kriterien der Entscheidung Therapieempfehlung vs. Modifiziertes Gutachterverfahren

Die Empfehlung zur Aufnahme einer Langzeittherapie wurde dann abgegeben, wenn für den Patienten mindestens eine Diagnose nach ICD-10 gestellt wurde und wenn der Patient in mindestens einem der drei zentralen Selbstbeurteilungsinstrumente einen klinisch auffälligen Wert erreichte. Dazu wurden die Rohwerte in T-Werte transformiert und mit den Referenzstichproben verglichen. Von einem klinisch auffälligen Wert spricht man, wenn die Wahrscheinlichkeit für den Patienten größer ist, zur klinischen Population zu gehören als zur funktionalen Normalpopulation (Cut-Off überschritten). Wenn der Wert größer ist als der Mittelwert der klinischen Referenzpopulation, kann man von einer schweren Belastung ausgehen. Dies ist in Abbildung 28 graphisch dargestellt.

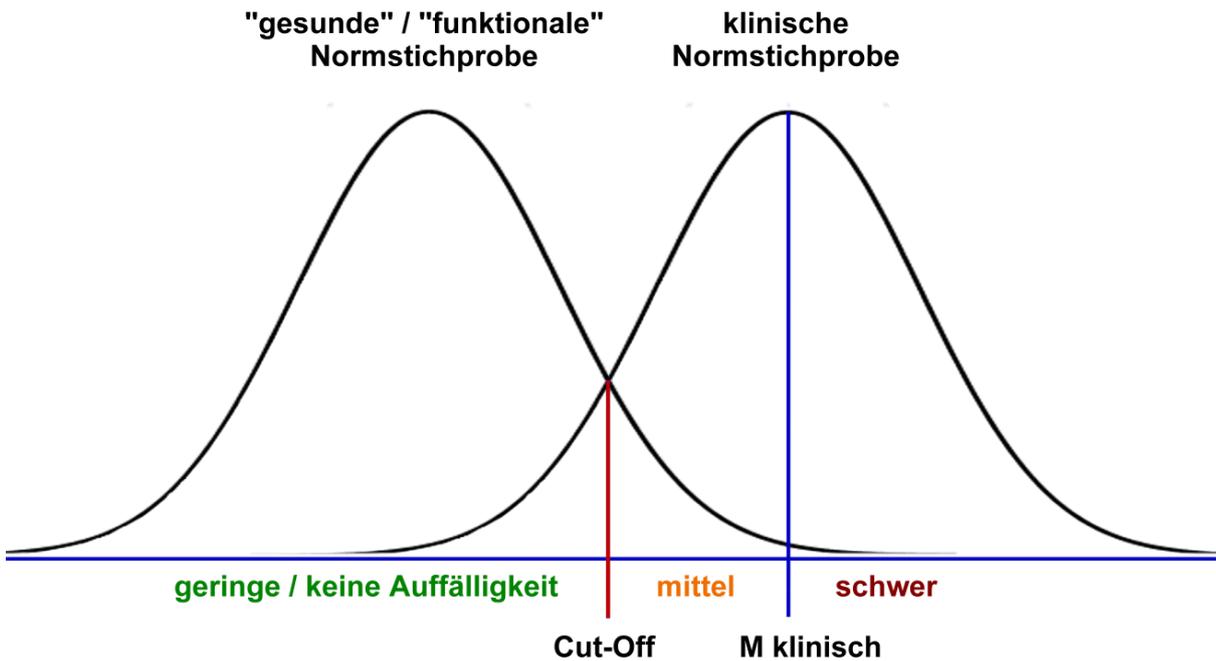


Abbildung 28. Veranschaulichung klinischer Auffälligkeiten

Für die Rückmeldungen während des Therapieverlaufs wurde das Konzept der klinisch bedeutsamen Veränderung von Jacobson und Truax (1991) verwendet. Um als klinisch signifikant verbessert zu gelten, müssen nach Jacobson und Truax zwei Kriterien erfüllt sein: Die Veränderung zwischen zwei Messzeitpunkten muss statistisch reliabel sein (die Wahrscheinlichkeit, dass die beobachtete Veränderung zufällig ist, beträgt höchstens 5%). Außerdem muss der Cut-Off Wert überschritten werden in Richtung funktionale Stichprobe. Gleichzeitig wurde im Rückmeldesystem der Anfangszustand des Patienten (Prämessung) mitberücksichtigt. Es ergeben sich verschiedene, in Tabelle 119 dargestellte Möglichkeiten (Kriterien A bis E), wie die Rückmeldung je nach Ausgangspunkt und je nach Veränderung aussehen kann (Lutz, et al., 2004).

Tabelle 119. Mögliche Variationen von Ausgangspunkt und Veränderung

<b>Kriterium A</b>	klinisch signifikante Veränderung
<b>Kriterium B</b>	Verbesserung, aber bereits zu Beginn im Normbereich
<b>Kriterium C</b>	reliable Verbesserung, am Ende nicht im Normbereich
<b>Kriterium D</b>	nicht reliabel verändert oder unverändert
<b>Kriterium E</b>	Reliabel verschlechtert

Im TK-Projekt wurden die Kriterien zusammen mit den Ausgangswerten in drei mögliche Rückmeldungen integriert: Positive Rückmeldung (+), neutrale (0) und negative (-) Rückmeldung. Dazu wurden die Patienten zunächst danach eingeteilt, ob sie zu Beginn der Therapie in dem jeweiligen Messinstrument leicht, mittel oder schwer belastet waren. Abhängig von dieser Ausgangslage mussten die Patientinnen und Patienten bis zum nächsten Erhebungszeitpunkt unterschiedliche Kriterien erfüllen, um in die Kategorie für eine mögliche positive Verlaufsrückmeldung fallen. Patienten, die bei der Prä-Messung unterhalb des Cut-Off Wertes lagen (damit also eher zur Normalpopulation gehören), wurden als „leicht“ belastete Patienten eingestuft. Diejenigen dagegen, deren Werte zwischen dem Cut-Off Wert und dem Mittelwert der klinischen Referenzgruppe lagen, wurden als „mittel-schwer“ eingestuft, und solche, die

oberhalb des Mittelwertes der klinischen Referenzgruppe lagen, als „schwer“. Patientinnen und Patienten, die Kriterium A erfüllten, müssen zu Therapiebeginn außerhalb des Bereichs der „Normalpopulation“ gelegen haben (d. h. eine mittlere oder schwere Belastung aufgewiesen haben) und lagen nach der Behandlung im Normbereich der „Gesunden“, sie haben sich reliabel verbessert. Das Kriterium B entspricht einer Verbesserung der Patienten (reliabel oder nicht), jedoch haben diese Patienten die Therapie bereits im Normbereich des jeweiligen Messinstruments begonnen. Tabelle 120 veranschaulicht das Prinzip.

Tabelle 120. Kombinierte Rückmeldeempfehlung für BSI, IIP-D und BDI

	Test	Prä-Wert	Post-Wert	erfülltes Kriterium	Rückmeldung Instrument	Rückmeldung gesamt
<b>BSI</b>						
Patient A	leicht	0,34	0,02	B	+	
Patient B	schwer	1,62	0,79	C	0	
Patient C	schwer	1,63	0,21	A	+	
Patient D	mittel	1,08	0,26	A	+	
Patient E	mittel	0,77	0,75	D	0	
Patient F	leicht	0,09	0,30	E	-	
etc.						
<b>IIP-D</b>						
Patient A	leicht	3,25	1,13	B	+	
Patient B	schwer	16,50	13,40	D	0	
Patient C	schwer	16,00	7,00	A	+	
Patient D	leicht	7,88	8,75	D	0	
Patient E	leicht	9,38	12,00	D	0	
Patient F	leicht	12,00	12,90	D	0	
etc.						
<b>BDI</b>						<b>BSI / IIP-D / BDI</b>
Patient A	leicht	4	0	B	+	(+,+,+) = +
Patient B	schwer	24	15	C	0	(0,0,0) = 0
Patient C	schwer	31	1	A	+	(+,+,+) = +
Patient D	leicht	6	2	B	+	(+,0,+) = +
Patient E	mittel	17	10	A	+	(0,0,+) = 0
Patient F	leicht	8	11	D	0	(-,0,0) = -
etc.						

Entsprechend der unterschiedlichen Ausgangslage der Patienten wurden spezifische Kriterien angelegt, um von einer klinisch bedeutsamen Verbesserung bei einem einzelnen Patienten zu sprechen. Schwer belastete Patienten mussten im Sinne der klinisch signifikanten Verbesserung sowohl eine reliable Verbesserung in ihrem Ausprägungsgrad aufweisen als auch den Cut-Off Wert hin zur Normalpopulation unterschreiten (Kriterium A). Patienten hingegen, die mit einer mittelschweren Belastung in die Therapie kamen, mussten entweder die Kriterien der klinisch signifikanten Verbesserung wie die schwer belasteten erfüllen, oder aber sie zeigten eine reliable Verbesserung und befanden sich am Ende noch nicht im Normbereich (Kriterium A oder C). Patienten, die lediglich leicht belastet in dem jeweiligen Bereich in die Therapie kamen, benötigten lediglich eine positive Veränderungsdifferenz in dem Messinstrument, um als klinisch bedeutsam verbessert eingestuft zu werden (Kriterium B). Unter Berücksichtigung der Ausgangswerte wurde also eingeschätzt, ob die Patienten das Kriterium für eine bedeutsame Veränderung in einem bestimmten Bereich erfüllten. Die Stabilität der Veränderung wurde anhand einer Erfolgsbewertung zum darauf folgenden Erhebungszeitpunkt untersucht, wobei wiederum die Differenz zum ersten Erhebungszeitpunkt als Bezugspunkt herangezogen wurde (Lutz et al., 2005). Abbildung 29 illustriert die Generierung der Rückmeldungen aus den Entscheidungskriterien.

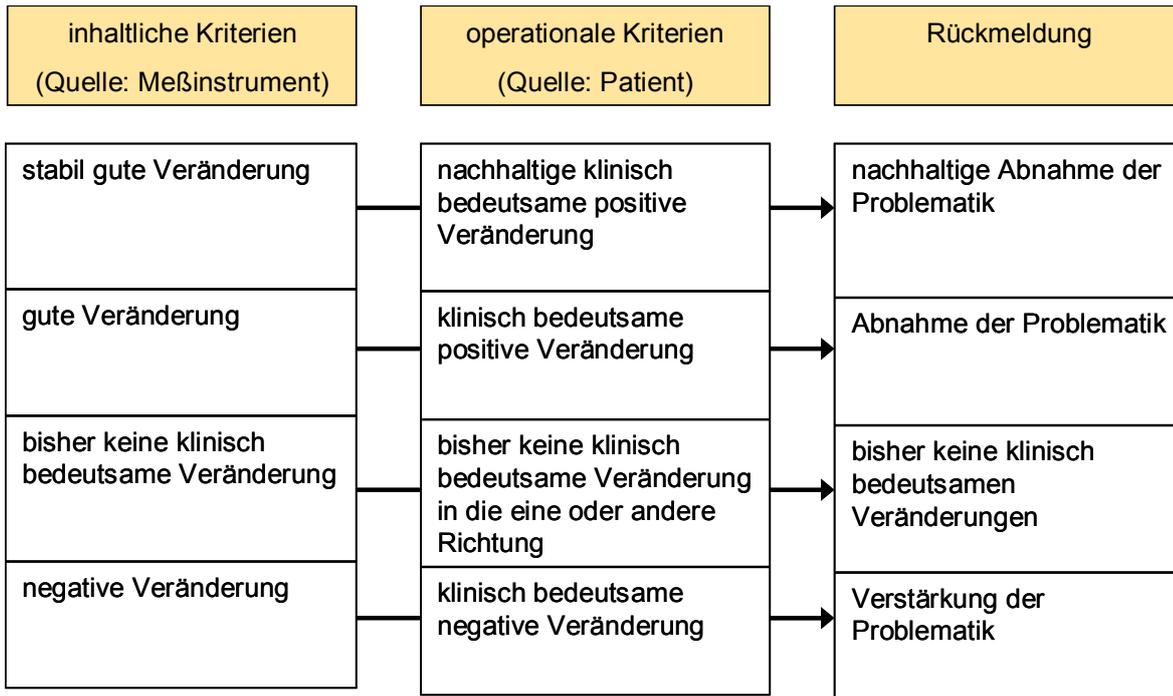


Abbildung 29. Inhaltliche und operationale Kriterien sowie die dazugehörigen Empfehlungen

Je nach Ergebnis der Rückmeldung wurden unterschiedliche Handlungsempfehlungen in Form von Textbausteinen abgegeben. Es handelt sich bei diesen Rückmeldungen um unverbindliche Empfehlungen, die je nach therapeutischem Hintergrund unterschiedlich interpretiert werden können. Es resultierten vier mögliche Gesamteinschätzungen der Patientenangaben: Verstärkung der Problematik, keine klinisch bedeutsame Veränderung, Abnahme der Problematik und nachhaltige Abnahme der Problematik. Die Therapeuten wurden in der Gesamtrückmeldung darauf aufmerksam gemacht, dass die Ergebnisse unterschiedliche Ursachen haben können. Entsprechend wurden sie aufgefordert, das bisherige Vorgehen zu überdenken und gegebenenfalls Anpassungen vorzunehmen. In Abbildung 30 ist der genaue Wortlaut der Textbausteine wiedergegeben.

#### **Gesamteinschätzung der Patientenangaben: Verstärkung der Problematik**

Die verschiedenen Messinstrumente zeigen im Vergleich zu den Werten bei Therapiebeginn bisher (noch) keine Abnahme der Problematik in der Selbsteinschätzung des Patienten, sondern deuten eher auf eine Verstärkung hin.

Dies kann unterschiedliche Ursachen haben. Bitte überdenken Sie daher den bisherigen Therapieverlauf unter dem Aspekt der Angemessenheit oder eventuellen Anpassung ihrer bisherigen therapeutischen Strategien an die Belange dieser Patientin/ dieses Patienten.

Berücksichtigen Sie dabei vor allem mögliche Erklärungen für die im Moment vorliegende stärker ausgeprägte Problematik. Sie kann auch ein erwünschter Teil des therapeutischen Prozesses sein (z.B. im Sinne einer notwendigen Destabilisierung).

#### **Gesamteinschätzung der Patientenangaben: Bisher keine klinisch bedeutsamen Veränderungen**

Die Erhebung in den verschiedenen Messinstrumenten zeigt keine eindeutige Entwicklung dieser Patientin / dieses Patienten im Sinne einer klinisch bedeutsamen Abnahme oder Verstärkung der Problematik in der subjektiven Selbsteinschätzung im Vergleich zu den Werten bei Therapiebeginn.

Dies kann unterschiedliche Ursachen haben. Bitte überdenken Sie den bisherigen Therapieverlauf auch mit Blick auf eine eventuelle Vermeidung problematischer Verhaltensweisen, auf Verdrängung der Hintergründe der psychischen Störung oder auf Probleme in der therapeutischen Beziehung.

#### **Gesamteinschätzung der Patientenangaben: Abnahme der Problematik.**

Die Erhebung in den verschiedenen Messinstrumenten zeigt eine klinisch bedeutsame Abnahme der Problematik in der Selbsteinschätzung dieser Patientin/ dieses Patienten im Vergleich zu den Werten bei Therapiebeginn.

Dies kann unterschiedliche Ursachen haben. Bitte überdenken Sie daher den bisherigen Therapieverlauf v.a. unter dem Aspekt der Angemessenheit oder eventuellen Anpassung ihrer bisherigen therapeutischen Strategien an die Belange dieser Patientin/ dieses Patienten. Berücksichtigen Sie dabei vor allem mögliche Erklärungen für die im Moment vorliegende positive Entwicklung. Es kann sich um eine produktive Entwicklung handeln, es kann aber auch sein, dass es sich um eine Vermeidung bzw. Verdrängung der Hintergründe der psychischen Störung handelt, durch die der verbesserte Zustand des Patienten nur vorübergehend/temporär wäre.

#### **Gesamteinschätzung der Patientenangaben: Nachhaltige Abnahme der Problematik.**

Die Erhebungen in den verschiedenen Messinstrumenten zeigen zum wiederholten Male eine klinisch bedeutsame und eventuell nachhaltige Abnahme der Problematik in der Selbsteinschätzung dieser Patientin / dieses Patienten im Vergleich zu den Werten bei Therapiebeginn.

Bitte überdenken Sie erneut den bisherigen Therapieverlauf v. a. unter dem Aspekt der Angemessenheit oder eventuellen Anpassung ihrer therapeutischen Strategien an die Belange dieser Patientin/ dieses Patienten. Berücksichtigen Sie dabei vor allem mögliche Erklärungen für die im Moment vorliegende nachhaltige Abnahme der Problematik. Auch diese könnte noch vorübergehend/temporär sein und würde dementsprechend eine weitere Auseinandersetzung mit der Problematik bzw. weitere therapeutische

Interventionen erfordern. Reflektieren Sie nochmals eventuell weiter bestehende selbst- und fremdschädigende Anteile/Verhaltensweisen des Patienten sowie mögliche Abwehr- oder Vermeidungsstrategien, welche gegebenenfalls einer dauerhaften Besserung im Wege stehen.

Prüfen Sie weitere therapeutische Schritte im Sinne einer Fortsetzung / Intensivierung der Therapie oder der Vorbereitung des Behandlungsabschlusses (Ablösungsprozess einleiten, Rückfallprophylaxe planen).

**Abbildung 30. Genauer Wortlaut der Textbausteine in der zum 01.04.2007 modifizierten Fassung (die ursprüngliche Fassung s. Anhang)**

Die Rückmeldung stützte sich auch bei den Verlaufsmessungen auf die drei zentralen Selbstbeurteilungsinstrumente. Für jedes der drei Instrumente wurde eruiert, ob sich eine positive, eine negative oder eine neutrale Rückmeldung ergab. Die Ergebnisse der drei Instrumente wurden im Anschluss zu einem Gesamturteil integriert, wobei die Einzelmessungen aufsummiert wurden. Negative Summenwerte führten zu einem negativen Gesamturteil, eine Summe von 0 oder +1 zu einem neutralen und 2 oder 3 + zu einem positiven Gesamturteil. Auch bei den Verlaufsrückmeldungen wurde zusätzlich zu den Handlungsempfehlungen ein graphischer Report für jedes der zentralen Instrumente erzeugt. Ein Beispiel für einen BSI-Report ist in Abbildung 31 zu sehen.

**Wie schätzt Ihre Patientin/ Ihr Patient die allgemeine psychische Belastung ein?**

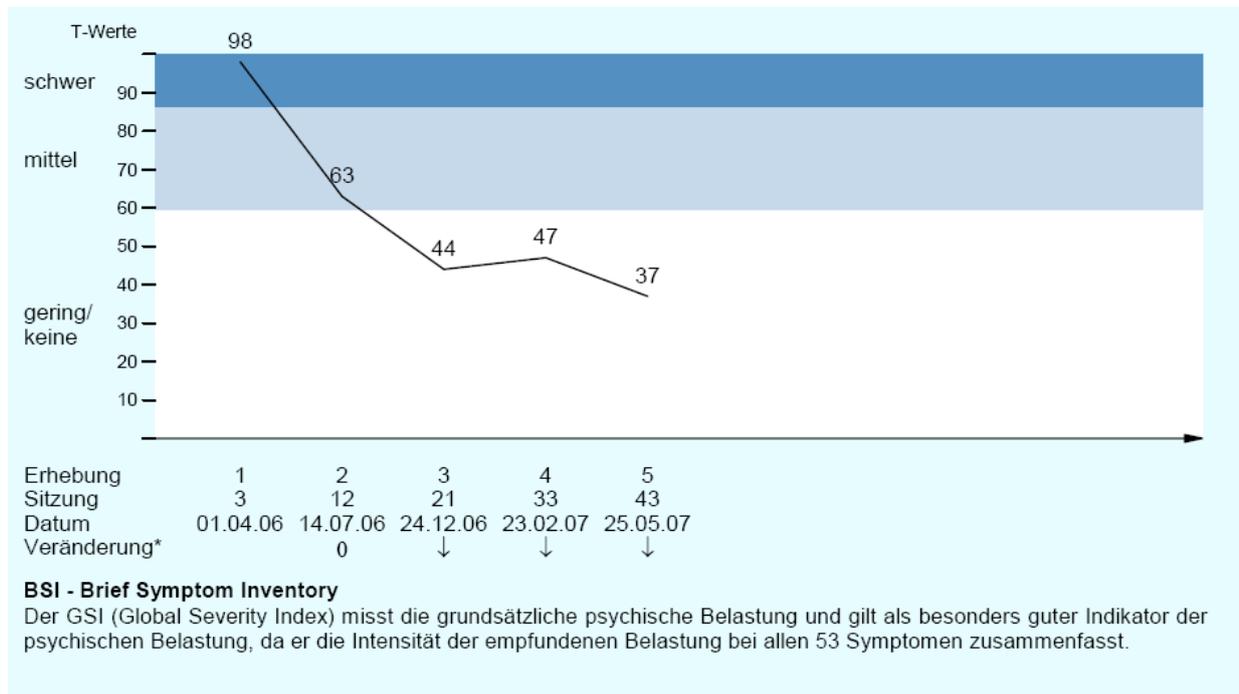


Abbildung 31. Beispiel eines BSI-Reports in der zum 01.04.2007 modifizierte Fassung (Beispiel-Report für die ursprüngliche Fassung s. Anhang).

Da sich der fiktive Patient auch in den anderen Instrumenten verbessert zeigte, ergab sich ein positives Gesamturteil und es wurde eine nachhaltige Abnahme der Problematik rückgemeldet.

Das für dieses Modellprojekt verwendete Rückmeldesystem kann im Rahmen der Diskussion um rationale vs. empirische Rückmelderegeln gesehen werden. Zu Projektbeginn wurde entschieden, ein rationales System zu verwenden. Darunter ist ein Rückmeldesystem zu verstehen, welches nach vorgegebenen Regeln entwickelt wurde und bei dem Grenzwerte aufgrund früherer Datensätze festgelegt wurden (vgl. Lambert et al, 2002). Die Grenzen wurden so festgelegt, dass die Anteile der verschiedenen Rückmeldungen in einer bestimmten Bandbreite liegen. Bei der Evaluation der Regeln geht es darum zu prüfen, ob die vorhergesagten Werte mit der neuen Stichprobe eingehalten werden (siehe Kap. Evaluation der Entscheidungsregeln). Für weitere Ausführungen zum Rückmeldesystem sei auf die Arbeiten von Lutz et al. (2004, 2005) sowie die aktuelle Publikation der Arbeitsgruppe zum Modellvorhaben (Lutz et al., 2009) verwiesen.

## 6.2 Evaluation der Entscheidungsregeln

In den folgenden Abschnitten werden die Entscheidungsregeln evaluiert. Dazu braucht es die Angaben zu den Eingangsdiagnosen (vgl. Abschnitt 4.2.2), welche zentral für die Auswahl des störungsspezifischen Instruments sind. Dieses wiederum ist ein zentrales Element in den Rückmeldungen. Die Eingangsmessungen dienen als Grundlage für die Entscheidungsregeln zu Beginn. Die Verlaufsmessungen dienen zur Generierung von Rückmeldungen im Verlauf. Schließlich wird noch auf abgeschlossene Messungen eingegangen.

### 6.2.1 Patienten mit und ohne Komorbidität bei den Eingangsdiagnosen

Eine Übersicht über das Auftreten bestimmter Komorbiditätsmuster in Interventions- und Kontrollgruppe bietet Tabelle 121. Der Anteil an Patienten mit nur einer Diagnose ist in der KG mit 62,8% gegenüber der IG mit 49,7% bedeutsam erhöht. Dementsprechend sind die drei hier definierten Komorbiditätsmuster (Komorbid mit Persönlichkeitsstörung, Affektive und Angststörung sowie Andere Muster) in der IG gegenüber der KG bedeutsam erhöht. Schränkt man die Datenanalyse allerdings nur auf diejenigen Fälle mit mindestens zwei Diagnosen (N = 677; IG: 449; KG: 228) ein, so zeigen sich keine signifikanten Abweichungen der beobachteten von der erwarteten Verteilung zwischen IG und KG mehr.

**Tabelle 121. Häufigkeiten und Anteile spezifischer Störungskombinationen in Interventions- und Kontrollgruppe**

	IG		KG		Vergleich	
	N	%	N	%	d	p
Keine gültige Diagnose vorhanden	13	1,4	15	2,3	-0,07	.192
Keine Komorbidität (nur eine Diagnose)	457	49,7	410	62,8	-0,26	<.001
Komorbid mit PK-Störung (F60-61)	100	10,9	47	7,2	0,12	.013
Affektive (F3) und Angststörung (F40-41)	124	13,5	51	7,8	0,18	<.001
Andere Muster	225	24,5	130	19,9	0,11	.033
Gesamt	919	100,0	653	100,0		

**Anmerkung. Stichprobe T1, Therapieeingangsdagnostik des Therapeuten (N = 1.572).**

Ein Vergleich der Patienten mit Einfachdiagnose mit komorbiden Patienten im Hinblick auf die Eingangsbelastung in den störungsunspezifischen Maßen BSI und IIP-D ist für die Gesamtgruppe aller Patienten in Tabelle 122. Die Mittelwertsunterschiede zwischen den Gruppen "keine Komorbidität" und "komorbid mit Persönlichkeitsstörung" sowie "keine Komorbidität" und "komorbide Angst- und Affektive Störung" werden für beide Maße signifikant. Dagegen unterscheiden sich Patienten mit anderen Störungskombinationen im Mittel nicht bedeutsam von solchen ohne Komorbidität.

Tabelle 122. Eingangsbelastung in BSI und IIP-D in Abhängigkeit von komorbiden Störungen

	keine Komorbidität (1)	komorbid mit PK-Störungen (2)	komorbid Angst und Affekt. Stör. (3)	komorbid andere Kombinationen (4)	ANOVA F (ges) / sign. Einzelvergleiche
<b>BSI</b>	M = 1,16 SD = 0,66 N = 802	M = 1,42 SD = 0,66 N = 135	M = 1,37 SD = 0,72 N = 167	M = 1,19 SD = 0,64 N = 338	F = 9,155 p<.001 (1-2), (1-3)
<b>IIP-D</b>	M = 1,47 SD = 0,54 N = 807	M = 1,75 SD = 0,55 N = 138	M = 1,64 SD = 0,52 N = 168	M = 1,55 SD = 0,54 N = 337	F = 13,273 p<.001 (1-2), (1-3)

Anmerkung. Alle Patienten mit vorhandener Diagnose und psychometrischer Messung zu Therapiebeginn. Die Darstellung erfolgt gemeinsam für alle Patienten der IG und KG.

Abbildung 32 zeigt eine Einteilung in Diagnosegruppen, wie sie für diverse Auswertungen vorgenommen wird. Hieraus geht ebenfalls deutlich hervor, dass Affektive Störungen, Anpassungsstörungen und verschiedene Angststörungen wie zu erwarten am häufigsten als Erstdiagnose gestellt werden. Die Affektiven Störungen gemeinsam genommen nehmen hier den größten Platz ein.

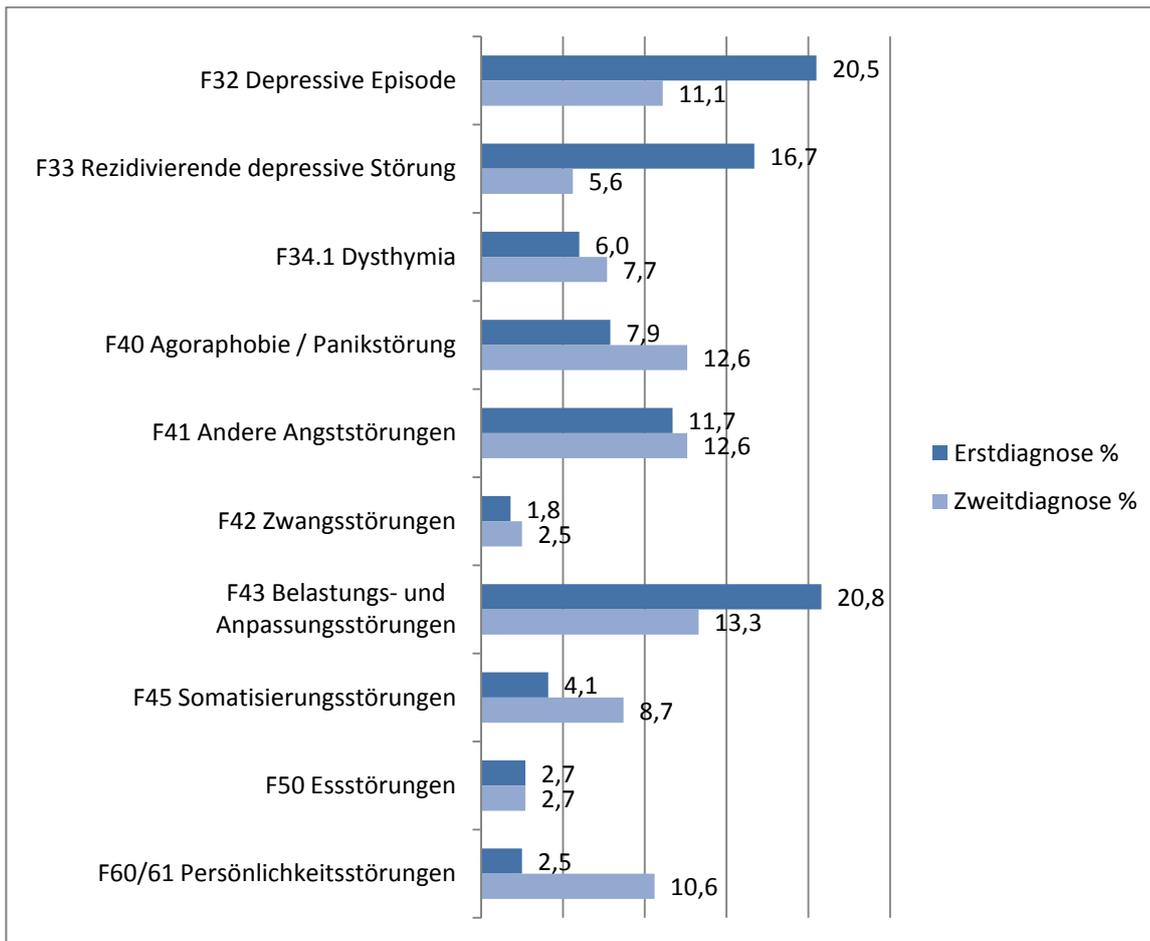


Abbildung 32. Häufigkeiten der Erst- und Zweitdiagnosen zu Therapiebeginn. Die prozentualen Anteile beziehen sich auf 1.544 Patienten mit vorhandener Erst- bzw. 677 Patienten mit vorhandener Zweitdiagnose. Die Darstellung erfolgt gemeinsam für alle Patienten der IG und KG.

## 6.2.2 Psychometrische Diagnostik zu Therapiebeginn: Überprüfung der Grenzwerte

Die Tabelle 123 und Tabelle 124 enthalten einen Vergleich der Eingangsmessungen in der TK-Studie mit Normstichproben. In dem allgemeinen Instrument IIP-D ergibt sich kein signifikanter Mittelwertsunterschied zwischen den ins Modellvorhaben eingegangenen Patienten und der ambulanten Normstichprobe. Im BSI ist der TK-Mittelwert (M=1,23) im Vergleich zur ambulanten Normstichprobe (M=1,14) leicht erhöht. Ebenso liegt für den BDI der TK-Mittelwert (M=20,60) eher im Bereich der stationären (M=19,32) als der ambulanten Normstichprobe (M=16,05). Die Tendenz zu leicht erhöhten Werten in der TK-Stichprobe im Vergleich zur Klinischen Vergleichsstichprobe setzt sich in den weiteren störungsspezifischen Instrumenten fort.

Tabelle 123. Mittelwerte der Eingangsmessungen und deren Vergleich mit Normstichproben

	Patienten TK-Studie (1)	Nichtklinische Stichprobe (2)	Ambulante Stichprobe (3)	Stationäre Stichprobe (4)	ANOVA <sup>a</sup> F (ges) / sign. Einzelvergleiche
<b>BSI</b>	M = 1,23 SD = 0,67 N = 1.577	M = 0,31 SD = 0,23 N = 600 <sup>b</sup>	M = 1,14 SD = 0,67 N = 1.235 <sup>c</sup>	--- <sup>d</sup>	F = 512,70 (1-2), (1-3), (2-3)
<b>IIP-D</b>	M = 1,55 SD = 0,55 N = 1.585	M = 1,28 SD = 0,51 N = 3.047 <sup>e</sup>	M = 1,55 SD = 0,49 N = 464 <sup>f</sup>	M = 1,91 SD = 0,45 N = 3.170 <sup>g</sup>	F = 844,42 (1-2), (1-4), (2-3), (2-4), (3-4)
<b>BDI</b>	M = 20,60 SD = 9,23 N = 1.155	M = 6,45 SD = 5,20 N = 86 <sup>h</sup>	M = 16,05 SD = 8,83 N = 445 <sup>i</sup>	M = 19,32 SD = 9,63 N = 4.342 <sup>j</sup>	F = 77,29 (1-2), (1-3), (1-4), (2-3), (2-4), (3-4)

Anmerkungen. Stichprobe (1) entspricht P1 = 1.598 Patienten mit vorhandener Erstmessung (IG: 1.011, KG: 587).

<sup>a</sup> Die Einzelvergleiche basieren auf einem .05 Alpha Niveau. Die Angaben in den Klammern zeigen die signifikanten Einzelvergleiche zwischen den Stichproben (1), (2), (3) und (4).

<sup>b</sup> Franke (2000): Mitarbeiterinnen und Mitarbeiter am Essener Institut für medizinischen Psychologie.

<sup>c</sup> Geisheim (2002): Drei ambulante Stichproben Christoph-Dornier-Stiftung (Braunschweig, Dresden, Marburg).

<sup>d</sup> Als stationäre Vergleichsstichprobe lagen keine adäquaten deutschen Stichproben vor. Vergleichbare stationäre Stichproben erhoben mit dem SCL-90-R ergaben GSI-Werte zwischen 1.24 (0,66) und 1.38 (0,63) (vgl. Tritt et al., 2003 und Schauenburg et al., 2001).

<sup>e</sup> Horowitz et al. (2000): Referenzpopulation bildet eine Stichprobe des Meinungsforschungsinstituts USUMA (Berlin). Auswahl der Personen erfolgt nach dem Random-Route-Verfahren.

<sup>f</sup> Berner Datensatz: Ambulante Stichprobe der Praxisstelle der Universität Bern.

<sup>g</sup> Mestel et al. (2001): Stationäre Stichprobe psychosomatische Klinik.

<sup>h</sup> Hautzinger et al. (2005): Kontrollgruppe sind psychiatrisch und psychosomatisch unauffällige und gesunde Personen

<sup>i</sup> Ehlers & Margraf (2001): Ambulante oder stationäre Patienten einer psychosomatischen Fachklinik; Kontrollgruppe ist ohne psych. Störung.

<sup>j</sup> Die Einzelvergleiche basieren auf einem .05 Alpha Niveau. Die Angaben in den Klammern zeigen die signifikanten Einzelvergleiche zwischen den Stichproben. 1. Patienten TK-Studie, 2. Nicht-klinische Stichprobe, 3. Ambulante Stichprobe, 4. Stationäre Stichprobe.

Tabelle 124. Mittelwerte der Eingangsmessungen und deren Vergleich mit Normstichproben

	Patienten TK-Studie (1)	Nichtklinische Stichprobe (2)	Klinische Stichprobe (3)	ANOVA <sup>a</sup> F (ges) / sign. Einzelvergleiche
ACQ (Panik, Agoraphobie)	M = 2,06 SD = 0,58 N = 64	M = 1,32 SD = 0,32 N = 208 <sup>a</sup>	M = 2,19 SD = 0,63 N = 1.231 <sup>b</sup>	F = 190,38 (1-2), (2-3)
BSQ (Panik, Agoraphobie)	M = 2,80 SD = 0,66 N = 63	M = 1,65 SD = 0,51 N = 208 <sup>a</sup>	M = 2,74 SD = 0,73 N = 1.215 <sup>b</sup>	F = 217,93 (1-2), (2-3)
ACQ (andere Angststörungen)	M = 1,88 SD = 0,61 N = 156	M = 1,32 SD = 0,32 N = 208 <sup>a</sup>	M = 1,74 SD = 0,51 N = 386 <sup>b</sup>	F = 70,99 (1-2), (1-3), (2-3)
BSQ (andere Angststörungen)	M = 2,36 SD = 0,79 N = 159	M = 1,65 SD = 0,51 N = 208 <sup>a</sup>	M = 2,06 SD = 0,69 N = 376 <sup>b</sup>	F = 50,96 (1-2), (1-3), (2-3)
EDI (Anorexia Nervosa)	M = 323,8 SD = 62,3 N = 5	M = 223,2 SD = 38,1 N = 186 <sup>c</sup>	M = 297,4 SD = 53,2 N = 146 <sup>d</sup>	-
EDI (Bulimia Nervosa)	M = 350,3 SD = 44,8 N = 5	M = 223,2 SD = 38,1 N = 186 <sup>c</sup>	M = 348,1 SD = 52,3 N = 217 <sup>d</sup>	-
HZI	M = 23,37 SD = 10,91 N = 43	M = 11,60 SD = 6,40 N = 77 <sup>a</sup>	M = 18,26 SD = 9,04 N = 43 <sup>b</sup>	F = 27,69 (1-2), (1-3), (2-3)
SOMS	M = 36,13 SD = 23,83 N = 87	M = 5,10 SD = 5,00 N = 101	M = 15,00 SD = 7,90 N = 586	F = 207,73 (1-2), (1-3), (2-3)

## Anmerkungen.

<sup>a</sup> Ehlers & Margraf (2001): Ambulante oder stationäre Patienten einer psychosomatischen Fachklinik; Kontrollgruppe ist ohne psych. Störung.

<sup>b</sup> Maß et al. (1997): Gesunde Probanden, im Arbeitsumfeld der Autoren rekrutiert

<sup>c</sup> Paul & Thiel (2005): BED-Patienten ohne psychogene Essstörung.

<sup>d</sup> Rief et al. (1997): Kontrollpersonen sind unauffällige Personen, im Rahmen einer Diplomarbeit erhoben. Klinische Stichprobe entstammt einer stationären psychosomatischen Klinik.

Die Mittelwerte zu Therapiebeginn liegen durchgängig über den Cut-off-Werten für mittlere Belastung (vgl. Tabelle 125). Bei vielen Instrumenten liegt die mittlere Belastung der TK-Patienten sogar im Bereich für starke Belastung. Für den EDI bei Männern liegen nicht genügend Daten vor, um die Werte interpretieren zu können.

Tabelle 125. Belastung zu Therapiebeginn in den rückmeldungsrelevanten Instrumenten und Vergleich mit Grenzwerten

Instrument	TK-Stichprobe		Grenzwert für mittlere Belastung (Cut-Off-Wert)	Grenzwert für starke Belastung (Mittelwert klinische Stichprobe)
	N	M (SD)		
BSI	1577	1,23 (0,67)	0,52	1,14
IIP-D	1588	1,55 (0,55)	1,54	1,79
BDI	1155	20,60 (9,23)	11,90	23,10
AKV	310	2,22 (0,60)	1,81	2,31
EDI (Frauen)	45	347,03 (44,89)	299,40	337,50
EDI (Männer)	2	218,00 (12,73)	267,00	297,10
HZI	43	23,37 (10,91)	14,36	18,26
SOMS	87	36,13 (23,83)	18,69	30,80

Anmerkung. Stichprobe P1 = 1.598 Patienten mit vorhandener Erstmessung (IG: 1.011, KG: 587).

Gemäß dem Verfahren im Modellvorhaben muss bei den Patienten der Interventionsgruppe für die Bewilligung einer Psychotherapie ohne modifiziertes Gutachterverfahren neben mindestens einer ICD-10-Diagnose bei Therapiebeginn auch eine klinisch bedeutsame Beeinträchtigung in mindestens einem der maßgebenden psychometrischen Selbstbeurteilungsinstrumente (BSI, IIP-D und störungsspezifisches Instrument) vorliegen (Lutz et al., 2005; Lutz et al., 2009).

Tabelle 126 zeigt, dass insgesamt 90,9% der Patienten zu Therapiebeginn eine mittlere oder schwere klinisch bedeutsame Beeinträchtigung als Gesamtrückmeldung erhalten haben. Für Patienten in der IG bedeutete dies die Bewilligung einer Langzeittherapie ohne zusätzliches Gutachterverfahren. Für die übrigen 9,1% lag keine klinisch bedeutsame Beeinträchtigung vor; für die Patienten der IG musste in diesem Fall ein modifiziertes Gutachterverfahren eingeleitet werden. KG - Patienten durchliefen unabhängig von diesen Ergebnissen das traditionelle Gutachterverfahren.

Tabelle 126. Initiale Belastung (Häufigkeit) nach Kategorien in den maßgebenden Messinstrumenten (alle Patienten).

Instrument	N	Klinisch bedeutsame Beeinträchtigung?		
		keine oder gering	mittel	schwer
BSI	1577	215 (13,6%)	589 (37,3%)	773 (49,0%)
IIP-D	1588	739 (46,5%)	299 (18,8%)	550 (34,6%)
BDI	1155	194 (16,8%)	543 (47,0%)	418 (36,2%)
AKV	310	88 (28,4%)	85 (27,4%)	137 (44,2%)
EDI (Frauen)	45	5 (11,1%)	13 (28,9%)	27 (60,0%)
HZI	43	9 (20,9%)	9 (20,9%)	25 (58,1%)
SOMS	87	16 (18,4%)	30 (34,5%)	41 (47,1%)
Rückmeldung Eingang: BSI+IIP-D+störungsspezifisch	1226	111 (9,1%)	1115 (90,9%)	
Hypothetische Rückmeldungen:				
BSI+IIP-D	1567	178 (11,4%)	1389 (88,6%)	
BSI+störungsspezifisch	1242	108 (8,7%)	1134 (91,3%)	
IIP-D+störungsspezifisch	1248	184 (14,7%)	1064 (85,3%)	

Anmerkungen. In diesen Fallzahlen sind nur die im Rahmen des TK-Modellvorhabens vorgegeben Instrumente enthalten. Darüber hinaus von den Therapeuten vorgelegte Instrumente sind nicht aufgeführt. Dies gilt auch für die folgenden Tabellen des Berichtes zur Mesoebene. <sup>a</sup>Der Fragebogen zu Körperbezogenen Ängsten, Kognitionen und Vermeidung setzt sich in dieser Studie aus den beiden Subinstrumenten ACQ und BSQ zusammen, d.h. aus dem Mittelwert der beiden Skalen.

Neben den tatsächlich verwendeten Rückmeldungen zeigt Tabelle 126 auch den jeweiligen Anteil der Patienten, die in der Interventionsgruppe eine Therapie ohne modifiziertes Gutachterverfahren bewilligt bekommen hätten, wenn nur zwei der eigentlich drei verwendeten maßgebenden Messinstrumente zum Einsatz gekommen wären. Dabei zeigt sich, dass der IIP-D nur in wenigen Fällen eine Therapieindikation nahe legt, wenn dies nicht schon der BSI oder das störungsspezifische Instrument tun.

Schließlich zeigt Tabelle 126 auch die initialen Beeinträchtigungen in den einzelnen Instrumenten. Während übereinstimmend mit dem eben ausgeführten Befund knapp 50% der Patienten bei Therapiebeginn im IIP-D keine oder geringe Belastung angaben, berichtete beispielsweise fast 50% der Teilnehmer im BSI sogar von einer schweren Beeinträchtigung, d. h. erzielte einen Score oberhalb des Mittelwertes einer Normstichprobe von ambulanten Psychotherapiepatienten.

Vergleicht man die initiale Belastung in der Interventions- und Kontrollgruppe (Tabelle 127 und Tabelle 128), so fällt auf, dass sich die Belastung nach den störungsunspezifischen Instrumenten BSI und IIP-D in der gleichen Größenordnung bewegt. In dem störungsspezifischen Instrument AKV zeigt sich eine Tendenz zu größerer Belastung in der Kontrollgruppe, in BDI und HZI hingegen eine Tendenz zu größerer Belastung in der Interventionsgruppe.

Tabelle 127. Initiale Belastung (Häufigkeit) nach Kategorien in den maßgebenden Messinstrumenten (nur IG).

Instrument (nur Interventionsgruppe)	N	Klinisch bedeutsame Beeinträchtigung?		
		keine oder gering	mittel	schwer
BSI	1000	131 (13,1%)	354 (35,4%)	515 (51,5%)
IIP-D	1006	462 (45,9%)	182 (18,1%)	362 (36,0%)
BDI	711	98 (13,8%)	342 (48,1%)	271 (38,1%)
AKV	226	66 (29,2%)	60 (26,5%)	100 (44,2%)
EDI (Frauen)	32	3 (9,4%)	11 (34,4%)	18 (56,3%)
HZI	30	5 (16,7%)	5 (16,7%)	20 (66,7%)
SOMS	67	12 (17,9%)	22 (32,8%)	33 (49,3%)
Rückmeldung Eingang: BSI+IIP-D+störungsspezifisch	753	61 (8,1%)	692 (91,9%)	
Hypothetische Rückmeldungen:				
BSI+IIP-D	994	105 (10,6%)	889 (89,4%)	
BSI+störungsspezifisch	760	56 (7,4%)	704 (92,6%)	
IIP-D+störungsspezifisch	763	96 (12,6%)	667 (87,4%)	

Tabelle 128. Initiale Belastung (Häufigkeit) nach Kategorien in den maßgebenden Messinstrumenten (nur KG).

Instrument (nur Kontrollgruppe)	N	Klinisch bedeutsame Beeinträchtigung?		
		keine oder gering	mittel	schwer
BSI	577	84 (14,6%)	235 (40,7%)	258 (44,7%)
IIP-D	582	277 (47,6%)	117 (20,1%)	188 (32,3%)
BDI	444	96 (21,6%)	201 (45,3%)	147 (33,1%)
AKV	84	22 (26,2%)	25 (29,8%)	37 (44,0%)
EDI (Frauen)	13	2 (15,4%)	2 (15,4%)	9 (69,2%)
HZI	13	4 (30,8%)	4 (30,8%)	5 (38,5%)
SOMS	20	4 (20,0%)	8 (40,0%)	11 (40,0%)
Rückmeldung Eingang: BSI+IIP-D+störungsspezifisch	473	50 (10,6%)	423 (89,4%)	
Hypothetische Rückmeldungen:				
BSI+IIP-D	573	73 (12,7%)	500 (87,3%)	
BSI+störungsspezifisch	482	52 (10,8%)	430 (89,2%)	
IIP-D+störungsspezifisch	485	88 (18,1%)	397 (81,9%)	

Differenziert man die Rückmeldungen nach Diagnose und Instrument (Tabelle 129 und Tabelle 130), so fällt auf, dass sich die Beeinträchtigung bei affektiven Störungen im allgemeinen Instrument BSI stärker niederschlägt als im störungsspezifischen Instrument BDI. Bei den Angststörungen zeigt sich ein ähnliches Bild, auch hier ergibt sich eine größere Beeinträchtigung im BSI als im AKV. Aus dem IIP-D würde sich wiederum nur vergleichsweise selten eine Therapieindikation ergeben.

Tabelle 129. Initiale Belastung unterteilt nach Diagnose (Störungsspezifisches Instrument BDI)

Instrument	Erstdiagnose	N	Klinisch bedeutsame Beeinträchtigung?		
			keine oder gering	mittel	schwer
BSI	Depressive Episode	283	23 (8,1%)	104 (36,7%)	156 (55,1%)
	Rezidivierende depressive Störung	217	17 (7,8%)	75 (34,6%)	125 (57,6%)
	Dysthymia	87	7 (8,0%)	31 (35,6%)	49 (56,3%)
	Posttraumatische Belastungsstörung	28	3 (10,7%)	9 (32,1%)	16 (57,1%)
	Anpassungsstörung	225	39 (17,3%)	97 (43,1%)	89 (39,6%)
IIP-D	Depressive Episode	282	116 (41,1%)	65 (23,0%)	101 (35,8%)
	Rezidivierende depressive Störung	218	79 (36,2%)	41 (18,8%)	98 (45,0%)
	Dysthymia	87	31 (35,6%)	23 (26,4%)	33 (37,9%)
	Posttraumatische Belastungsstörung	30	14 (46,7%)	7 (23,3%)	9 (30,0%)
	Anpassungsstörung	227	123 (54,2%)	45 (19,8%)	59 (26,0%)
BDI	Depressive Episode	278	38 (13,7%)	121 (43,5%)	119 (42,8%)
	Rezidivierende depressive Störung	212	22 (10,4%)	97 (45,8%)	93 (43,9%)
	Dysthymia	85	13 (15,3%)	39 (45,9%)	33 (38,8%)
	Posttraumatische Belastungsstörung	29	3 (10,3%)	19 (65,5%)	7 (24,1%)
	Anpassungsstörung	219	55 (25,1%)	105 (47,9%)	59 (26,9%)

Tabelle 130. Initiale Belastung unterteilt nach Diagnose (Störungsspezifisches Instrument AKV)

Instrument	Erstdiagnose	N	Klinisch bedeutsame Beeinträchtigung?		
			keine oder gering	mittel	schwer
BSI	Agoraphobie ohne Panikstörung	20	3 (15,0%)	1 (5,0%)	16 (80,0%)
	Soziale Phobie	61	13 (21,3%)	22 (36,1%)	26 (42,6%)
	Spezifische Phobie	29	14 (48,3%)	9 (31,0%)	6 (20,7%)
	Panikstörung	46	6 (13,0%)	20 (43,5%)	20 (43,5%)
	GAD	50	6 (12,0%)	19 (38,0%)	25 (50,0%)
	Angst und depressive Störung gemischt	44	6 (13,6%)	19 (43,2%)	19 (43,2%)
IIP-D	Agoraphobie ohne Panikstörung	20	10 (50,0%)	2 (10,0%)	8 (40,0%)
	Soziale Phobie	62	30 (48,4%)	11 (17,7%)	21 (33,9%)
	Spezifische Phobie	29	23 (79,3%)	2 (6,9%)	4 (13,8%)
	Panikstörung	46	32 (69,6%)	7 (15,2%)	7 (15,2%)
	GAD	50	22 (44,0%)	8 (16,0%)	20 (40,0%)
	Angst und depressive Störung gemischt	44	27 (61,4%)	3 (6,8%)	14 (31,8%)
AKV	Agoraphobie ohne Panikstörung	19	3 (15,8%)	4 (21,1%)	12 (63,2%)
	Soziale Phobie	58	21 (36,2%)	12 (20,7%)	25 (43,1%)
	Spezifische Phobie	27	15 (55,6%)	7 (25,9%)	5 (18,5%)
	Panikstörung	44	5 (11,4%)	17 (38,6%)	22 (50,0%)
	GAD	43	11 (25,6%)	10 (23,3%)	22 (51,2%)
	Angst und depressive Störung gemischt	12	5 (41,7%)	5 (41,7%)	2 (16,7%)

## 6.3 Verlaufserhebungen und Rückmeldungen

Tabelle 131 bietet einen Überblick über die bisher vorliegenden Verlaufsmessungen in den einzelnen Instrumenten. Diese Angaben beziehen sich nur auf die IG, da in der KG keine Verlaufsmessungen vorgenommen werden. Es gibt noch eine zehnte Verlaufsmessung, diese enthält jedoch nur drei Messungen im BSI und eine Messung im IIP-D.

Tabelle 131. Anzahl der Erhebungen (IG)

	Eingangsmessung	Verlaufsmessungen								
		1.	2.	3.	4.	5.	6.	7.	8.	9.
<b>BSI</b>	1000	861	779	599	383	193	84	33	15	9
IIP-D	1006	868	778	594	378	186	78	28	12	8
BDI	711	614	542	419	251	117	41	14	4	2
AKV	226	183	161	119	81	40	19	7	6	3
EDI	32	23	21	16	10	9	4	1	-	-
HZI	30	21	19	14	9	4	3	-	-	-
SOMS	67	40	33	22	16	8	2	-	-	-

### 6.3.1 Rückmeldung zu Beginn der Therapie

Zu Beginn der Therapie wurde in der Interventionsgruppe eine Rückmeldung durch das Testsystem generiert, die entweder einen direkten Beginn der Therapie oder die Einleitung des modifizierten Gutachterverfahrens nahelegen sollte. Werden die Entscheidungsregeln zu Therapiebeginn auf Patienten der Interventions- und Kontrollgruppe angewendet (Generierung hypothetischer Rückmeldungen), so wird für insgesamt 111 Fälle (9,1%) das modifizierte Gutachterverfahren empfohlen und für 1.115 (90,9%) Fälle ist dies nicht der Fall. Für 482 Patienten (28,2%), lagen nicht genügend Angaben im HTS vor. In diese Berechnungen gehen nur Fälle ein, für die in den entsprechenden Messinstrumenten mindestens 80% der Items beantwortet wurden. Tabelle 132 zeigt die Ergebnisse noch einmal in der Zusammenschau (zu den Details vgl. auch Tabelle 126 bis Tabelle 128). Es bestehen keine Unterschiede vor der Durchführung des MVPT zwischen den beiden Gruppen in der Häufigkeit, mit der das modifizierte Gutachterverfahren empfohlen würde ( $\chi^2 = 2,15$ ;  $df = 1$ ;  $N = 1.226$ ;  $p = .140$ ).

Tabelle 132. Einleitung des modifizierten Gutachterverfahrens nach den Gruppen

	Modifiziertes Gutachterverfahren	Kein Gutachterverfahren
<b>Interventionsgruppe</b>	61 (8,1%)	692 (91,9%)
<b>Kontrollgruppe*</b>	50 (10,6%)	423 (89,4%)

Anmerkung. \*Für die Kontrollgruppe nur *hypothetische* Angabe!

Nach Anwendung dieser Regeln in der IG begannen von den Patienten mit angezeigtem modifizierten Gutachterverfahren ( $n=61$ ) nach der Begutachtung noch 25 (41,0%) eine Therapie. Bei der KG wurde das traditionelle Gutachterverfahren unabhängig von den Ergebnissen der psychometrischen Testung angewandt.

Tabelle 133 präsentiert die Ergebnisse der Entscheidungsregeln zu den ersten fünf Zwischenmessungen, für die genügend Daten vorliegen. Es werden dabei nur die Fälle berücksichtigt, für welche die jeweilige Erhebung nicht eine Abschluss- oder Katamnesemessung darstellt. Bei jeweils um 50% der Fälle in Behandlung lautet die Rückmeldung zu jedem Messzeitpunkt „bisher keine klinisch bedeutsame Veränderung“. Der Anteil der Patienten mit einer temporären Verstärkung der Problematik liegt tendenziell zwi-

schen 15% und 20%. Der Anteil der nachhaltigen Abnahme der Problematik nimmt mit zunehmender Anzahl der Messzeitpunkte leicht zu. Die folgende Tabelle zeigt das Bild für die Abschlüsse und Katamnesen.

**Tabelle 133. Gesamtrückmeldungen zu den ersten fünf Verlaufsmesszeitpunkten in der Interventionsgruppe; ohne Katamnese oder Abschluss (s. Text)**

Verlaufsmessung	N	Nachhaltige Abnahme der Problematik	Abnahme der Problematik	Bisher keine klinisch bedeutsamen Veränderungen	Temporäre Verstärkung der Problematik
1.	581	-	123 (21.2%)	310 (53.4%)	148 (25.5%)
2.	444	65 (14.6%)	68 (15.3%)	227 (51.1%)	84 (18.9%)
3.	286	49 (17.1%)	43 (15.0%)	147 (51.4%)	47 (16.4%)
4.	137	29 (21.2%)	20 (14.6%)	68 (49.6%)	20 (14.6%)
5.	56	11 (19.6%)	7 (12.5%)	28 (50,0%)	10 (17.9%)

**Anmerkung. Gesamtrückmeldung (BSI, IIP-D, störungsspezifische Instrument).** In den Fallzahlen sind nur die im Rahmen des TK-Modellvorhabens vorgegeben Instrumente enthalten. Darüber hinaus von den Therapeuten vorgelegte Instrumente sind nicht aufgeführt. Dies gilt auch für die folgenden Tabellen des Berichtes zur Mesoebene.

Zu den Therapieabschluss- und Katamnesezeitpunkten können zwischen 60% und 70% der Fälle eine (nachhaltige) Abnahme der Problematik im Vergleich zum Therapiebeginn verzeichnen. Zwischen 30% und 35% der Patienten erreichen in den gewählten Instrumenten keine Abnahme der Problematik, wobei 6,8% die Therapie mit einer Verstärkung der Problematik verlassen (Tabelle 134).

**Tabelle 134. Gesamtrückmeldungen zu den Abschlüssen und Katamnesen in der IG (im Vergleich zum Therapiebeginn)**

Messung	N	Nachhaltige Abnahme und Abnahme der Problematik	Bisher keine klinisch bedeutsamen Veränderungen	Verstärkung der Problematik
Abschluss	383	238 (62.1%)	119 (31.1%)	26 (6.8%)
Katamnesen	225	158 (70,2%)	48 (21.3%)	19 (8.4%)

### 6.3.2 Rückmeldung und Fortschritt im Therapieverlauf

Tabelle 135 zeigt die Entwicklung der Gesamtrückmeldungen von der ersten zur zweiten Verlaufsmessung. Es werden nur Fälle berücksichtigt, die vollständige Messungen zum 1. und zum 2. Verlaufsmesszeitpunkt haben und sich noch in Behandlung befinden (keine Abschlussmessung oder Katamnese; N=439). Im linken Teil der Tabelle sind die Anteile der Rückmeldungen zur ersten Verlaufsmessung aufgeführt. Für die jeweils in einer Zeile stehende Patientengruppe mit einer bestimmten Rückmeldung zu Zeitpunkt 1 wird die Art der Rückmeldung zur 2. Verlaufsmessung in der rechten Hälfte der Abbildung aufgeführt. So erhalten z.B. von den 87 Patienten mit einer Abnahme der Problematik zur ersten Zwischenmessung 65 die Rückmeldung einer „nachhaltigen Abnahme der Problematik“ sowie 18 „keine Veränderung“ zur zweiten Zwischenmessung. Bei vier Personen hat sich die Problematik verstärkt.

Tabelle 136 zeigt deutlich, wie die Effektgrößen über die Behandlungszeit und der damit verbundenen Sitzungszahl zunehmen. Der BDI zeigt schon bei der ersten Verlaufsmessung eine prägnante Effektgröße mit  $d = 0,70$  an, dieses Instrument scheint für die TK-Patienten, denen er vorgelegt wurde (Indikation Depressive Störungen) besonders veränderungssensitiv zu sein.

Tabelle 135. Entwicklung der Gesamtrückmeldungen zwischen der ersten und zweiten Verlaufsmessung in der Interventionsgruppe (ohne Katamnesen und Abschlüsse).

Erste Verlaufsmessung		Zweite Verlaufsmessung			
Rückmeldung		Nachhaltige Abnahme der Problematik	Abnahme der Problematik	Bisher keine klinisch bedeutsamen Veränderungen	Verstärkung der Problematik
Abnahme der Problematik	87 (19,8% <sup>a</sup> )	65 (74,7%)	0 (0,0%) <sup>c</sup>	18 (20,7%)	4 (4,6%)
Bisher keine klinisch bedeutsamen Veränderungen	237 (54,0% <sup>a</sup> )	0 (0,0%) <sup>b</sup>	58 (25,5%)	151 (63,7%)	28 (11,8%)
Verstärkung der Problematik	115 (26,2% <sup>a</sup> )	0 (0,0%) <sup>b</sup>	10 (8,7%)	56 (48,7%)	49 (42,6%)

Anmerkung. <sup>a</sup> Die Prozentwerte in der Spalte addieren sich zu 100. <sup>b</sup> Nach den Entscheidungsregeln nur dann vorgesehen, wenn bei der ersten Verlaufsmessung „Abnahme der Problematik“ rückgemeldet wurde. <sup>c</sup> Führt hier zur Klassifikation „Nachhaltige Abnahme der Problematik“, wenn bereits bei der ersten Verlaufsmessung „Abnahme der Problematik“ rückgemeldet wurde.

Tabelle 136. Veränderungen der Mittelwerte in den Messinstrumenten und Effektgrößen zwischen der Eingangsmessung und den Verlaufsmessungen in der Interventionsgruppe (wenn nicht Katamnese oder Abschluss)

Instrument	N	Eingangsmessung	Erste Verlaufsmessung	Effektgröße d
BSI	759	1,30 (0,68)	1,01 (0,64)	0,43
IIP-D	758	1,61 (0,56)	1,48 (0,57)	0,24
BDI	534	21,78 (9,00)	15,51 (9,37)	0,70
Instrument	N	Eingangsmessung	Zweite Verlaufsmessung	Effektgröße d
BSI	575	1,32 (0,68)	0,91 (0,66)	0,60
IIP-D	574	1,64 (0,55)	1,39 (0,60)	0,45
BDI	402	22,07 (8,92)	13,27 (8,76)	0,99
Instrument	N	Eingangsmessung	Dritte Verlaufsmessung	Effektgröße d
BSI	363	1,31 (0,69)	0,86 (0,66)	0,65
IIP-D	362	1,63 (0,51)	1,30 (0,62)	0,65
BDI	243	22,45 (9,22)	11,53 (9,56)	1,18
Instrument	N	Eingangsmessung	Vierte Verlaufsmessung	Effektgröße d
BSI	178	1,31 (0,68)	0,78 (0,56)	0,78
IIP-D	177	1,63 (0,49)	1,25 (0,58)	0,78
BDI	114	21,84 (9,00)	11,09 (9,24)	1,19
Instrument	N	Eingangsmessung	Fünfte Verlaufsmessung	Effektgröße d
BSI	75	1,39 (0,70)	0,83 (0,57)	0,80
IIP-D	75	1,70 (0,48)	1,29 (0,57)	0,85
BDI	40	23,25 (9,02)	11,58 (7,82)	1,29

Anmerkung. Die Effektgröße wurde berechnet, indem die Differenz zwischen beiden Messungen durch die Streuung der Eingangsmessung geteilt wurde.

### 6.3.3 Verlaufsprofile und Rückmeldung

Abbildung 33 zeigt die Verlaufsprofile des BSI als wichtigstes generisches Instrument für IG und KG. Verwendet wurden dafür die Daten der Stichprobe P3 (fett gedruckt) und der P4 als Angabe für den Katamnesezeitpunkt (Wechsel der Stichprobe durch Absetzung des Linienformates gekennzeichnet). Hinsichtlich der Belastung im BSI bestehen für die Stichprobe P3 zu keinem der ersten beiden (sowie in P4 für die Katamnese) Messzeitpunkte statistisch signifikante Gruppenunterschiede zwischen IG und KG. Beide Gruppen verbessern sich im Therapieverlauf deutlich mit einem großen Prä/Post-Effekt (t-Test für Messwiederholungen;  $p < .001$ ).

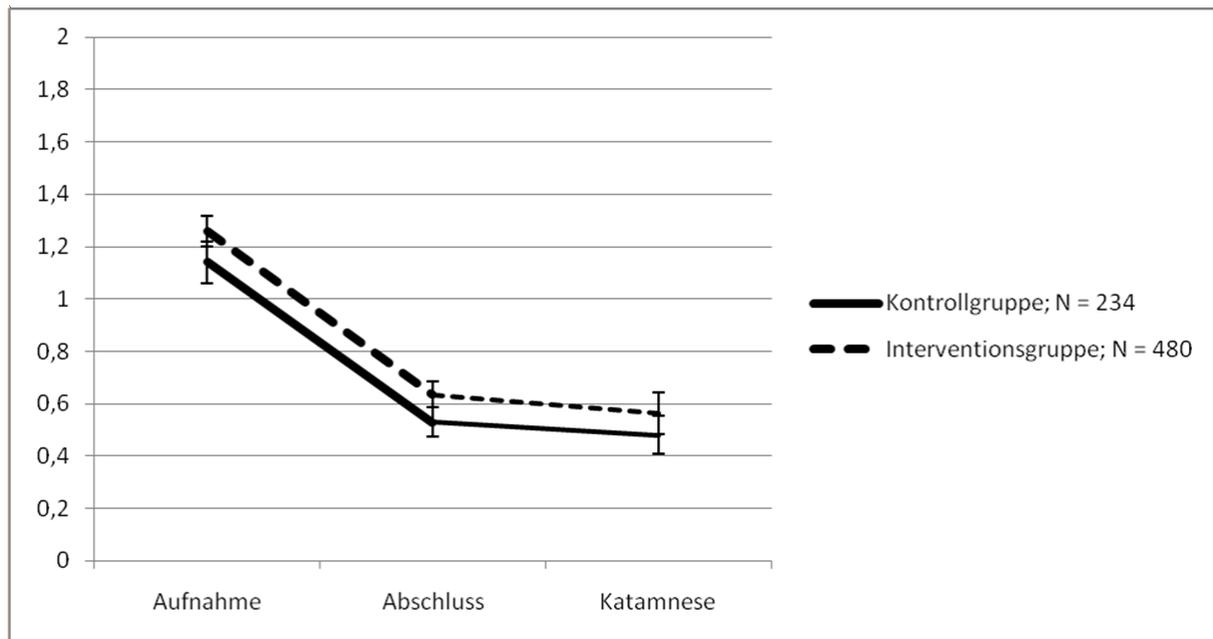


Abbildung 33. Verlaufsprofile BSI für Kontroll- und Interventionsgruppe; angegebenes N bezieht sich auf Aufnahme und Abschluss (P2); Katamnesen (P4): N = 120 in der Kontrollgruppe; N = 216 in der Interventionsgruppe; 95%-Konfidenzintervalle

Abbildung 34 zeigt die BSI-Verläufe für die Patienten der IG aufgeteilt nach den Rückmeldungen zur ersten (meist zur 10. genehmigungspflichtigen Sitzung) durchgeführten Verlaufsmessung. Für die KG wurden keine Verlaufsmessungen vorgenommen. Patienten mit Rückmeldung einer Abnahme der Problematik zum Zeitpunkt der ersten Verlaufsmessung weisen eine deutliche Symptomreduktion auf, die dann auch beibehalten wird. Dies könnte ein Ausdruck des Phänomens früher Verbesserungen in der Psychotherapie sein, da dieser Effekt zur zweiten Verlaufsmessung weniger deutlich auftritt (vgl. Abbildung 35) - der Verlauf weist hier eher auf eine kontinuierliche Besserung hin. Die Verläufe für Patienten ohne klinisch bedeutsame Veränderungen bis zur jeweiligen Zwischenerhebung zeigen im Mittel kontinuierliche Verläufe zur Besserung. Die Patienten mit temporären Verstärkungen der Problematik weisen zur zweiten Verlaufsmessung eine im Mittel stärkere Verschlechterung als zur ersten Verlaufsmessung auf ( $d = -0,12$  für die erste Verlaufsmessung;  $d = -0,31$  für die zweite Verlaufsmessung). Nach der Rückmeldung dieser temporären Verschlechterung an den Therapeuten erreichen jedoch auch diese Patienten noch eine Verbesserung gegenüber dem Eingangswert ( $d = 0,66$  für eine negative Rückmeldung zur ersten Verlaufsmessung;  $d = 0,56$  für eine negative Rückmeldung zur zweiten Verlaufsmessung). Jedoch erreichen diese Patienten nicht ganz das Abschlussniveau der Patientengruppen ohne negative Verlaufsrückmeldung.

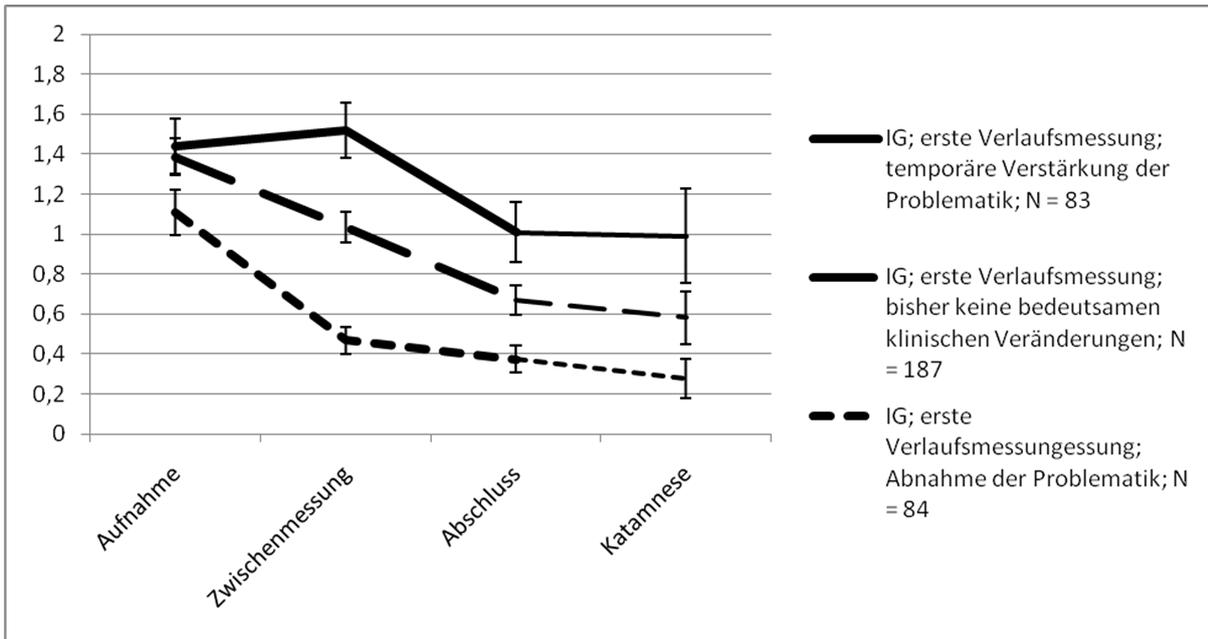


Abbildung 34. Verlaufsprofile BSI in der Interventionsgruppe für Patienten mit unterschiedlichen Rückmeldungen zur ersten Verlaufsmessung; angegebenes N bezieht sich auf Aufnahme und Abschluss (P2); Katamnesen (P4): N = 36 bei negativer Rückmeldung; N = 48 bei positiver Rückmeldung; N = 73 bei keiner Veränderung; 95%-Konfidenzintervalle.

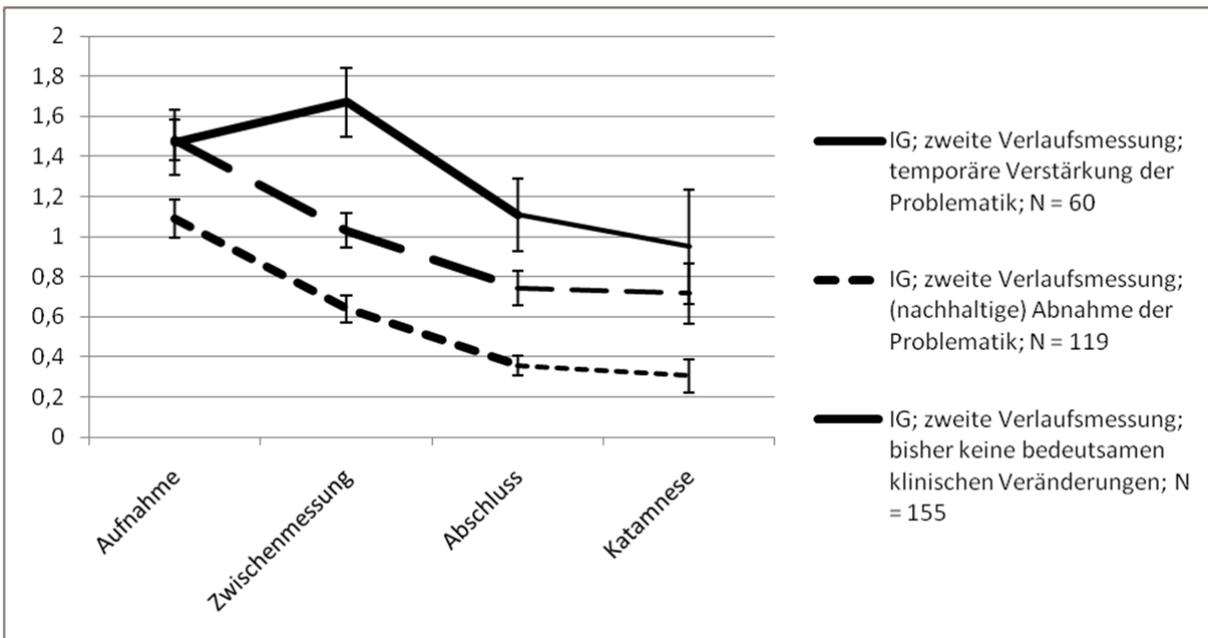


Abbildung 35. Verlaufsprofile BSI in der Interventionsgruppe für Patienten mit unterschiedlichen Rückmeldungen zur zweiten Verlaufsmessung; angegebenes N bezieht sich auf Aufnahme und Abschluss (P2); Katamnesen (P4): N = 28 bei negativer Rückmeldung; N = 67 bei positiver Rückmeldung; N = 63 bei keiner Veränderung; 95%-Konfidenzintervalle.

Die Verläufe entsprechen den bisherigen Befunden in der Forschung zu Rückmeldungen im Therapieverlauf (z.B. Lambert, 2007). Dass Patienten mit einer temporären Verstärkung der Problematik sich im Anschluss an diese Rückmeldung verbessern, kann als Effekt der Rückmeldungen gedeutet werden. Leider fehlt zum Beleg dieser Interpretation eine Kontrollgruppe, in der lediglich wiederholte Messungen ohne die Bereitstellung einer Rückmeldung durchgeführt wurden.

Abbildung 36 zeigt die mittleren Sitzungszahlen (inkl. probatorische Sitzungen) wie sie aus den abgerechneten Leistungen rekonstruiert wurden (Stichprobe P2; für alle Patienten basierend nur auf den sozioökonomischen Daten ab S. 139). Für die Interventionsgruppe ergibt sich im Mittel eine Sitzungszahl

von 42,82 Sitzungen (N = 463; Median = 43,0; 25%-Perzentil = 27,0; 75%-Perzentil = 52,0) und für die Kontrollgruppe ein Mittel von 36,09 Sitzungen (N = 216; Median = 29,0; 25%-Perzentil = 25,0; 75%-Perzentil = 48,0). Zwischen beiden besteht ein statistisch signifikanter Unterschied ( $p < .001$ ;  $d = 0,32$ ); die Therapien in der Interventionsgruppe dauern länger.

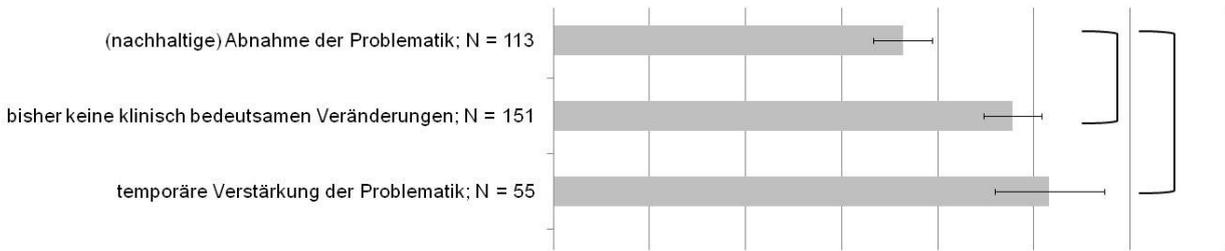
Mit Blick auf die erste Verlaufsmessung (die meist zur 10. Sitzung stattfand) zeigt sich in der Interventionsgruppe ein statistisch signifikanter Unterschied:

- Patienten, die bis zu diesem Zeitpunkt eine Abnahme der Problematik aufweisen, nehmen etwas weniger Sitzungen bis zum Therapieende in Anspruch (N = 81; M = 36,99; Median = 38,0; SD = 18,93; 25%-Perzentil = 20,0; 75%-Perzentil = 48,0)
- als Patienten, die zur ersten Verlaufsmessung keine Veränderung (N = 176; M = 43,36; Median = 43,5; SD = 20,69; 25%-Perzentil = 28,0; 75%-Perzentil = 53,0) oder eine temporäre Verschlechterung verzeichnen (N = 79; M = 48,53; Median = 48,0; SD = 18,68; 25%-Perzentil = 36,0; 75%-Perzentil = 61,0).

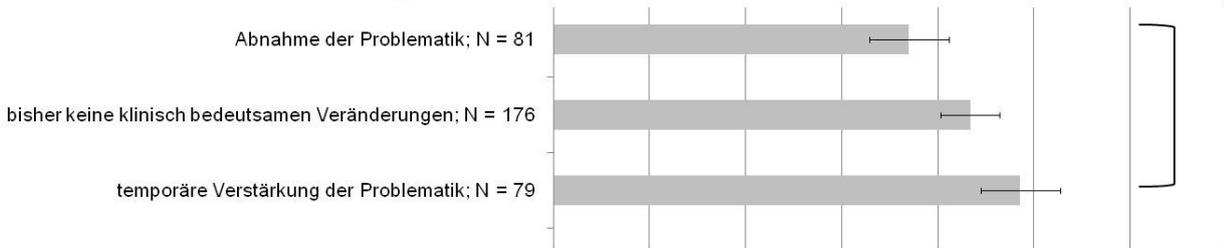
Mit Blick auf die zweite Verlaufsmessung (die meist zur 20. Sitzung stattfand) zeigt sich in der Interventionsgruppe statistisch signifikant,

- dass Patienten, die bis zu diesem Zeitpunkt eine (nachhaltige) Abnahme der Problematik aufweisen, etwas weniger Sitzungen bis zum Therapieende in Anspruch nehmen (N = 113; M = 36,35; Median = 36,0; SD = 16,67; 25%-Perzentil = 22,0; 75%-Perzentil = 47,5)
  - als Patienten, die zur zweiten Verlaufsmessung eine temporäre Verschlechterung verzeichnen (N = 55; M = 51,6; Median = 51,0; SD = 21,67; 25%-Perzentil = 41,0; 75%-Perzentil = 62,0);
  - Patienten, die keine Veränderung zu diesem Zeitpunkt gegenüber dem Therapiebeginn verzeichnen, zeigen keinen signifikanten Unterschied zu den anderen beiden Gruppen (N = 151; M = 47,81; Median = 46,0; SD = 18,91; 25%-Perzentil = 35,0; 75%-Perzentil = 58,0).
-

### Erste Verlaufsmessung - nur IG -



### Zweite Verlaufsmessung - nur IG -



### Alle Patienten

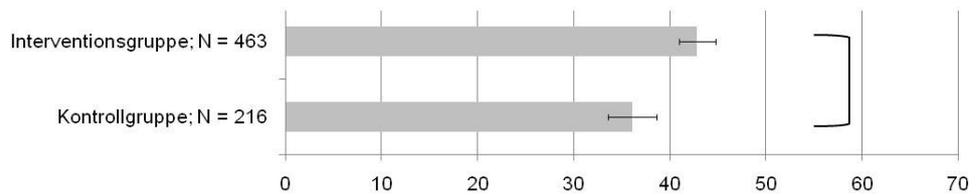


Abbildung 36. Mittlere abgerechnete Sitzungsanzahlen für die Fälle mit einer Aufnahme- und Abschlussmessung (P2); IG = Interventionsgruppe; 95%-Konfidenzintervalle

Abbildung 37 zeigt die Verlaufskurven für den IIP-D als zweites generisches Verlaufsinstrument zunächst für die IG und KG. Wieder beziehen sich die Angaben für die ersten drei Messzeitpunkte auf die P3 und die Katamnese ist mit den Daten der P4 für die entsprechenden Fälle erstellt (Kennzeichnung in der Abbildung durch Fettsetzung der P3-Linien). Beide Gruppen erreichen bis zum Therapieende eine deutliche Verbesserung mit großen Effekten, die auch über das Therapieende hinaus stabil bleibt. Abbildung 38 und Abbildung 39 geben wieder die Verläufe in Abhängigkeit von der Rückmeldung zur 1. bzw. 2. Zwischenmessung wieder. Beim IIP-D entsteht der Eindruck, dass bei einem Feedback „unverändert“ bzw. „verbessert“ zum Zeitpunkt der 1. Zwischenmessung auch nach Therapieende dann eine weitere Abnahme der zwischenmenschlichen Probleme im IIP-D erfolgt – ein Indiz dafür, dass hier auch nach Ende der Therapie neue Strategien im interpersonalen Bereich weiter erfolgreich realisiert bzw. dysfunktionale Bindungsmuster, die nach allgemeiner Auffassung ja relativ traitnah sind und daher Zeit zur Veränderung benötigen, erfolgreich modifiziert werden können. Bei den Patienten, die zur 1. bzw. 2. Verlaufsmessung eine temporäre Verstärkung ihrer zwischenmenschlichen Probleme berichten, zeigt sich danach ein ähnliches Muster wie beim BSI: Bis zum Therapieende nimmt die Problematik zwar auch ab, nach der Therapie besteht dann aber insgesamt noch eine höhere Belastung als in den beiden Gruppen mit positiver bzw. neutraler Rückmeldung in der Zwischenmessung.

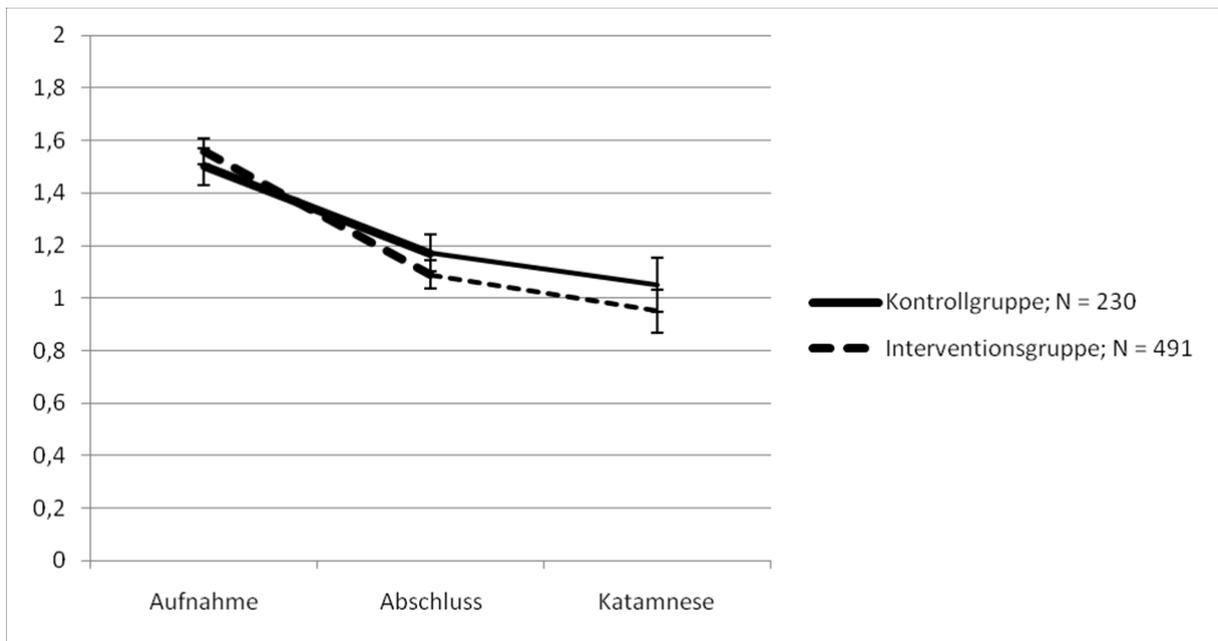


Abbildung 37. Verlaufsprofile IIP-D für Kontroll- und Interventionsgruppe; angegebenes N bezieht sich auf Aufnahme und Abschluss (P2); Katamnesen (P4): N = 117 in der Kontrollgruppe; N = 216 in der Interventionsgruppe; 95%-Konfidenzintervalle.

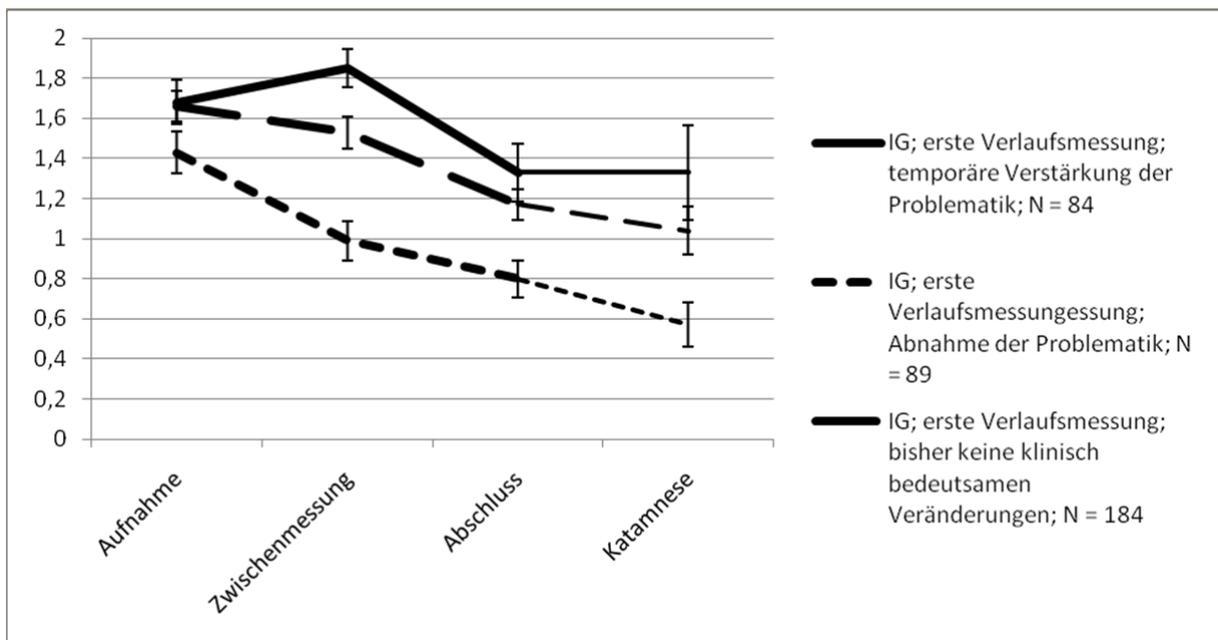


Abbildung 38. Verlaufsprofile IIP-D für in der Interventionsgruppe für Patienten mit unterschiedlichen Rückmeldungen zur ersten Verlaufsmessung; angegebenes N bezieht sich auf Aufnahme und Abschluss (P2); Katamnesen (P4): N = 36 bei negativer Rückmeldung; N = 52 bei positiver Rückmeldung; N = 73 bei keiner Veränderung; 95%-Konfidenzintervalle.

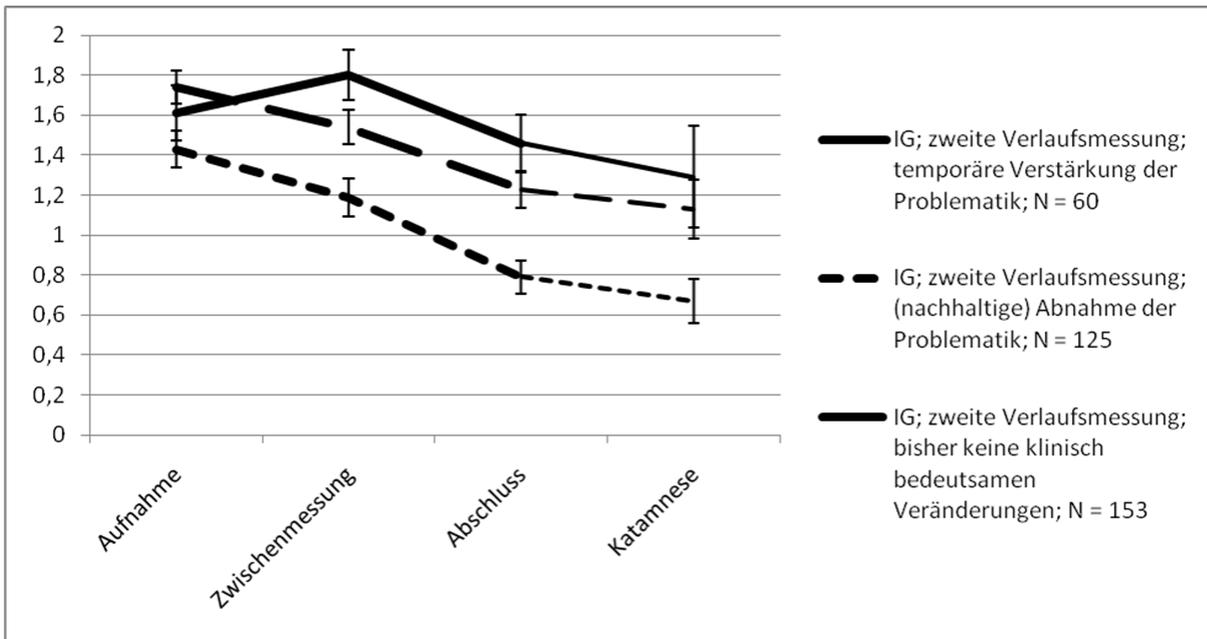


Abbildung 39. Verlaufprofile IIP-D für in der Interventionsgruppe für Patienten mit unterschiedlichen Rückmeldungen zur zweiten Verlaufsmessung; angegebenes N bezieht sich auf Aufnahme und Abschluss (P2); Katamnesen (P4): N = 28 bei negativer Rückmeldung; N = 71 bei positiver Rückmeldung; N = 63 bei keiner Veränderung; 95%-Konfidenzintervalle

Um deutlicheren Aufschluss über die Verteilungen der Therapiedauern in den Gruppen in den Gruppen zu bekommen, sind im Folgenden die Verteilungen verschiedener Kenngrößen in Histogrammen dargestellt (alle inklusive Probatorik). Es werden hierfür nur die sozio-ökonomischen Daten genutzt, d.h. es werden alle dort als abgerechnet dokumentierten Therapien berücksichtigt. Für die Information, wie viele Therapien in den jeweiligen Verfahrensgruppen stattfanden, wird auf Tabelle 22 verwiesen.

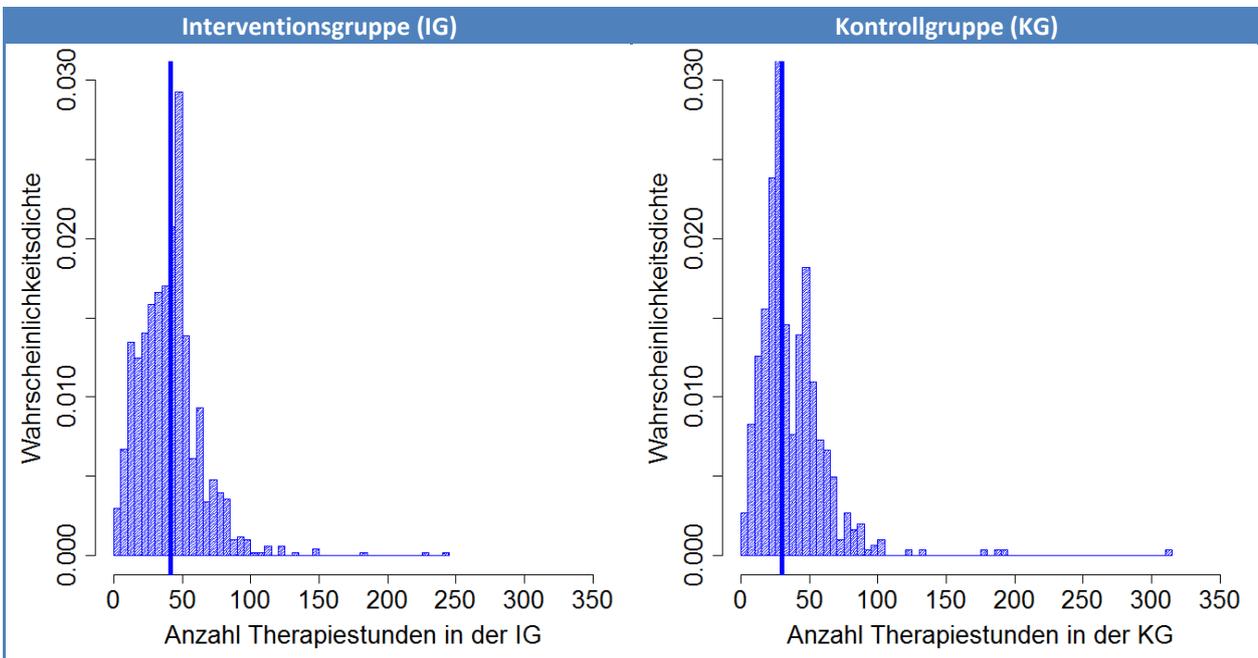


Abbildung 40. Verteilung der abgerechneten Therapiesitzungen (inklusive probatorische Sitzungen) in der Interventionsgruppe (links) und der Kontrollgruppe (rechts); die senkrechte Linie kennzeichnet den Median beider Gruppen

Die Verteilungen der abgerechneten Therapiestunden (inklusive probatorischer Sitzungen) stellen sich in den Gruppen wie folgt dar: In der Interventionsgruppe liegt der Mittelwert bei 41,29 Sitzungen (N = 1011; Median = 41,0; SD = 23,29; 25%-Perzentil = 26,0; 75%-Perzentil = 51,0); in der Kontrollgruppe liegt der Mittelwert bei 36,95 Sitzungen (N = 604; Median = 30,0; SD = 24,84; 25%-Perzentil = 23,0; 75%-Perzentil = 48,0). Der Unterschied zwischen den Gruppenmittelwerten ist statistisch signifikant ( $t = 3,47$ ;  $df = 1204,41$ ;  $p = ,001$ ;  $d = ,17$ ). Es kann kein Unterschied zwischen den Streuungen der Gruppen festgestellt werden (Levene's Test auf Varianzgleichheit:  $F = ,02$ ;  $p = ,89$ )

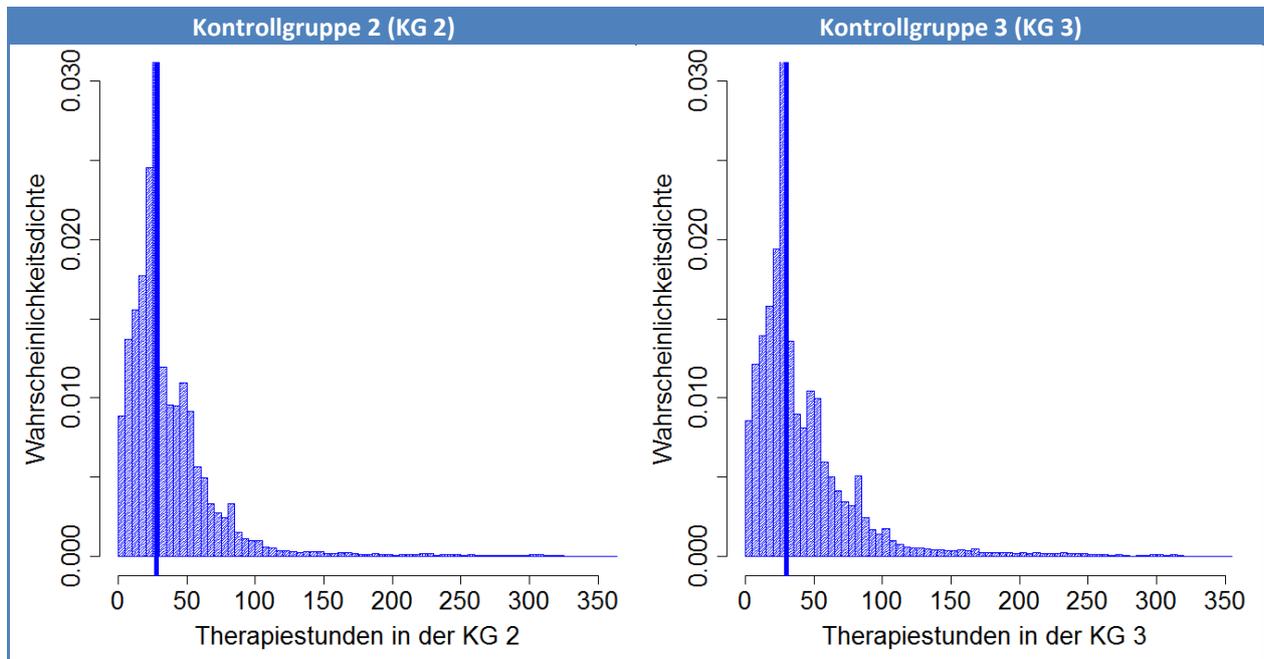


Abbildung 41. Verteilung der abgerechneten Therapiestunden (inklusive probatorische Sitzungen) in der Kontrollgruppe 2 (links) und der Kontrollgruppe 3 (rechts); die senkrechte Linie kennzeichnet den Median beider Gruppen

Die Verteilung der abgerechneten Sitzungen für die beiden Kontrollgruppen der nichtteilnehmenden Therapeuten in den Modellregionen (KG 2) und der nicht im Bereich der teilnehmenden KVen Versicherten (KG 3) ist in Abbildung 41 dargestellt. In der Kontrollgruppe 2 liegt der Mittelwert bei 37,68 Sitzungen (N = 31543; Median = 28,0; SD = 34,76; 25%-Perzentil = 19,0; 75%-Perzentil = 47,0); in der Kontrollgruppe 3 liegt der Mittelwert bei 42,78 Sitzungen (N = 25915; Median = 30,0; SD = 39,72; 25%-Perzentil = 20,0; 75%-Perzentil = 53,0).

Zum Vergleich der Anzahl bewilligter Therapiestunden standen nur Daten bis einschl. Dezember 2009 zur Verfügung (Abbildung 42). In der Interventionsgruppe wurden gemäß der modifizierten Vorgehensweise im Modellvorhaben von vorneherein 45 (VT), 50 (TP) bzw. 155 Stunden (AT) genehmigt. In der Kontrollgruppe wurden die Therapiebewilligungen und das Antragsverfahren nach den üblichen Bedingungen durchgeführt. Dementsprechend liegt der Median der Interventionsgruppe bei 45 Sitzungen (N = 941; M = 51,0; SD = 19,1; 25%-Perzentil = 45; 75%-Perzentil = 50). Für die Kontrollgruppe liegt der Median bei 25 Sitzungen (N = 612; M = 36,8; SD = 22,4; 25%-Perzentil = 25; 75%-Perzentil = 45).

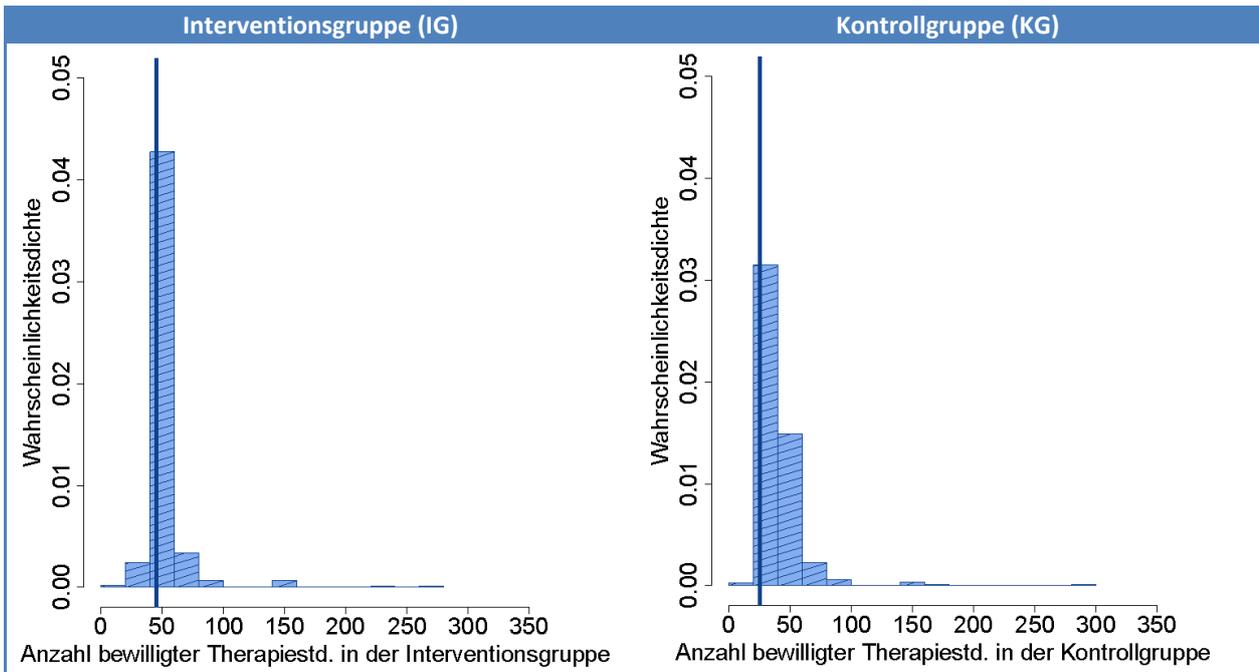
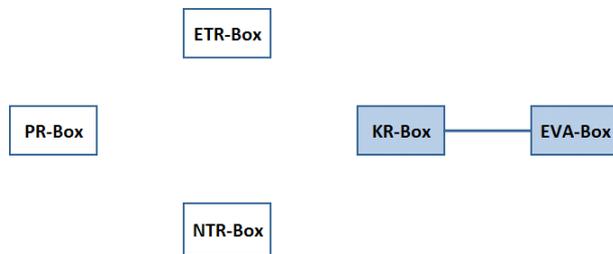


Abbildung 42. Verteilung der Anzahl bewilligter Therapiesitzungen in der Interventionsgruppe (links) und der Kontrollgruppe (rechts); die senkrechte Linie kennzeichnet den Median beider Gruppen (Stand: Dezember 2009)

### 6.3.4 Modifiziertes Gutachterverfahren im Therapieprozess und im Verlauf

Die in Abschnitt 6.3.3 dargestellten Ergebnisse beziehen sich auf die Fälle, für die nicht das modifizierte Gutachterverfahren eingeleitet wurde. Die Zahl der Fälle, für die ein modifiziertes Gutachterverfahren angezeigt war, die eine Therapie begonnen haben und für die auch eine Abschlussmessung vorliegt, ist gering. Dies trifft auf  $N = 25$  im BSI ( $N = 25$  im IIP-D) Personen in der Interventionsgruppe zu. Die mittlere Ausgangsbelastung dieser Patientinnen und Patienten liegt bei  $M = 0,33$  im BSI ( $SD = 0,13$ ; im IIP-D:  $M = 0,89$ ;  $SD = 0,38$ ). Die Prä-Post Effektgrößen für diese Patienten lagen in der Interventionsgruppe für den BSI bei  $d = 0,85$  und für den IIP-D bei  $d = 0,45$ . Die beobachteten Veränderungen zeigen zunächst, dass auch für diese Gruppe im Mittel eine Verbesserung beobachtet werden konnte. Dies ist jedoch insofern zu relativieren, als dass die Fallzahl hier sehr gering ist und die Größe der Effektstärken in Zusammenhang mit der eingeschränkten Ausgangsvarianz steht (die Fälle wurden durch diese Instrumente selektiert). Eine weitere Perspektive ist die Frage der klinischen Relevanz dieser Veränderungen, da keiner dieser Patientinnen und Patienten zu Beginn der Therapie oberhalb der Cut Off-Werte der Instrumente liegen konnte (sonst wäre nicht das modifizierte Gutachterverfahren empfohlen worden).

## 7 Praktikabilität und Akzeptanz des Modellvorhabens (Mikroebene)



Wie bewerten die drei unmittelbar an der Umsetzung des Modellvorhabens beteiligten Stakeholdergruppen

- Therapeuten
- Patienten
- Gutachter

die Realisierung des Modellvorhabens und die eingesetzten Instrumente zur Qualitätssicherung insgesamt? Dieser wichtige Aspekt ist nach (Rossi, Freeman, & Lipsey, 1999) der formativen Evaluation zuzuordnen und liefert wichtige Informationen darüber, ob ein Programm auch wie geplant umgesetzt werden konnte und wie es von den Beteiligten akzeptiert wurde. Bestehen hier Defizite, so kann man von einer Gefährdung der internen Validität aufgrund mangelnder Umsetzung der eigentlich für die IG geplanten Vorgaben im Sinne eines wirksam alternativen Vorgehens im Vergleich zum „Business as usual“ (KG) ausgehen. Dies würde sich in einem solchen Fall wiederum mindernd auf die bei der summativen Evaluation gemessenen Unterschiede zwischen IG und KG auswirken. Dort wo mögliche Defizite aufgedeckt werden, können für künftige ähnliche Vorhaben darüber hinaus Vorkehrungen bereits bei der Planung und Umsetzung implementiert werden.

### 7.1 Evaluation aus Therapeutensicht

Nachfolgend sind die Ergebnisse der Evaluationsfragen wiedergegeben, die nur an die Therapeuten der IG gerichtet und für jeden Patienten gesondert beantwortet wurden. Zunächst werden die Ergebnisse deskriptiv beschrieben und sodann gemäß Abschnitt 2.4.1.3 d im Evaluationsplan in Bezug zur Anzahl der durchgeführten Therapien gesetzt und entsprechende Zusammenhangsanalysen durchgeführt.

#### 7.1.1 Praktikabilitätseinschätzung durch die Therapeuten

Tabelle 137 gibt die Ergebnisse von 12 Einzelfragen wieder, die zu einer Gesamtskala „Bewertung des TK-Modells aus Therapeutensicht“ zusammengefasst wurden (N = 539; M = 2,42; SD = 0,58; Cronbachs Alpha = .81). Die grafischen Rückmeldungen stimmen in der Regel gut mit dem klinischen Eindruck überein, haben jedoch nicht unbedingt einen Einfluss auf die Behandlungsstrategie. Negative Auswirkungen wurden im Hinblick auf die therapeutische Beziehung lediglich in 4,1% aller Fälle bzw. den Behandlungsverlauf lediglich in 3,1% aller Fälle überhaupt geäußert, diese Befürchtung wurde im Vorfeld der Studie ja häufig genannt und erwies sich somit als unbegründet. Verbesserungsbedarf besteht noch beim zeitlichen Aufwand für die Informationsgewinnung, so gaben 54,0% aller Therapeuten an, dass der zeitliche Aufwand für das TK-Modell im Verhältnis zum Informationsgewinn zu groß war.

Tabelle 137. Evaluationsfragen Therapeutensicht zum TK-Modell (nur IG)

Beurteilungsaspekt	1	2	3	4	5	N	M	SD
Ich fand die Datenerhebung und -handhabung mit PC angenehmer als die sonst übliche Bearbeitung in Papierform <sup>a</sup>	172 42,5%	171 42,2%	53 13,1%	9 2,2%	0 0,0%	405 100,0%	1,75	0,76
Das Hogrefe TestSystem hat während des gesamten Verlaufes gut funktioniert <sup>a</sup>	204 48,6%	147 35,0%	29 6,9%	37 8,8%	3 0,7%	420 100,0%	1,78	0,96
Die grafischen Rückmeldungen des TK-Reports stimmen gut mit meinem klinischen Eindruck bei dieser Patientin/ diesem Patienten überein	103 19,0%	304 56,1%	71 13,1%	53 9,8%	11 2,0%	542 100,0%	2,20	0,93
Die Handlungsempfehlungen des TK-Reports stimmen gut mit meinem klinischen Eindruck bei dieser Patientin / diesem Patienten überein	66 12,6%	221 42,3%	156 29,9%	79 15,1%	0 0,0%	522 100,0%	2,48	0,90
Die diagnostischen Bausteine (SKID-Screening und IDCL) waren für die Behandlung dieser Patientin/ dieses Patienten hilfreich	40 7,4%	226 41,5%	183 33,6%	79 14,5%	16 2,9%	544 100,0%	2,64	0,92
<u>Nur für die psychodynamischen Verfahren:</u> Die Durchführung von OPD und Heidelberger Umstrukturierungsskala war für die Behandlung dieser Patientin/ dieses Patienten hilfreich	36 20,1%	68 38,0%	50 27,9%	23 12,8%	2 1,1%	179 100,0%	2,37	0,98
Die grafischen Rückmeldungen und Handlungsempfehlungen waren hilfreich für die Behandlung dieser Patientin/ dieses Patienten	22 4,8%	186 41,0%	136 30,0%	89 19,6%	21 4,6%	454 100,0%	2,78	0,97
Die grafischen Reports und Handlungsempfehlungen haben einen Einfluss auf die Behandlungsstrategie bei dieser Patientin/ diesem Patienten gehabt	4 0,8%	90 18,9%	167 35,3%	121 25,5%	93 19,6%	475 100,0%	3,44	1,03
Der zeitliche Aufwand zur Durchführung des TK-Modells bei dieser Patientin/ diesem Patienten steht in einem angemessenen Verhältnis zum Informationsgewinn.	26 5,3%	194 39,7%	87 17,8%	130 26,6%	52 10,6%	489 100,0%	2,98	1,14
Die grafischen Rückmeldungen und die Handlungsempfehlungen haben sich negativ auf die Beziehung zur Patientin/ zum Patienten ausgewirkt <sup>b</sup>	2 0,4%	18 3,7%	77 16,0%	97 20,2%	287 59,7%	481 100,0%	1,65	0,91
Die grafischen Rückmeldungen und die Handlungsempfehlungen haben sich negativ auf den Behandlungsverlauf ausgewirkt <sup>b</sup>	0 0,0%	15 3,1%	84 17,5%	89 18,5%	292 60,8%	480 100,0%	1,63	0,88
Der Aufwand zum TK-Modell war im Verhältnis zum Nutzen zu groß <sup>b</sup>	62 12,8%	151 31,2%	102 21,1%	109 22,5%	60 12,4%	484 100,0%	3,10	1,24

Anmerkungen. Codierung: 1 – trifft voll zu, 2 – trifft eher zu, 3 – weder / noch, 4 – trifft eher nicht zu, 5 – trifft überhaupt nicht zu. <sup>a</sup> Nur bei selbstadministrierter Testung. <sup>b</sup>Item wurde vor Berechnung des Mittelwertes umgepolt nach der Beziehung  $y = 6 - x$ .

### 7.1.2 Bewertung der eingesetzten Assessment-Instrumente durch die Therapeuten

Es liegen zu 501 beendeten IG-Therapien Informationen zur Nutzung des Rückmeldesystems vor. Bei den durch die Patienten ausgefüllten psychometrischen Testverfahren zur Selbstbeurteilung wurde der störungsspezifische Test mit 50,1% als besonders wichtig eingestuft, gefolgt vom BSI mit 47,9% (Tabelle 138). Weniger wichtig ist den Teilnehmern offenbar der SF-12 mit 14,6%.

Bei den wichtigsten Verfahren zur Fremdbeurteilung führt der Fragebogen zur Eingangsdiagnostik die Liste mit 68,5% klar an, gefolgt von der Basisdokumentation mit 39,5% (Tabelle 139). Klares Schlusslicht bildet der Fragebogen zur Psychopharmaka-Behandlung mit lediglich 3,2%. Möglicherweise hätte es hier also genügt, einfach die momentan eingenommenen Medikamentengruppen abzufragen und auf eine weitergehende Differenzierung für die Routinedokumentation zu verzichten.

**Tabelle 138. Welche psychodiagnostischen Tests waren am wichtigsten für Sie (nur IG)?**

<i>Mehrere Angaben sind möglich</i>	IG						
	N	%					
Störungsspezifischer Test	254	50,1					
BSI	240	47,9					
IIP-D	182	36,3					
HAQ	110	22,0					
SF-12	73	14,6					
Gesamt	501	100,0					

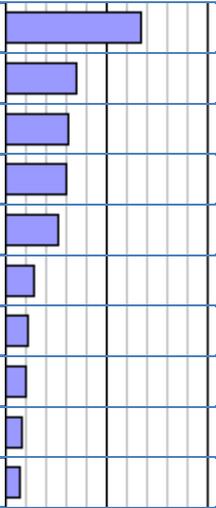
**Tabelle 139. Welche Elemente der Fremddokumentation waren am wichtigsten für Sie (nur IG)?**

<i>Mehrere Angaben sind möglich</i>	IG						
	N	%					
Eingangsdiagnostik	343	68,5					
Basisdokumentation	198	39,5					
Therapeuten-Ratings	150	30,0					
Abschlussdokumentation	92	18,4					
Psychopharmaka-Behandlung	16	3,2					
Gesamt	501	100,0					

### 7.1.3 Nutzung der Rückmeldungen durch die Therapeuten

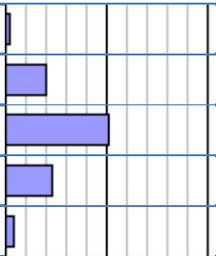
Weiter wurde erfasst, was die Therapeuten mit den Rückmeldungen durch das TK-Modell angefangen haben (Tabelle 140). Ausschließlich das Item „überhaupt keine Veränderungen in der Behandlung vorgenommen“ wurde in 149 Fällen angekreuzt, was einer Quote von 29,7% entspricht. In immerhin zwei von fünf Fällen wurden die Fragebogenergebnisse mit den Patienten besprochen. Bei jedem fünften Patienten wurde versucht, die diagnostischen Informationen anzupassen bzw. die Ressourcen des Patienten mehr zu fördern (Mehrfachangaben waren möglich).

**Tabelle 140. Nutzung der Rückmeldungen durch die Therapeuten**

Ich habe aufgrund der Rückmeldungen...	IG		
	N	%	
<i>Mehrere Angaben sind möglich</i>			
... die Ergebnisse in den Fragebögen mit dem Patienten besprochen	234	66,5	
... versucht, die Ressourcen des Patienten zu fördern	123	34,9	
... versucht meine therapeutischen Informationen anzupassen	110	31,3	
... die interpersonalen Probleme des Patienten besprochen	106	30,1	
... das Therapieende vorbereitet	92	26,1	
... versucht, die Therapiemotivation des Patienten zu erhöhen	49	13,9	
... die Sitzungsintervalle variiert	38	10,8	
... versucht, die therapeutische Beziehung zu verbessern	35	9,9	
... zusätzliche Hilfe hinzugezogen (Supervision, Literatur, Fortbildung)	28	8,0	
... neue Hausaufgaben erprobt	23	6,5	
Gesamt	352	100,0	

Abschließend wurde im Evaluationsfragebogen auch die Gesamtzufriedenheit mit dem Modellvorhaben erfragt (Tabelle 141). Es wird deutlich, dass die Zufriedenheit insgesamt geringer ausfällt als bei den Patienten, immerhin äußern sich aber  $\frac{3}{4}$  aller Befragten als zufrieden bis vollkommen zufrieden mit dem Modellvorhaben. Bei den Patienten beträgt dieser Anteil mehr als 90% (vgl. Tabelle 143).

**Tabelle 141. Zufriedenheit mit dem TK-Modellvorhaben im Hinblick auf diesen Patienten insgesamt...**

	IG		
	N	%	
<i>Mehrere Angaben sind möglich</i>			
Vollkommen zufrieden	10	2,1	
Sehr zufrieden	99	20,4	
Zufrieden	246	50,7	
Weniger zufrieden	109	22,5	
Unzufrieden	21	4,3	
Gesamt	485	100,0	

### 7.1.4 Praktikabilitätseinschätzung durch die Therapeuten und Anzahl der Therapien

Die Anzahl der durchgeführten Therapien pro Therapeut ist in der IG höher als in der KG, wie durch Abbildung 27 veranschaulicht wird (t-Test:  $p < .001$ ;  $d = 0,67$ ). Es wurden hier nur Therapien berücksichtigt, für die eine gültige Abschlussdokumentation durch die Therapeuten vorliegt (Stichprobe T2). Für die nachfolgenden Auswertungen ist natürlich nur die IG relevant, da nur diese die Therapie nach dem TK-Modell durchgeführt und den Evaluationsfragebogen ausgefüllt hat.

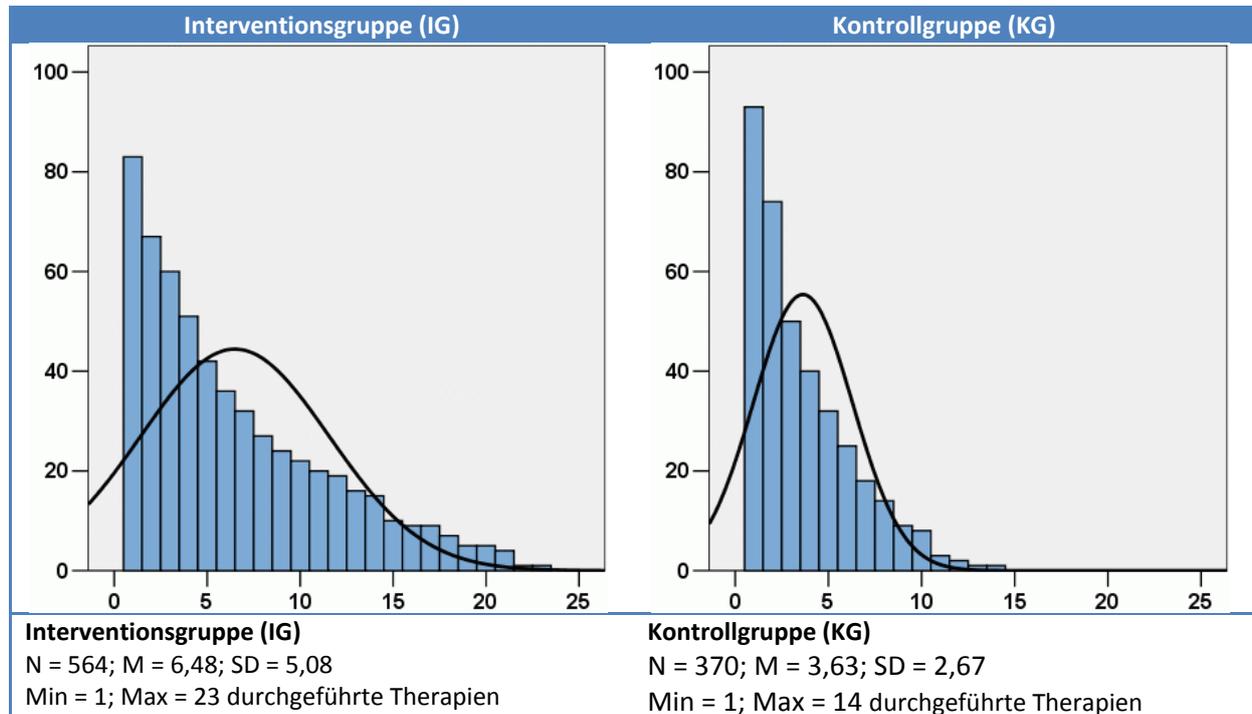


Abbildung 43. Anzahl der durchgeführten Therapien mit Abschlussdokumentation

Setzt man die Anzahl der durchgeführten Therapien in der IG in Beziehung mit der Praktikabilitätseinschätzung des Modellvorhabens (vgl. Tabelle 137), so ergeben sich uneinheitliche Ergebnisse. So beträgt die Produkt-Moment-Korrelation zwischen der explorativ gebildeten Gesamtskala zu den Evaluationsfragen (1=sehr gut... 5 = sehr schlecht) aus Therapeutensicht und der Anzahl der durchgeführten Therapien  $r = -.04$  ( $N = 538$ ;  $p = .319$ ). Bei mehr abgeschlossenen Therapien fällt die Evaluation des Modellvorhabens durch die Therapeuten also nicht besser oder schlechter aus.

Bei lediglich zwei Items ergeben sich signifikante Korrelationen, die angesichts des nicht signifikanten Ergebnisses der Summenskala allerdings nicht überbewertet werden sollten:

Je mehr Therapien durchgeführt wurden, desto eher...

- ...hatte die computergestützte Testung gut funktioniert ( $r = -.18$ ;  $p < .001$ ;  $N = 419$ )
- ...wurde die OPD und Heidelberger Umstrukturierungsskala als hilfreich erlebt ( $r = .15$ ;  $p = .047$ ;  $N = 178$ )

Bei der 1-Item-Frage zur Gesamtzufriedenheit mit dem Modellvorhaben beträgt der Zusammenhang mit der Anzahl der dokumentierten Therapien  $r = -.11$  ( $N = 485$ ;  $p = .019$ ; kleinere Werte stehen für eine höhere Zufriedenheit) und ist somit ebenfalls statistisch signifikant. Dies bedeutet, dass mit jeder als abgeschlossen dokumentierten Therapie in der IG die Gesamtzufriedenheit mit dem Modellvorhaben etwas steigt. Rechnet man  $r$  in  $d$  um, so ist die Effektgröße allerdings klein mit  $d = 0,22$ .

## 7.2 Evaluation aus Patientensicht

Insgesamt lagen Evaluationen von 601 Patienten mit einer abgeschlossenen Therapie vor. Neben der Evaluation des Modellvorhabens selbst wurde auch eine direkte Veränderungseinschätzung des Befindens durch die Patienten (BESS-Skala) bzw. die Patientenzufriedenheit mit der Behandlung (ZUF-8-Skala) der Patienten erfragt.

### 7.2.1 Bewertung des Modellvorhabens durch die Patienten

Wie Tabelle 142 zeigt, bewerteten die Patienten der IG das TK-Modellvorhaben „Qualitätssicherung in der ambulanten Psychotherapie“ als vorwiegend sehr positiv und der dadurch anfallende Aufwand wurde zum größten Teil als akzeptabel empfunden. Bildet man aus allen sieben Items explorativ eine Gesamtskala, so hat diese einen Mittelwert von 1,53 (SD = 0,42), was im Durchschnitt einer sehr guten bis guten Bewertung entspricht (N = 596 Patienten mit mindestens 6 beantworteten Items; Cronbachs Alpha = .64).

Tabelle 142. Evaluation des TK-Modellvorhabens aus Patientensicht (nur IG).

Beurteilungsaspekt	1	2	3	4	5	N	M	SD
Ich finde die Idee gut, ein Projekt zum Qualitätsmonitoring in der ambulanten Psychotherapie durchzuführen	374 62,6%	177 29,6%	41 6,9%	3 0,5%	2 0,3%	597 100,0%	1,46	0,68
Ich finde es wichtig, dass psychotherapeutische Behandlungsergebnisse überprüft werden	399 66,8%	156 26,1%	30 5,0%	8 1,3%	4 0,7%	597 100,0%	1,43	0,71
Der Zeitaufwand, den ich für die Beantwortung der Fragen eingesetzt habe, war vertretbar	389 65,2%	181 30,3%	14 2,3%	12 2,0%	1 0,2%	597 100,0%	1,42	0,66
Die Fragen waren verständlich	384 64,4%	181 30,3%	14 2,3%	12 2,0%	1 0,2%	596 100,0%	1,42	0,65
Ich hatte Mühe, die Fragen zu beantworten, da sie mir zu nahe gingen <sup>a</sup>	21 3,5%	32 5,4%	50 8,4%	159 26,8%	332 55,9%	594 100,0%	1,74	1,05
Die Fragen haben alle Themen und Problembereiche abgedeckt, die mir wichtig erschienen	154 26,0%	302 41,0%	84 14,2%	44 7,4%	8 1,4%	592 100,0%	2,07	0,90
Ich habe die Fragen immer spontan und ehrlich beantwortet	492 82,4%	100 16,8%	2 0,3%	3 0,5%	0 0,0%	597 100,0%	1,19	0,44

Anmerkungen. Codierung: 1 – trifft voll zu, 2 – trifft eher zu, 3 – weder / noch, 4 – trifft eher nicht zu, 5 – trifft überhaupt nicht zu. <sup>a</sup>Item wurde vor Berechnung des Mittelwertes umgepolt nach der Beziehung  $y = 6 - x$ .

Insgesamt 96,2% aller befragten Patienten äußerten sich als zufrieden bis vollkommen zufrieden mit dem Modellvorhaben insgesamt (Tabelle 143). Statistisch signifikante Unterschiede zwischen IG (N = 276) und KG (N = 221) bestehen dabei nicht.

Tabelle 143. Patientenzufriedenheit mit dem Modellvorhaben insgesamt

Wie zufrieden sind Sie mit dem Modellvorhaben „Qualitätsmonitoring in der ambulanten Psychotherapie“ insgesamt?	IG und KG							
	N	%						
Vollkommen zufrieden	112	19,6						
Sehr zufrieden	234	40,8						
Zufrieden	205	35,8						
Weniger zufrieden	18	3,1						
Unzufrieden	4	0,7						
Gesamt	573	100,0						

Tabelle 144. Evaluation der Rückmeldungen durch die Patienten (nur IG)

Beurteilungsaspekt	1	2	3	4	5	N	M	SD
Mein Therapeut/meine Therapeutin hat mir die grafischen Rückmeldungen und Reports zu den Fragebögen gezeigt und mit mir besprochen	182 66,2%	46 16,7%	10 3,6%	10 3,6%	27 9,8%	275 100,0%	1,74	1,29
Die Informationen waren sehr wichtig für mich	85 32,9%	105 40,7%	35 13,6%	13 5,0%	20 7,8%	258 100,0%	2,14	1,16
Die Informationen waren gut verständlich	118 45,9%	94 36,6%	25 9,7%	8 3,1%	12 4,7%	257 100,0%	1,84	1,04
Der Informationsumfang war angemessen	97 37,7%	99 38,5%	41 16,0%	8 3,1%	12 4,7%	257 100,0%	1,98	1,04
Die Rückmeldungen waren sehr hilfreich für mich	73 28,3%	98 38,0%	52 20,2%	16 6,2%	19 7,4%	258 100,0%	2,26	1,15
Die Rückmeldungen haben mit meiner eigenen Einschätzung übereingestimmt	62 24,1%	119 46,3%	44 17,1%	18 7,0%	14 5,4%	257 100,0%	2,23	1,06
Die Atmosphäre bzw. die Arbeitsbeziehung zwischen meinem Therapeuten /meiner Therapeutin und mir hat sich durch die Rückmeldungen zum Positiven verändert	44 17,4%	59 23,3	103 40,4%	17 6,7%	30 11,9%	253 100,0%	2,72	1,18
Ich hatte das Gefühl, durch die Rückmeldungen meinen Therapeuten/meiner Therapeutin gegenüber eine stärkere Rolle einzunehmen	27 10,8%	39 15,5%	114 45,4%	33 13,1%	38 15,1%	251 100,0%	3,06	1,15
Ich hatte das Gefühl, durch die Rückmeldungen selbstverantwortlicher mit meiner Erkrankung umgehen zu können	46 18,0%	75 29,3%	72 28,1%	26 10,2%	37 14,5%	256 100,0%	2,74	1,28
Es kam vor, dass ich Fragen anders beantwortet habe, wenn es um die Weiterbewilligung meiner Therapie ging <sup>a</sup>	2 0,8%	7 2,8%	44 17,9%	27 11,0%	166 67,5%	246 100,0%	1,59	0,93

Anmerkungen. Codierung: 1 – trifft voll zu, 2 – trifft eher zu, 3 – weder / noch, 4 – trifft eher nicht zu, 5 – trifft überhaupt nicht zu. <sup>a</sup>Item wurde vor Berechnung des Mittelwertes umgepolt nach der Beziehung  $y = 6 - x$ .

Den Therapeuten war es frei gestellt, die Rückmeldungen zu den Eingangs- und Verlaufsmessungen mit den Patienten zu besprechen. Tabelle 144 zeigt die Bewertung der Rückmeldungen durch die Patienten. Dieser Fragenkomplex wurde nur durch die Patienten der IG beantwortet, da nur hier die Rückmeldungen gemäß Studienplan Verwendung fanden, woraus sich die kleinere Fallzahl erklärt. Die Therapeuten nahmen sich nach Angaben der Patienten im Durchschnitt fünf Minuten Zeit, um die Rückmeldungen mit ihnen zu besprechen (Range = 0 bis 60; M = 4,62; SD = 8.67) und der überwiegende Teil der Therapeuten hat laut Patienten von dieser Möglichkeit Gebrauch gemacht.

Die Patienten bewerteten die Rückmeldungen überwiegend als nützlich, verständlich und hilfreich, den Einfluss auf die therapeutische Beziehung und ihr Patientenverhalten bewerteten sie eher neutral. Dies ist besonders relevant in Bezug auf die Mesoebene, da die Rückmeldungen einen wichtigen Teil der Interventionen in der IG darstellen. Die insgesamt positive Evaluation der Rückmeldungen zeigt sich auch in der explorativ berechneten Gesamtskala (8 von 10 Items mussten beantwortet sein; N = 256; M = 2,20; SD = 0,80; Cronbachs Alpha = .90).

Für weitere Fragen, Kritik oder Anregungen der Patienten zu den Rückmeldungen lagen 51 Angaben vor. Die größte Gruppe von Angaben machte dabei Kritik (27,5%) aus, welche sich vor allem auf eine wahrgenommene Irrelevanz für die eigentliche Therapie und mangelnde Verständlichkeit bezogen. Die zweitgrößte Gruppe (17,6%) lobte hingegen den Einsatz von Rückmeldungen aufgrund ihres Informationsgehalts und die Möglichkeit, die psychometrischen Auswertungen mit dem Therapeuten zu besprechen. Schließlich gab ein weiterer Teil Vorschläge zur Verbesserung an (15,7%), z.B. zur grafischen Gestaltung oder zur häufigeren Durchführung von Erhebungen.

Zum Modellvorhaben allgemein machten 128 Patienten Anmerkungen. Diese bezogen sich größtenteils auf die Formulierung von Fragebogenitems (21,1%), etwa dass die Fragen zu allgemein, doppelt oder auf ein unklares Zeitfenster bezogen gestellt seien. Kritik am Projekt (7,9%) beinhaltete überwiegend einen erhöhten Zeitaufwand durch das Ausfüllen der Fragebögen. Lob (6,3%) richtete sich auf die Rückmeldungen als eine den Therapieprozess begleitende Information für die Patienten. Alle Angaben im Original sind im Anhang aufgeführt.

## 7.2.2 Akzeptanz der ambulanten Therapie durch die Patienten

Die in Tabelle 145 wiedergegebenen Aspekte wurden sowohl durch die Patienten der IG als auch durch die Patienten der KG evaluiert. Bei der Summenskala des ZUF-8, der rechnerisch Werte zwischen 8 und 32 Punkte annehmen kann, zeigt sich ein statistisch signifikanter Unterschied zwischen der IG und KG in einer Größenordnung von  $d = 0,27$ . Der beobachtete Punktwert von 28 steht aber in beiden Gruppen für eine insgesamt sehr hohe Akzeptanz und Patientenzufriedenheit, wenn man Vergleichsuntersuchungen verschiedener Indikationsgruppen im Bereich der stationären Rehabilitation zur Bewertung als Maßstab heranzieht.

Tabelle 145. Akzeptanz der ambulanten Psychotherapie durch die Patienten der IG und KG

	IG			KG			Vergleich	
	N	M	SD	N	M	SD	d	p
ZUF-8 Skala Patientenzufriedenheit	286	28,90	3,37	229	27,87	4,44	0,27	.004
BESS Skala Gesundheitliche Veränderung	285	1,71	0,67	231	1,77	0,66	0,09	.286
Skala Evaluation des Modellvorhabens	285	1,55	0,44	229	1,50	0,39	-0,11	.230
Zufriedenheit mit dem Modellvorhaben	276	2,25	0,85	221	2,29	0,79	0,05	.604

Anmerkung. Codierung: ZUF-8 Wertebereich 8 (unzufrieden) – 32 (sehr zufrieden). BESS 1 (viel besser) – 5 (viel schlechter). Positive Evaluation des Modellvorhabens 1 (trifft voll zu) – 5 (trifft nicht zu). Zufriedenheit mit dem Modellvorhaben 1-vollkommen zufrieden, 2-sehr zufrieden, 3-zufrieden, 4-weniger zufrieden, 5-unzufrieden.

Bei der BESS-Skala zur Selbsteinschätzungen der gesundheitlichen Veränderungen, die aus 5 Items gebildet wird (Veränderung des gesundheitlichen, seelischen, körperlichen und Allgemeinbefindens sowie der Leistungsfähigkeit), gaben die Patienten im Durchschnitt an, sich viel bis etwas besser zu fühlen. Signifikante Gruppenunterschiede bestehen nicht, ebenso wenig wie bei der explorativ gebildeten Skala zur Evaluation des Modellvorhabens bzw. der 1-Item-Frage zur Zufriedenheit mit dem Modellvorhaben insgesamt.

Tabelle 146 gibt die Einzelheiten für die Items der BESS-Skala wieder. 94,4% aller Patienten nehmen eine Verbesserung ihrer seelischen Verfassung im Vergleich zum Behandlungsbeginn wahr, was ja unabhängig von der konkreten Symptomatik das wichtigste Anliegen einer Psychotherapie darstellt. Aber auch die körperliche Verfassung und die Leistungsfähigkeit – eine wichtige Voraussetzung für die Bewältigung der täglichen Rollenanforderungen in Beruf und Privatleben – haben sich aus Sicht der Befragten deutlich verbessert.

**Tabelle 146. Items der BESS-Skala zur direkten Veränderungsmessung des gesundheitlichen Befindens (IG und KG gemeinsam)**

Beurteilungsaspekt	1	2	3	4	5	N	M	SD
Gesundheitliches Befinden	252 48,6%	178 34,4%	73 14,1%	14 2,7%	1 0,2%	518	1,71	0,82
Seelisches Befinden	341 66,0%	147 28,4%	17 3,3%	11 2,1%	1 0,2%	517	1,42	0,68
Körperliches Befinden	167 31,9%	177 34,4%	143 27,8%	28 5,4%	2 0,4%	514	2,08	0,92
Allgemeinbefinden	275 53,4%	185 35,9%	42 8,2%	13 2,5%	0 0,0%	515	1,60	0,75
Leistungsfähigkeit	200 38,8%	195 37,9%	102 19,8%	15 2,9%	3 0,6%	515	1,89	0,86

Anmerkungen. Codierung: 1 – viel besser, 2 – etwas besser, 3 – unverändert, 4 – etwas schlechter, 5 – viel schlechter.

In Tabelle 147 sind die Häufigkeiten für die Beantwortungen der ZUF-8-Items detailliert wiedergegeben. Diese offenbaren eine beeindruckend hohe Gesamtzufriedenheit der Studienpatienten mit der ambulanten Psychotherapie. So äußern sich auf allen acht Items mehr 90%, bei vier Items sogar mehr als 95% der Befragten zufrieden mit der Behandlung.

**Tabelle 147. Items der ZUF-8-Skala zur Patientenzufriedenheit – Häufigkeiten für alle Patienten (IG und KG gemeinsam)**

Beurteilungsaspekt	1	2	3	4	N	M	SD
Gute Qualität der Behandlung	5 1,0%	25 4,8%	273 52,7%	215 41,5%	518	3,35	0,62
Behandlung erhalten, die Sie wollten	5 1,0%	16 3,1%	202 39,1%	293 56,8%	516	3,52	0,61
Behandlung Bedürfnissen entsprochen	3 0,6%	20 3,9%	223 43,2%	270 52,3%	516	3,47	0,60
Freund Therapeut empfehlen	8 1,5%	22 4,3%	108 20,9%	379 73,3%	517	3,66	0,64
Zufrieden mit Ausmaß an Hilfe	12 2,3%	20 3,9%	200 38,7%	285 55,1%	517	3,47	0,68
Behandlung bei Problemumgang geholfen	1 0,2%	7 1,4%	90 17,5%	417 81,0%	515	3,79	0,45
Zufrieden mit Behandlung	4 0,8%	18 3,5%	176 34,1%	318 61,6%	516	3,57	0,60
Wieder zu diesem Therapeuten kommen	17 3,3%	28 5,4%	96 18,7%	373 72,6%	514	3,61	0,74

Anmerkungen. Codierung: 1 – eindeutig verneint, 2 – eher verneint, 3 – eher bejaht, 4 – eindeutig bejaht.

Im Gegensatz zur BESS Skala und den beiden Instrumenten zur Evaluation des Modellvorhabens zeigt sich in der ZUF-8 Skala im Gruppenvergleich ein signifikanter Unterschied. Der Unterschied ist als klein einzustufen ( $d = .27$ ; im Mittel knapp ein Punkt) doch statistisch signifikant ( $p = .004$ ; s. Tabelle 145). Die ZUF-8-Skala reagiert erfahrungsgemäß sehr sensitiv auf Qualitätsunterschiede gerade im oberen Bereich hoher Prozessqualität und findet in Routinebefragungen im stationären Bereich daher als Frühindikator für Qualitätsveränderungen Anwendung vgl. (Steffanowski, Nübling, Schmidt, & Löschmann, 2006). Eine explorative Betrachtung der Unterschiede zwischen den Gruppen könnte erste Hinweise darauf liefern, wo Effekte durch die modifizierte Vorgehensweise in der IG gegenüber der KG erzielt werden konnten. Tabelle 148 zeigt die Ergebnisse. Der Trend zu einer etwas besseren Bewertung in der IG zeigt sich auf allen Items der ZUF-8-Skala (alle Effektstärken sind positiv). Die deutlichsten Unterschiede zeigen sich bei der Frage, ob die Behandlung den Bedürfnissen des Patienten entsprochen hat ( $d = .28$ ) und ob die Behandlung beim Problemumgang geholfen hat ( $d = .28$ ). Beides kann ein Indiz dafür sein, dass in der IG tatsächlich eine etwas stärkere Orientierung an den individuellen Bedürfnissen und Problemen des Patienten durch das Modellvorhaben erreicht wird.

**Tabelle 148. Items der ZUF-8-Skala zur Patientenzufriedenheit – Vergleich zwischen IG und KG**

	IG			KG			Vergleich
	N	M	SD	N	M	SD	d
Gute Qualität der Behandlung	287	3,39	0,56	231	3,29	0,68	0,16
Behandlung erhalten, die Sie wollten	285	3,56	0,56	231	3,46	0,66	0,17
Behandlung Bedürfnissen entsprochen	287	3,55	0,55	229	3,38	0,65	0,28
Freund Therapeut empfehlen	287	3,72	0,56	230	3,58	0,71	0,22
Zufrieden mit Ausmaß an Hilfe	287	3,52	0,64	230	3,40	0,73	0,17
Behandlung Problemumgang geholfen	285	3,85	0,37	230	3,72	0,53	0,28
Zufrieden mit Behandlung	285	3,63	0,54	231	3,49	0,67	0,23
Wieder zu diesem Therapeuten kommen	285	3,67	0,69	229	3,52	0,79	0,20

Anmerkungen. Codierung: 1 – eindeutig verneint, 2 – eher verneint, 3 – eher bejaht, 4 – eindeutig bejaht.

### 7.3 Evaluation aus Gutachtersicht

Es liegen insgesamt N = 201 Evaluationen des TK-Modellvorhabens durch Gutachter vor. Die Gutachter stehen dem Modellvorhaben insgesamt eher kritisch gegenüber (Tabelle 149). Die explorativ aus allen Items gebildete Gesamtskala hat einen Mittelwert von 2,94 (N = 195; SD = 0,62; Cronbachs Alpha = .62). Immerhin gibt ein Teil der Gutachter (35,9%) an, dass die grafischen Rückmeldungen und Handlungsempfehlungen des TK-Reports hilfreich für die Erstellung der Gutachten sind.

Tabelle 149. Evaluation des TK-Modellvorhabens durch die Gutachter

Beurteilungsaspekt	1	2	3	4	5	N	M	SD
Zeitlicher Aufwand für Gutachtenerstellung angemessen	38 19,3%	87 44,2%	11 5,6%	50 25,4%	11 5,6%	197	2,54	1,22
Grafische Rückmeldungen stimmten gut mit klinischem Bericht der Therapeuten überein	6 3,1%	60 30,8%	25 12,8%	62 31,8%	42 21,5%	195	3,38	1,21
Handlungsempfehlungen stimmten gut mit klinischem Bericht der Therapeuten überein	5 2,6%	41 21,7%	42 22,2%	61 32,3%	40 21,2%	189	3,48	1,13
Bei meinem Gutachten vor allem grafischen TK-Reports berücksichtigt	7 3,6%	34 17,4%	53 27,2%	66 33,8%	35 17,9%	195	3,45	1,08
Bei meinem Gutachten vor allem die verschiedenen Fremdbeurteilungen berücksichtigt	23 11,8%	72 36,9%	48 24,6%	36 18,5%	16 8,2%	195	2,74	1,14
Bei meinem Gutachten vor allem klinischen Kurzbericht berücksichtigt	75 38,5%	74 37,9%	24 12,3%	22 11,3%	0 0,0%	195	1,96	0,98
Grafische Rückmeldung für Gutachtenerstellung hilfreich	7 3,6%	63 32,3%	59 30,3%	38 19,5%	28 14,4%	195	3,09	1,11

Anmerkungen. Codierung: 1 – trifft voll zu, 2 – trifft eher zu, 3 – weder / noch, 4 – trifft eher nicht zu, 5 – trifft überhaupt nicht zu.

Tabelle 150. Gutachterzufriedenheit mit dem Modellvorhaben insgesamt

Wie zufrieden sind Sie mit dem Modellvorhaben „Qualitätsmonitoring in der ambulanten Psychotherapie“ insgesamt?	N	%	
Vollkommen zufrieden	1	0,5	
Sehr zufrieden	3	1,6	
Zufrieden	48	25,7	
Weniger zufrieden	89	47,6	
Unzufrieden	46	24,6	
Gesamt	187	100,0	

Auch die Gesamtzufriedenheit mit dem Modellvorhaben wird durch die Gutachter deutlich zurückhaltender bewertet als durch die Patienten oder Therapeuten (Tabelle 150).

Gemäß den Gutachtern dauerte die Erstellung eines Gutachtens zwischen 10 und 70 Minuten (N = 134; M=25,1; SD=12,9).

Auch wenn die Gutachter das Modellvorhaben insgesamt eher kritisch bewerten, so werden die verwendeten Assessment-Instrumente tendenziell als hilfreich für die Gutachtenerstellung bewertet (siehe Tabelle 151). Die explorativ erstellte Skala aus allen Einzelbewertungen hat einen Mittelwert von 2,77 (N = 153, SD = 0,54, Cronbachs Alpha = .69).

**Tabelle 151. Gutachtereinschätzungen der Nützlichkeit der psychodiagnostischen Tests und der Fremdbeurteilungen für die Erstellung der Gutachten**

Beurteilungsaspekt	1	2	3	4	5	N	M	SD
BSI hilfreich für Gutachtenerstellung	13 7,2%	95 52,5%	35 19,3%	19 10,5%	19 10,5%	181	2,65	1,10
IIP-D hilfreich für Gutachtenerstellung	10 5,7%	85 48,3%	43 24,4%	23 13,1%	15 8,5%	176	2,70	1,05
Störungstest hilfreich für Gutachtenerstellung	12 7,2%	79 47,3%	45 26,9%	17 10,2%	14 8,4%	167	2,65	1,04
HAQ hilfreich für Gutachtenerstellung	11 6,3%	82 46,6%	53 30,1%	23 13,1%	7 4,0%	176	2,62	0,93
SF12 hilfreich für Gutachtenerstellung	9 5,1%	67 38,3%	63 36,0%	26 14,9%	10 5,7%	175	2,78	0,96
Basisdokumentation hilfreich für Gutachtenerstellung	10 6,1%	59 35,8%	62 37,6%	28 17,0%	6 3,6%	165	2,76	0,93
Therapieeingangsdagnostik hilfreich für Gutachtenerstellung	9 5,4%	77 46,4%	56 33,7%	19 11,4%	5 3,0%	166	2,60	0,87
Therapeuteneinschätzungen hilfreich für Gutachtenerstellung	35 20,8%	66 39,3%	47 28,0%	15 8,9%	5 3,0%	168	2,34	1,00
Psychopharmakologische Behandlung hilfreich für Gutachtenerstellung	2 1,4%	12 8,7%	61 44,2%	16 11,6%	47 34,1%	138	3,68	1,08

**Anmerkungen. Codierung: 1 – sehr hilfreich, 2 – eher hilfreich, 3 – weder / noch, 4 – eher nicht hilfreich, 5 – überhaupt nicht hilfreich.**

Wie auch die Patienten konnten Gutachter Anmerkungen und Hinweise im Evaluationsbogen äußern. Von den 91 Angaben bezog sich der Großteil (24,2%) auf zu wenig durch das Modellprojekt generierte Informationen für das Gutachten. Kritik an fehlenden Unterlagen (19,8%) machte die zweitgrößte Gruppe aus, gefolgt von Schwierigkeiten durch unleserliche Handschrift (8,8%). Eine vollständige Auflistung der Gutachteranmerkungen befindet sich im Anhang.

## 8 Beantwortung der Studienfragestellungen

### 8.1 Hauptfragestellungen

Die zentralen Fragestellungen des Projektes (Evaluationsplan von 2003, vgl. Abschnitt 2.1) sind:

1. Führt der Einsatz von Qualitätsmonitoringinstrumenten des TK-Modells in ambulanten Psychotherapien zu einer höheren Ergebnisqualität (Effektivität) im Vergleich zur traditionellen Richtlinienpsychotherapie ?
2. Kann das TK-Modell die Effizienz der ambulanten Psychotherapie nachhaltig verbessern?

**Zu Fragestellung 1: Nein. Die Vorgehensweisen nach der traditionellen Richtlinienpsychotherapie und nach dem TK-Modell führen zu keinen statistisch signifikanten Unterschieden bei der Ergebnisqualität.**

**Zu Fragestellung 2: Nein. Die Datenanalysen ergaben keinen Hinweis, dass hinsichtlich der erreichten gesundheitlichen Veränderungen bei den Patienten pro eingesetzter Therapiestunde Gruppenunterschiede zwischen der traditionellen Richtlinienpsychotherapie und dem TK-Modell zu erwarten sind.**

Nachfolgend wird zu den in Abschnitt 2.4 spezifizierten Hypothesen, welche neben den beiden Hauptfragestellungen auch explorative Nebenfragestellungen ansprechen, detailliert Stellung genommen:

#### **Makroebene - Wirksamkeit und Wirtschaftlichkeit**

a) Das Qualitätsmonitoring führt bei allen nicht monetären Kriterienmaßen zu signifikant besseren Ergebnissen als bei der Kontrollgruppe 1, für die diese Kriterien ebenfalls erhoben werden (für die Kontrollgruppen 2 und 3 werden diese Informationen nicht erhoben).

**Nein: Weder bei den Therapeutenangaben (Vergleich zwischen A- und E-Messung), noch bei den Patientenangaben (Vergleich zwischen A- und E- bzw. A- und K-Messung) zeigen sich – auch bei Kontrolle möglicher Selektionseffekte (Confoundervariablen) - statistisch signifikante Gruppenunterschiede zwischen der Interventions- und Kontrollgruppe auf dem multiplen Ergebnisindikator.**

b) Das Qualitätsmonitoring, das bei der Experimentalgruppe verwendet wird, führt zu höheren Effektgrößen auf dem EMEK-monetär bei allen Vergleichen von EX-Grp. mit KON-Grp. 1, und auf dem EMEK-monetär bei der Katamnese zu allen Kontrollgruppen 1-3.

**Diese Fragestellung lässt sich aufgrund der Datenlage nicht beantworten.**

### 8.2 Nebenfragestellungen

#### **Makroebene - Wirksamkeit und Wirtschaftlichkeit (Fortsetzung)**

c) Wir erwarten für alle vier Gruppen im EMEK-monetär Effektgrößen  $d > .20$  im Vorher-Nachher-Vergleich, d.h. für den Vergleich des Einjahreszeitraumes vor der Therapie und den der Katamnese.

**Diese Fragestellung lässt sich aufgrund der Datenlage nicht beantworten.**

---

## Mesoebene - Diagnostische und adaptive Entscheidungsstrategien

a) Das TK-Modell (EX-Grp.) führt zu einem bedarfsgerechteren Einsatz der therapeutischen Ressourcen (Indikator: Varianz der Behandlungsdauer in der EX-Grp. ist größer als in der KON-Grp. 1).

**Nein: Der Test auf Varianzgleichheit bei der Anzahl der abgerechneten Therapiestunden ergab, dass die IG und KG sich hinsichtlich der Varianz der abgerechneten Therapiestunden nicht statistisch signifikant voneinander unterscheiden. Unabhängig von dem Ergebnis ist zu diskutieren, ob die Varianz tatsächlich im Sinne von Bedarfsgerechtigkeit interpretiert werden kann. Beim Gruppenmittelwert der abgerechneten Therapiestunden besteht ein statistisch signifikanter Unterschied zwischen IG und KG: in der IG werden mehr Sitzungen abgerechnet.** Der Unterschied ist im Vergleich zur Varianz in der Versorgung nicht ungewöhnlich. So finden sich in den Versorgungskontrollgruppen, die nicht am MVPT teilnahmen, ähnliche Werte.

b) Die Indikationsstellung bei der Aufnahme von Psychotherapie kann in der EX-Grp. durch den Einsatz verschiedener diagnostischer Instrumente gegenüber der KON-Grp. I verbessert werden.

**Diese Hypothese kann mit den erhobenen Daten nicht schlüssig beantwortet werden.**

c) Die Therapeuten in der EX-Grp. nutzen die Handlungsempfehlungen in unterschiedlicher Weise.

**Ja: Während ein Teil der Therapeuten die Handlungsempfehlungen überhaupt nicht nutzt, reicht bei anderen Therapeuten das Spektrum vom reinen Besprechen der Ergebnisse (häufiger angekreuzt) bis hin zur Anpassung der therapeutischen Informationen und auch Vorbereitung des Therapieendes (seltener angekreuzt).**

d) Die beteiligten Gutachter der EX-Grp. nutzen die Handlungsempfehlungen in unterschiedlicher Weise.

**Ja: Die Gutachter nutzen vor allem den klinischen Kurzbericht und verlassen sich weniger auf die grafischen Rückmeldungen.**

e) Die Konsistenz der eingesetzten Entscheidungsregeln und Handlungsoptionen unterscheidet sich aus Patienten-, Therapeuten- und Gutachtersicht.

**Ja: Am positivsten wird das Modellvorhaben von den Patienten bewertet, das hier auf gute Akzeptanz stößt. Die Therapeuten nehmen eine Mittelstellung ein. Die Gutachter bewerten das Rückmeldesystem hingegen eher kritisch.**

f) Die in der EX-Grp. eingesetzten Messinstrumente decken die zentralen Problembereiche aus Patientensicht ab.

**Ja: 67,0% der Patienten bejahten diese Frage, 14,2% waren sich unsicher und 8,8% verneinten.**

g) Der Umfang des Erhebungsinventars muss für einen praktikablen Einsatz in der Routine angepasst werden.

**Ja: Der SF12 wird z.B. kaum für die Rückmeldungen genutzt, liefert auf der anderen Seite aber Referenzwerte für die Vergleichbarkeit mit anderen Studien. Dieser Punkt wäre unter Einbeziehung der niedergelassenen Therapeuten zu diskutieren. Darüber hinaus empfiehlt sich für einen Routineeinsatz aus Kostengründen die Verwendung lizenzfreier Instrumente und Verfahren.**

---

### **Mikroebene - Praktikabilität und Akzeptanz**

a) Die Patienten in der Experimentalgruppe erleben die Rückmeldung als eine Stärkung ihrer Befähigung (Empowerment).

**Ja. 73,6% aller Patienten bewerteten die Informationen aus den Rückmeldungen als für sie persönlich wichtig. 66,3% empfanden die Rückmeldungen als hilfreich. Immerhin 47,3% gaben an, durch die Rückmeldungen selbstverantwortlicher mit ihrer Erkrankung umgehen zu können.**

b) Es wird eine relativ hohe Akzeptanz der ambulanten Psychotherapie bei Experimental- und Kontrollgruppe auf den Messinstrumenten ZUF-8 und SERVQUAL erwartet.

**Ja: Auf der ZUF-8-Skala liegt die beobachtete Patientenzufriedenheit mit 28 Punkten im oberen Bereich der 8 Items umfassenden Skala, die Werte zwischen 8 und 32 Punkten annehmen kann. Die SERVQUAL-Skala war im Evaluationsplan zwar ursprünglich vorgesehen, wurde dann aber nicht in das Assessment übernommen, so dass hier keine Daten erhoben wurden.**

c) Experimental- und Kontrollgruppe unterscheiden sich im Spiegel der Patientensicht nicht signifikant voneinander.

**Teilweise: Auf der ZUF-8-Skala ergab sich in der IG eine um einen Punkt höhere Patientenzufriedenheit als in der KG, was einer kleinen Effektgröße von  $d = 0,27$  entspricht. Bei der BESS-Skala zur direkten Veränderungsmessung des gesundheitlichen Befindens, bei der explorativ gebildeten Gesamtskala zur Evaluation des Modellvorhabens sowie bei der 1-Item-Einschätzung der Gesamtzufriedenheit mit der Studie zeigten sich hingegen keine statistisch signifikanten Unterschiede zwischen IG und KG (vgl. Tab. 145).**

d) Die Praktikabilitätseinschätzung des Qualitätsmonitorings aus Therapeutesicht steigt mit der Anzahl der Therapien, die mit diesem System durchgeführt werden.

**Teilweise: Therapeuten mit mehr abgeschlossenen („Modell“-)Therapien sind insgesamt etwas zufriedener mit dem Modellvorhaben, wobei die Effektgröße mit  $d = 0,22$  allerdings klein ausfällt. Die explorativ gebildete Skala zur Praktikabilitätseinschätzung korreliert hingegen nicht signifikant mit der Anzahl der abgeschlossenen Therapien. Immerhin ergeben sich Hinweise dahingehend, dass eine elektronische Dokumentation der traditionellen Papier- und Kugelschreiber-Fragebogenmethode (mit der damit verbundenen umständlichen Auswertung) vorgezogen wird, insgesamt 84,7% aller Befragten bejahten dies. Je besser das PC-gestützte Datenerfassungssystem funktionierte, desto mehr Behandlungen wurden dabei dokumentiert.**

## 9 Diskussion

Die wichtigsten Ergebnisse der Studie lassen sich wie folgt zusammenfassen: Bezogen auf die Hauptfragestellungen zeigen sich auch bei Berücksichtigung möglicher Confoundervariablen keine statistisch signifikanten Unterschiede zwischen IG und KG auf den zur Bewertung der Ergebnisqualität herangezogenen Skalen und multiplen Ergebniskriterien.

### 9.1 Effektivität

#### 9.1.1 Angaben zum gesundheitlichen Befinden (EMEK-nonmonetär)

Weder bei den Therapeutenangaben (Vergleich zwischen A- und E-Messung), noch bei den Patientenangaben (Vergleich zwischen A- und E- bzw. A- und K-Messung) zeigen sich – auch bei Kontrolle möglicher Selektionseffekte (Confoundervariablen) – statistisch signifikante Gruppenunterschiede zwischen der Interventions- (IG) und Kontrollgruppe (KG) auf dem multiplen Ergebnisindikator (EMEK nonmonetär). Einzelne statistisch signifikante Vergleiche sollten vor dem Hintergrund einer möglichen Alpha-Fehler-Inflation nicht überbewertet werden: So profitierten die Patienten der IG aus Sicht der Therapeuten (A/E-Vergleich) etwas deutlicher bei der Problembewältigung im privaten Bereich und bei der Selbstverantwortung. Dieser Befund deckt sich mit den etwas größeren Zuwächsen im IIP-D aus Sicht der Patienten (ebenfalls A/E-Vergleich). Im A/K-Vergleich zeigt sich beim BDI ein etwas besseres Abschneiden der IG. Die Gruppenunterschiede sind allerdings klein und der numerische Unterschied im BDI ist bei Berücksichtigung klinischer Bewertungskriterien zudem irrelevant. Da sich die Prüfhypothese jedoch auf alle Einzelaspekte bezieht, muss diese hier klar zurückgewiesen werden. Dies bedeutet, dass Interventions- und Kontrollgruppe in der Gesamtschau gleich gut abschneiden.

#### 9.1.2 Gesundheitsökonomische Daten (EMEK-monetär)

Die im Evaluationsplan ursprünglich vorgesehene vergleichende Beurteilung der Wirtschaftlichkeit (EMEK monetär) zwischen IG und KG anhand der bei der TK dokumentierten Leistungsausgaben (GKV-Routinedaten) erwies sich im Studienverlauf als nicht realisierbar. Die hierzu verfügbare Datenbasis hat sich durch strukturelle Änderungen in den Routinedaten der Krankenkassen während des Studienzeitraumes so stark verändert, dass keine reliablen und validen Prä-Post-Vergleiche des 1-Jahres-Zeitraumes vor und nach der Therapie möglich sind. Diese Einschränkung ist unter anderem auf Veränderungen der gesetzlichen Rahmenbedingungen im Studienzeitraum zurückzuführen (u.a. Einführung des morbiditätsorientierten Risikostrukturausgleichs), welche die Voraussetzungen definieren, unter denen die Verbrauchsdaten zu dokumentieren sind.

## 9.2 Effizienz

#### 9.2.1 Kosten-Nutzen-Verhältnis: Effektivität pro Therapiestunde

Die erste Hauptfragestellung nach Unterschieden bei der Effektivität zwischen IG und KG wurde bereits verneint. Hinsichtlich der Anzahl abgerechneter Therapiesitzungen ergab sich zwar ein statistisch signifikanter Unterschied zwischen IG und KG. In der IG und KG kam es aufgrund unterschiedlicher Bewilligungsmodi zu unterschiedlichen Therapielängen. Die Kontrollgruppen aus der Versorgung (KG2 und KG3) zeigen aber auch, dass ohne ein Modellvorhaben eine vergleichbare Bandbreite von Therapielängen realisiert wird. Einen Unterschied zwischen IG und KG bezüglich des Kosten-Nutzen-Verhältnisses (Effektivität pro Therapiestunde) abzuleiten, ist daher nicht möglich.

#### 9.2.2 Dokumentationsaufwand für die Therapeuten

Die Frage nach der Effizienz berührt aber auch den Dokumentationsaufwand für die Therapeuten, der sich ja nicht in der Zahl der abgerechneten Therapiesitzungen niederschlägt, sondern unabhängig davon

---

zu leisten ist. So steht dem verringerten Aufwand durch die verkürzte Begutachtung in der IG ein höherer Aufwand durch die psychometrischen Messungen gegenüber.

## 9.3 Nebenfragestellungen

### 9.3.1 Kategoriale Diagnostik

Die Ergebnisse der kategorialen Diagnostik zu Therapiebeginn mit dem größten Anteil Diagnosen im Bereich der Angst- und Affektiven Störungen entsprechen den Erwartungen nach den Prävalenzzahlen dieser Störungsgruppen in der Allgemeinbevölkerung (Jacobi, Klose & Wittchen, 2004). Während in der KG keine Vorgaben bezüglich des diagnostischen Vorgehens gemacht wurden, wurden die Therapeuten der IG dazu angehalten, zur Unterstützung der Diagnosestellung auf IDCL-Checklisten (Hiller, et al., 1995) zurückzugreifen. Zwischen Interventions- und Kontrollgruppe konnten Unterschiede in Art und Anzahl diagnostizierter Störungen gefunden werden, die Hinweise auf ein verändertes Diagnoseverhalten in der Interventionsgruppe durch die Anwendung von Diagnosechecklisten liefern. Jedoch ist einschränkend festzuhalten, dass keine Information zur Reliabilität der gestellten Diagnosen vorliegt. Somit lassen sich alleine aus der Anzahl vergebener Diagnosen keine Aussagen über die Qualität der Diagnostik oder der Therapie an sich ableiten. Darüber hinaus ist anzumerken, dass der Anteil an Patienten mit Dritt- und Viertdiagnosen auch in der IG gering ist und dass Patienten, für die keinerlei diagnostische Information vorliegt, deutlich häufiger der IG zugehörig sind als der KG.

### 9.3.2 Ausgangsbelastung und Psychotherapiebedarf der Patienten

Die Patienten in dieser Studie zeigen sich sowohl in allgemeinen als auch in störungsspezifischen Instrumenten durchgängig mindestens genauso belastet wie klinische Vergleichsstichproben. Sie überschreiten im Mittel deutlich die Grenzwerte für klinisch relevante Belastung, in einigen Instrumenten liegt der Mittelwert sogar im Bereich für starke Belastung (vergleichbar mit stationär behandelten Patienten). Der Vergleich zwischen IG und KG hinsichtlich der Eingangsbelastung ergibt teils signifikante Unterschiede, mit jeweils etwas größerer Belastung in der IG, wobei hier allerdings aufgrund der Stichprobengröße auch kleine Effekte signifikant werden. Insgesamt ist festzuhalten, dass aufgrund der festgestellten klinisch relevanten Beeinträchtigung der hier untersuchten Patientenstichprobe sowohl in der IG als auch in der KG der psychotherapeutische Behandlungsbedarf evident ist.

### 9.3.3 Verlaufsmessungen

Die Effektgrößen zu den einzelnen Verlaufsmessungen zeigen, dass im Therapieverlauf bereits Veränderungen eintreten. Jede der Verlaufsmessungen zeigt größere Effektgrößen als die vorangegangene, auch wenn die Unterschiede zwischen den Verlaufsmessungen etwas zurückgehen. Dieser Befund passt zu einer Reihe von Theorien und vorigen Befunden zur Entwicklung von Patienten im Verlauf der Therapie (z.B. Lutz, 2002; Lutz, Stulz, Köck, 2009). Darüber hinaus zeigen die Instrumente sehr unterschiedliche Effektgrößen im Verlauf, was verdeutlicht, dass sie unterschiedliche Prozesse der Veränderung messen und die Therapie so über unterschiedliche Aspekte informieren können. So ist der BDI ein Instrument zur Bewertung sehr spezifischer Therapieerfolge und zeigt in der relevanten Population große Effektgrößen. Der BSI als ein Maß für die Gesamtbelastung der Patienten ist weniger sensitiv für spezifische Veränderungen in der Therapie, doch hält er Informationen zur Gesamtsymptombelastung bereit, die auch Ziel therapeutischer Veränderungsprozesse sein können – was die Ergebnisse im Übrigen auch belegen. Der IIP-D gibt eine zusätzliche Perspektive über die interpersonale Problemlage des Patienten und verändert sich erst später als zum Beispiel die Symptommaße. Dies betont noch einmal die Wichtigkeit der multidimensionalen Bewertung von Veränderungsprozessen in der Therapie, und dass für verschiedene Abschnitte der Therapie unterschiedliche Ergebnisdimensionen interessant sein können. Als Alternative zu den im Modellvorhaben genutzten (meist proprietären) psychometrischen Instrumenten könnten

---

daher auch Instrumente, welche spezifisch zur multidimensionalen Erfassung des Therapieverlaufs entwickelt wurden, eingesetzt werden (Howard et al., 1993; Schulte, 1993; Lutz et al., 2009).

Die Behandlungsverläufe in Relation zu den Rückmeldungen sind vergleichbar zu bisherigen Befunden in der Forschung zu Rückmeldungen im Therapieverlauf (z.B. Lambert, 2007). Dass Patienten mit einer temporären Verstärkung der Problematik sich im Anschluss an diese Rückmeldung verbessern, könnte als Effekt der Rückmeldungen gedeutet werden. Zum Beleg dieser Interpretation fehlt jedoch eine Kontrollgruppe, in der lediglich wiederholte Messungen ohne die Bereitstellung einer Rückmeldung durchgeführt wurden. Um den Einfluss des Feedbacks von dem der reinen mehrfachen Datenerhebung zu trennen, sollte daher in zukünftigen Projekten eine Kontrollgruppe mit Messwiederholung, aber ohne Feedback erhoben werden.

#### 9.3.4 Prä-Post-Effekte

Insgesamt zeigt die Effektgrößenanalyse einen deutlichen positiven Effekt ambulanter Psychotherapie auf allen verwendeten Messinstrumenten sowie die Analyse der Zufriedenheit eine hohe Patientenzufriedenheit bzgl. der psychotherapeutischen Maßnahmen. Die Bewertung anhand von Effektgrößen hängt auch z.B. von den Streuungen in den untersuchten Stichproben ab und berücksichtigt weder die Ausgangsbelastung noch den Abschlussstatus der Patienten. Die in dieser Studie daher zusätzlich vorgenommene kriteriumsorientierte Einstufung der Veränderung in der Therapie nach dem Konzept der klinisch signifikanten Veränderung basierte auf drei Dimensionen psychischer Belastung: allgemeine Symptombelastung, störungsspezifische Belastung und interpersonale Probleme. Dieses strengere Erfolgskriterium zeigte ebenfalls die Effektivität der Therapie in den Verlaufs-, Abschluss- und Katamneseerhebungen. In den Verlaufsmessungen zeigen sich zu jedem Erhebungszeitpunkt knapp 30% der Patienten gegenüber der Eingangsbelastung als zumindest reliabel gebessert. Werden abgeschlossene Therapien betrachtet, erfüllen über 60% der Fälle das strengste Erfolgskriterium „(nachhaltige) Abnahme der Problematik“ (über 70% zum Katamnesezeitpunkt).

Sowohl bei der IG als auch bei der KG ergeben sich sehr hohe und stabile Prä-Post-Effektgrößen in einer Größenordnung von 1,20 auf den zur Gesamtevaluation verwendeten multiplen Status- und Ergebnisdikatoren. Kritisch muss hier angemerkt werden, dass bei einer großen Zahl von Patienten entweder keine Entlassungsmessung (E), keine Katamneseermessung (K) oder beides nicht vorlag. Über die Gründe wurde in Kapitel 3.8 berichtet. Bei fehlender Entlass- oder Katamneseermessung liegt aber für viele Patienten die jeweils andere Messung vor, so dass die Möglichkeit bestand, die Qualitätspunktwerte der multiplen Ergebniskriterien zwischen verschiedenen Dropout-Gruppen der Interventionsgruppe und der Kontrollgruppe zu vergleichen. Der Unterschied zwischen IG und KG liegt bei etwa 2 Qualitätspunkten und die Streuung innerhalb der Gruppen beträgt rund 10 Qualitätspunkte. Daher liegen diese Verzerrungen umgerechnet in Cohen's d Effektstärken bei  $2 / 10 = 0,20$ . Die Prä-Post Effektstärken sind deshalb als mit einem positiven Bias versehen zu bewerten. Zieht man diese positive Überschätzung des Prä-Post Therapieerfolges von der KG bzw. der IG ab, verbleibt immer noch eine Effektgröße d für den A/E-Vergleich von 1,01 bei der IG und 0,92 bei der KG bzw. für den A/K-Vergleich von 1,07 für die IG und 0,93 für die KG. Das Vorhandensein der E- bzw. K-Messung wurde in den Regressionsgleichungen zur statistischen Kontrolle möglicher Selektionseffekte berücksichtigt, so dass auch die biaskorrigierten Effektgrößen der IG und KG als gleich groß anzusehen sind.

---

### 9.3.5 Abschätzung der Kosten-Nutzen-Relation anhand der Prä-Post-Effektgrößen

Die im Modellvorhaben beobachtete Effektgröße von 1,0 dokumentiert eine beeindruckende Leistung der ambulanten psychotherapeutischen Versorgung, die im oberen Bereich internationaler metaanalytischer Vergleichsuntersuchungen liegt. Auch wenn dies keine Hauptfragestellung des Evaluationsplanes darstellt, sollte dieser Punkt bei der Bewertung des Modellvorhabens deutlich hervorgehoben werden.

Eine ökonomische Bewertung von Interventionen ist möglich, auch wenn keine direkten und indirekten Kosten und Nutzenaspekte vorliegen. Man muss dann jedoch einige Annahmen treffen, die auf ihre Plausibilität hin bewertet werden müssen. Hilfreich ist in diesem Zusammenhang eine sogenannte Sensitivitätsanalyse bei der einzelne Parameter zwischen den beiden Polen konservativ und optimistisch variiert werden, um die Auswirkungen dieser Variationen auf die Bewertung und die Robustheit von Schlussfolgerungen zu überprüfen.

Die Brogden/Cronbach/Gleser Kosten-Nutzengleichung gilt für rein monetäre Ergebnismaße ebenso wie für nichtmonetäre Ergebnismaße, wie die von uns eingesetzten multiplen Ergebniskriterien. Nichtmonetäre Ergebnismaße korrelieren nicht perfekt aber meist substantiell mit monetären Ergebnismaßen. Im Sinne pfadanalytischer Terminologie kann man sich den Wirkungsmechanismus einer Intervention, wie hier der ambulanten Psychotherapie, auf ein reines monetäres Kriterium so vorstellen, dass der Effekt vollkommen über ein nichtmonetäres Kriterium vermittelt wird. Im englischen Sprachraum spricht man von einem „fully mediated effect“. Ein solcher indirekter Effekt berechnet sich aus dem Produkt der Effekte Intervention auf das nichtmonetäre Kriterium multipliziert mit dem Effekt des nichtmonetären Kriteriums auf das monetäre Kriterium. Neben diesem indirekten Effekt besteht theoretisch aber auch die Möglichkeit eines direkten Effektes von Psychotherapie auf das monetäre Ergebniskriterium, den wir für die folgende Betrachtung konservativ auf den Wert Null setzen. Den Effekt ambulanter Psychotherapie auf das nichtmonetäre multiple Ergebniskriterium kennen wir, er beträgt korrigiert um den Bias rund  $d=1,0$ , dieser Effekt war um mehr als das Doppelte größer als wir in der Eingangshypothese (2.4.3) vermutet hatten. Den Effekt nichtmonetär-monetär kennen wir nicht, schätzen ihn aber z.B. mit .30, .40, .50 oder .60. In der Kosten-Nutzengleichung ist dieser Effekt mit dem Parameter  $a$  bezeichnet.

In früheren Evaluationsstudien zur Kosten-Effektivität von stationärer Psychotherapie von psychosomatisch erkrankten Versicherten haben wir vergleichbare Berechnungen durchgeführt (Steffanowski et al., 2007). Im dortigen Bereich konnten wir die Ergebnisse unserer Methode mit einer Analyse vergleichen, die auf reinen monetären Ergebnismaßen beruhte (Zielke, 1993, Zielke et al., 2004). Die Ergebnisse waren vergleichbar und führten zu identischen Schlussfolgerungen über die Kosten-Nutzen Relationen, deshalb können wir hier auch von einer hohen Plausibilität der Annahmen hinter dieser Sensitivitätsanalyse ausgehen.

Unter 5.4.12 haben wir eine erste explorative Abschätzung der Kosten-Nutzen-Relation mit dem Parameter  $a=.50$  vorgenommen. Die Modellrechnung in Abschnitt 5.4.12 demonstriert, dass sich jeder in die Psychotherapie investierte Euro gesamtgesellschaftlich gesehen mehr als zweifach auszahlt - selbst wenn man davon ausgeht, dass der Therapieeffekt nur ein Jahr lang anhält. Setzen wir nun die anderen drei Schätzungen dieses Parameters ein, so führen wir eine Sensitivitätsanalyse durch. Diese Sensitivitätsanalyse erbringt für alle vier Werte eine Kosten-Nutzen-Relation von 1,95; 2,60; 3,26 bzw. 3,90. Die Bandbreite reicht deshalb von Schätzungen zwischen 2 bis 4 Euro an gesamtgesellschaftlichem Nutzen, pro in die ambulante Psychotherapie investiertem Euro. Angesichts der Tatsache, dass die hier gemessenen Therapieeffekte zwischen Entlassung und Katamnese stabil sind, könnte man ebenso von einem Anhalten der Effekte über mindestens zwei oder auch drei Jahre hinweg ausgehen. Ambulante Psychotherapie ist also keine Kostenverschwendung, ambulante Psychotherapie rechnet sich!

Dies bedeutet für erfolgreich behandelte Patienten nicht nur eine nachhaltige Verbesserung ihrer Lebensqualität, sondern auch ihrer Arbeitsfähigkeit. Somit zeugen die im Modellvorhaben gemessenen Effektstärken nicht zuletzt von einem erheblichen und (volks-) wirtschaftlich relevanten Kosten-Nutzen Potential der ambulanten psychotherapeutischen Versorgung. Den wirtschaftlichen Hauptnutzen haben in erster Linie die Arbeitgeber durch die verbesserte Produktivität ihrer Mitarbeiter, in zweiter Linie aber

---

auch die Krankenkassen durch geringere Ausgaben beim Krankengeld sowie höhere Beiträge der Versicherten in einem regulären Arbeitsverhältnis. Es empfiehlt sich daher, deutlich mehr Kostentransparenz über eine Evidenzbasierung von Maßnahmen im Gesundheitswesen zu etablieren, wie sie hier beispielhaft im TK-Modellprojekt aufgezeigt wurde.

### 9.3.6 Zufriedenheit der Patienten und Therapeuten

Auch wenn die Anzahl an durchgeführten IG-Therapien in einem leicht positiven Zusammenhang mit der Gesamtzufriedenheit mit dem Modellvorhaben steht, ist dabei der erhöhte Aufwand für Therapeut und Patient durch den Umfang der psychometrischen Messungen zu berücksichtigen. Für eine Optimierung des Feedbacksystems wären die Hinweise aus der Patientenbefragung (formative Evaluation) zu berücksichtigen, dass nämlich Messungen in den im Modellprojekt praktizierten Zeitabständen in vielen Fällen den Verlauf der Therapie nicht adäquat abbilden. Das Qualitätsmonitoring sollte - so die Patienten - häufiger, aber in geringerem Umfang durchgeführt werden. Auch wird ggf. ein Bedarf zur Weiterentwicklung der grafischen und inhaltlichen Gestaltung der Rückmeldungen bzw. zur Optimierung des Erhebungssystems gesehen, so dass Abläufe erleichtert und die Rückmeldungen für alle Beteiligten verständlicher und besser interpretierbar werden.

Bei der Projektevaluation durch die Patienten wird in der IG eine etwas höhere Patientenzufriedenheit im ZUF-8 berichtet, wobei der Unterschied zur KG klein ist. Möglicherweise gelang durch das Modellvorhaben bei einigen Patienten der IG eine bessere Abstimmung zwischen individuellem Anliegen und Behandlungsangebot durch den Therapeuten als beim traditionellen Vorgehen in der KG. Zudem empfinden Patienten die Rückmeldung der psychometrischen Testergebnisse offenbar als Zugewinn an Transparenz, welche sie grundsätzlich begrüßten, was sich wiederum in leicht höheren Zufriedenheitswerten mit dem Modellvorhaben äußert. Auf der anderen Seite könnte dies aber auch im Sinne eines Hawthorne-Effektes interpretiert werden: Dadurch, dass den Therapeuten der IG durch die Checklisten, den (kostenlosen) HTS-Zugang, die Schulung usw. sowie den Patienten der IG durch die Verlaufsmessungen, Evaluationsfragen usw. mehr Aufmerksamkeit zukam, bewerten diese die therapeutische Beziehung und Therapie insgesamt etwas positiver, schneiden bei der Ergebnisqualität dann aber dennoch nicht signifikant besser ab als die KG. Weiter ist die unterschiedliche Vorgehensweise bei der Diagnostik in der IG (Verwendung von Diagnose-Checklisten) gegenüber der KG als mögliche Erklärung für die beobachteten Unterschiede im ZUF-8 zu berücksichtigen.

### 9.3.7 Therapiedauer und Varianzen

Ein Befund des Projektes ist, dass sich die Therapien in der KG und IG in den bewilligten und den abgerechneten Sitzungen unterschieden. So beantragten die Therapeuten der IG für ihre am Modellvorhaben teilnehmenden Patienten in der Regel sofort Langzeittherapien, welche die TK aufgrund der Symptombelastung meist ohne (modifiziertes) Gutachten mit den üblichen Kontingenten bewilligte, während in der KG sowohl Kurz- als auch Langzeittherapien beantragt wurden. Ein weiterer Unterschied zwischen den Gruppen findet sich in der um ca. 0,5 Stunden längeren Probatorik in der IG, der weitestgehend durch die strukturiertere Eingangsdiagnostik durch IDCL-Checklisten bedingt sein dürfte.

Die Therapiedauer - ausgedrückt in der Zahl abgerechneter Therapiesitzungen - zeigt in den beiden Gruppen keine Unterschiede in der Variabilität. So veränderte die de facto breitere Kontingentierung in der IG die mittlere Sitzungszahl, jedoch nicht die Variabilität. Ob Unterschiede in der Varianz der Therapiesitzungen überhaupt höhere Bedarfsgerechtigkeit repräsentieren, muss offen bleiben, da wohl eher die unterschiedlichen Kontingentierungen und die unterschiedlichen Richtlinienverfahren die Variabilität beeinflussen.

---

## 9.4 Methodische Aspekte

Bei der Überprüfung der Randomisierung ergaben sich, abgesehen von einzelnen signifikanten Einzelvergleichen, keine Anhaltspunkte dafür, dass Patienten mit bestimmten Merkmalen bei der Studienteilnahme von vorneherein systematisch bevorzugt oder ausgeschlossen worden wären. Die gebildeten multiplen Statusindikatoren sowie die auf deren Prä/Post-Differenzwerten beruhenden Ergebnisindikatoren sind im Großen und Ganzen normalverteilt. Dies ist ein Anhaltspunkt dafür, dass weder in der IG noch in der KG systematisch bestimmte Messresultate ausgeschlossen wurden.

Weiterhin muss die insgesamt hohe Datenausfallrate bei der Interpretation der Studienergebnisse berücksichtigt werden sowie die Tatsache, dass von den IG-Therapeuten insgesamt deutlich mehr Patienten pro Praxis in die Studie einbezogen wurden als von den KG-Therapeuten. Hier besteht grundsätzlich die Gefahr einer hohen Selektivität der gewonnenen Daten. Auch kann nicht ausgeschlossen werden, dass die Gesamtgruppe der am Modellprojekt teilnehmenden Therapeuten aufgrund der Selbstselektion (Bereitschaft zur Teilnahme an diesem Projekt) möglicherweise nicht repräsentativ für die Gesamtgruppe niedergelassener Therapeuten ist, da das Modellprojekt möglicherweise besonders motivierte und erfolgreiche Therapeuten angesprochen haben könnte.

Zur Prüfung möglicher Effekte der hohen Drop-Out-Rate wurde die Mindesteffektgröße bei einer Power von 80% und einem Alpha-Niveau von 5% für den E-Zeitpunkt (Stichproben T2 und P2) bzw. K-Zeitpunkt (Stichprobe P3) neu berechnet. Bei 564 (IG) und 370 (KG) Patienten (Stichprobe T2) beträgt die entsprechende Mindesteffektgröße  $d = 0,19$  und für die Stichprobe P2 (IG: 383; KG: 214)  $d = 0,24$  bzw. für P3 (IG: 306; KG: 162)  $d = 0,27$ . Die Stichprobengröße ist also gerade noch hinreichend, um auch kleine Effekte beim Vergleich zwischen IG und KG in dieser Größenordnung bei einer Power von 80% zu entdecken. Die empirisch in der Studie beobachteten und statistisch nicht signifikanten Effektgrößen auf dem multiplen Ergebniskriterium zum Vergleich zwischen IG und KG liegen mit 0,10 (Stichprobe T2), 0,06 (P2) bzw. 0,15 (P3) deutlich darunter. Weiterhin zeigten die Drop-Out-Analysen zwar einen deutlichen Zusammenhang zwischen der Ausgangsbelastung zu Therapiebeginn und dem Fehlen der E- bzw. K-Messung, immerhin war jedoch der Anteil fehlender Messungen in der IG und KG, bezogen auf die Ausgangsstichproben T1 und P1, nicht unterschiedlich. Um dennoch jegliche Möglichkeit auszuschließen, dass die Studienergebnisse durch systematische Effekte zwischen Datenausfall und bestimmten Confoundervariablen verfälscht werden, wurden diese für die multiplen Ergebnisindikatoren simultan zusammen mit der Zugehörigkeit zur IG bzw. KG in multiplen Regressionsgleichungen berücksichtigt. Auch hier ergab sich keine abweichende Interpretation hinsichtlich der multiplen Ergebniskriterien im Vergleich zu einer Interpretation ohne Berücksichtigung der Confoundervariablen, d.h. es ergeben sich keine statistisch signifikanten Unterschiede zwischen IG und KG.

Die Abhängigkeit der Messungen innerhalb der Cluster wurde bei den Gruppenvergleichen berücksichtigt und die Signifikanzwerte entsprechend korrigiert. Die Tatsache, dass die Messungen der Patienten aufgrund der Clusterrandomisierung nicht vollständig unabhängig voneinander sind, ändert allerdings nichts an der Aussage, dass sich IG und KG hinsichtlich der Ergebnisqualität nicht systematisch voneinander unterscheiden.

---

## 9.5 Fazit

Wie steht es nun um den differentiellen Gewinn durch das Modellvorhaben? Es wurde ausgeführt, dass sich keine statistisch signifikanten Unterschiede zwischen IG und KG bei den multiplen Ergebnisindikatoren ergeben haben. Das bedeutet mit Blick auf das Gutachterverfahren bei der Genehmigung von Langzeittherapien: Geht man davon aus, dass es sich dabei ebenso wie beim TK-Modell um eine Form der Qualitätsbeurteilung handelt, lassen sich im Vergleich beider Vorgehensweisen keine Unterschiede in den Behandlungsergebnissen feststellen. Anders formuliert: Die Ergebnisqualität psychotherapeutischer Behandlungen ist mit und ohne Gutachterverfahren gleich hoch, ebenso mit und ohne psychometrische Messungen.

Zu beachten ist der durch psychometrische Testungen für Therapeuten und Patienten erhöhte (Dokumentations-) Aufwand in der Alltagspraxis. Eine erhebliche Rolle spielen hier u.a. die zur Datenerhebung ggf. eingesetzte Software und die zum Assessment eingesetzten Testbatterien. Im Vergleich zum bisherigen Vorgehen wäre eine durch psychometrische Messungen gestützte Routinediagnostik sicherlich vorteilhaft mit Blick auf Transparenz und Nachvollziehbarkeit der zur Bewertung des gesundheitlichen Befindens der Patienten herangezogenen Beurteilungskriterien. Führte man psychometrische Testungen in die ambulante psychotherapeutische Regelversorgung ein, wäre zu prüfen, inwieweit der zeitliche (und für die Kostenträger auch finanzielle) Mehraufwand durch Änderungen der bisherigen Antrags- und Bewilligungspraxis kompensiert werden könnte. Dies sollte ggf. unter Einbeziehung aller beteiligten Stakeholder (Krankenkassen, Therapeuten, Gutachter, Patientenvertreter usw.) diskutiert werden.

In jedem Fall lieferte das Modellvorhaben wertvolle Daten zur Versorgungsforschung, die angesichts einer sehr guten Prozess- und Ergebnisqualität der durchgeführten Behandlungen die gesamtgesellschaftliche Bedeutung der ambulanten Psychotherapie hervorheben.

---

## 10 Literaturverzeichnis

- Arbeitskreis OPD. (1996). Operationalisierte Psychodynamische Diagnostik. Grundlagen und Manual. Bern: Huber.
- Bassler, M., Potratz, B., & Krauthauser, H. (1995). Der "Helping Alliance Questionnaire" (HAQ) von Luborsky. *Psychotherapeut*, 40 (1), S. 23-32.
- Beckmann, U., Klosterhuis, H., & Lindow, B. (2007). Bewertungskonzept für medizinische Reha-Einrichtungen. In Deutsche Rentenversicherung Bund, DRV-Schriften: Tagungsband 17. Rehabilitationswissenschaftliches Kolloquium vom 3.-5. März 2008 in Bremen (S. 146-148). Berlin: Deutsche Rentenversicherung Bund.
- Brodgen, H. (1949). When testing pays off. *Personnel Psychology*, 2, 171-185.
- Bullinger, M., Kirchberger, I. et al. (1998). Der SF-36 Fragebogen zum Gesundheitszustand. Handbuch für die deutschsprachige Fragebogenversion. Göttingen: Hogrefe.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), S. 155-159.
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7 (3), S. 249-253.
- Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression / correlation analysis for the behavioral sciences* (3rd Ed.). Mahwah: Lawrence Erlbaum Associates.
- CONSORT (2010). The Consort-Statement. Internetpublikation (Zugriff am 16.05.2011). <http://www.consort-statement.org/consort-statement/overview0/>
- Cook, D., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field setting*. Chicago: Rand McNelly.
- Cronbach, L. & Gleser, G. (1965). *Psychological Tests and Personnel Decisions* (2nd ed.). Urbana, IL: University of Illinois Press.
- Ehlers, A., & Margraf, J. (2001). AKV Fragebogen zu körperbezogenen ängsten, Kognitionen und Vermeidung. Göttingen: Beltz Test Gesellschaft.
- Endicott, J., Spitzer, R., Fleiss, J., & Cohen, J. (1976). The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry*, 33, S. 766-771.
- Franke, G. (2000). Brief Symptom Inventory von L. R. Derogatis (Kurzform der SCL-90-R) -Deutsche Version. Göttingen: Beltz Test Gesellschaft.
- Glass, G., McGaw, B., & Smith, M. (1978). *Meta-analysis in social research*. London: Sage Publications.
- Grosse Holtforth, M., & Grawe, K. (2002). Bern Inventory of Treatment Goals: Part 1. Development and first application of a taxonomy of treatment goal themes. *Psychotherapy Research*, 12 (1), S. 79-99.
- Hautzinger, M., Bailer, M., Worall, H., & Keller, F. (1994). BDI. Beck-Depressions-Inventar. Bern: Huber.
- Hiller, W., Zaudig, M., Mombour, W. et al. (1995). IDCL - Internationale Diagnosen Checklisten für ICD-10 und DSM-IV. Bern: Huber.
- Horowitz, L., Strauss, B., & Kordy, H. (2000). Inventar zur Erfassung interpersonaler Probleme (IIP-D). Handanweisung (2. Aufl.). Weinheim: Beltz Test Gesellschaft.
- Howard, K., Lueger, R., Maling, M., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology*, 61 (4), S. 678-685.
- Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59 (1), S. 12-19.
- Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks: Sage Publications.
-

- Lutz, W. et al. (1997). Evaluation eines Qualitätssicherungsprogrammes in der Psychotherapie. Universität Heidelberg, Fakultät für Sozial- und Verhaltenswissenschaften. Regensburg: Roderer.
- Lutz, W., Köck, K., Böhnke, J. et al. (2009). Die Wirkung von Rückmeldesystemen aus ambulanten Settings: Das Modellvorhaben zur Psychotherapie der Techniker Krankenkasse und Wege in die stationäre Praxis. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 22, S. 118-125.
- Lutz, W., Tholen, S., Kosfelder, J. et al. (2004). Modelle und Konzepte zur empirischen Unterstützung der Evaluation des Psychotherapiebedarfs und der Therapieverlängerung. *Verhaltenstherapie und Verhaltensmedizin*, 25 (4), S. 503-527.
- Lutz, W., Tholen, S., Kosfelder, J., Grawe, K., Schulte, D. et al. (2005). Zur Entwicklung von Entscheidungsregeln in der Psychotherapie. Die Validierung von Vorhersagemodellen mit einer sequenzanalytischen Methode. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 34 (3), S. 165-175.
- Paul, T., & Thiel, A. (2005). EDI-2 Eating Disorder Inventory. Manual. Göttingen: Hogrefe.
- Rief, W., Hiller, W., Heuser, J. et al. (1997). SOMS Das Screening für somatoforme Störungen. Bern: Huber.
- Rossi, P., Freeman, H., & Lipsey, M. (1999). *Evaluation: A systematic approach* (6th Edition). Thousand Oaks: Sage.
- Schepank, H. (1995). Beeinträchtigungs-Schwere-Score (BSS). Ein Instrument zur Bestimmung der Schwere einer psychogenen Erkrankung. Göttingen: Hogrefe.
- Schmidt, F., Hunter, J., & Pearlman, K. (1982). Assessing the economic impact of personnel programs on workforce productivity. *Personnel Psychology*, 35 (2), S. 333-347.
- Schmidt, J. (1991). Evaluation einer Psychosomatischen Klinik. Frankfurt a. M.: VAS - Verlag für Akademische Schriften.
- Schmidt, J., Bernhard, P., Wittmann, W., & Lamprecht, F. (1987). Die Unterscheidung zwischen singulären und multiplen Ergebniskriterien. In F. Lamprecht, *Spezialisierung und Integration in Psychosomatik und Psychotherapie*. Deutsches Kollegium für psychosomatische Medizin, 6.-8. März 1986 (S. 293-299). Berlin: Springer.
- Schmidt, J., Lamprecht, F., & Wittmann, W. (1989). Zufriedenheit mit der stationären Versorgung. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 39 (7), S. 248-255.
- Schmidt, J., Steffanowski, A., Nübling, R., Lichtenberg, S., & Wittmann, W. (2003). Ergebnisqualität stationärer psychosomatischer Rehabilitation. Regensburg: Roderer.
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Co.
- Steffanowski, A., Nübling, R., Schmidt, J., & Löschmann, C. (2006). Patientenbefragungen in der medizinischen Rehabilitation - Computergestütztes Routinemonitoring der Struktur-, Prozess- und Ergebnisqualität. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 71, S. 35-46.
- Steffanowski, A., Löschmann, C., Schmidt, J., Wittmann, W. W., & Nübling, R. (2007). Meta-Analyse der Effekte stationärer psychosomatischer Rehabilitation - MESTA-Studie. Bern: Huber.
- WHO. (2010). ICD-10. Genf: World Health Organization.
- Wittchen, H., Lieb, R., Wunderlich, U., Schuster, P. et al. (1999). Comorbidity in primary care: Presentation and consequences. *Journal of Clinical Psychiatry*, 60, S. 29-38.
- Wittchen, H., Nelson, C., Lachner, G. et al. (1998). Prevalence of mental disorders and psychosocial impairments in adolescents and young adults. *Psychological Medicine*, 28, S. 109-126.
- Wittmann, W. (1990). Brunswik-Symmetrie und die Konzeption der Fünf-Datenboxen. *Zeitschrift für Pädagogische Psychologie*, 4 (4), S. 241-251.
-

Wittmann, W. (1995). Evaluation in der Rehabilitation: Methoden, Ergebnisse und Folgerungen für die Praxis. In K. Siek, F. Pape, W. Blumenthal, & M. Schmollinger, Erfolgsbeurteilung in der Rehabilitation - Begründungen, Möglichkeiten, Erfahrungen (S. 77-88). Ulm: Universitätsverlag.

Wittmann, W., & Walach, H. (2001). Evaluating complementary medicine: Lessons to be learned from evaluation research. In G. Lewith, W. Jonas, & H. Walach, Clinical research in complementary therapies: Principles, problems and solutions (S. 93-108). London: Churchill Livingstone.

Wittmann, W., Nübling, R., & Schmidt, J. (2002). Evaluationsforschung und Programmevaluation im Gesundheitswesen. Zeitschrift für Evaluation , 1, S. 39-60.

Zaworka, W., Hand, I., Jauernig, G., & Lünenschloß, K. (1993). HZI - Testmanual. Göttingen: Beltz Test Gesellschaft.

Zielke, M. (1993). Wirksamkeit stationärer Verhaltenstherapie. Weinheim: Psychologie Verlags Union.

Zielke, M., Borgart, E. J., Carls, W., Herder, F., Lehenhagen, J., Leidig, S., Limbacher, K., Meermann, R., Reschenberg, I., & Schwickeratz, J. (2004). Ergebnisqualität und Gesundheitsökonomie verhaltensmedizinischer Psychosomatik in der Klinik. Lengerich: Pabst.

## 11 Anhang: Fragebogen Nichtteilnehmer - Sonstige Gründe für die Nichtteilnahme am Modellprojekt aus Therapeuten- und Patientensicht im Original-Wortlaut

### Sonstige Gründe für Nichtteilnahme aus Therapeutensicht

1998 bis 1999 VT bei mir. Jetzt Dgn, Rezidivierende depressive Störung. Antidepressive Medikation ausreichend

ab 8 / 2008 meinerseits 1 J. Erziehungszeit; deswegen unklar, ob Behandlung von mir beendet werden kann oder vom Entlastungsassistenten übernommen werden muss

ab 8/2008 meinerseits 1 Jahr Erziehungszeit, deswegen unklar, ob Ende der Behandlung nicht von jemand anderem übernommen werden muss

aktuell PT nicht sinnvoll, erst Stabilisierung bezügl. Der Suchtproblematik (innerhalb 3 J. 10 Entgiftungen, die letzte vor 2 Wochen)

Alkoholiker, nicht trocken, muss erst trocken sein.

Ambivalenz des Patienten (Therapie ja/nein. Überzeugung, die Beschwerden seien somatisch bedingt an VT überwiesen, meinerseits psychodynamisch arbeitend

Anfängliche Zustimmung mit Unterschrift unter Modell, Ausfüllen der Fragebögen, Genehmigung durch TK, Inanspruchnahme von 2 Sitzungen, dann Bruch, da P. erzwingen wollte, dass der Therapeut in einem Begehren nach Be.. (unleserliches Wort) nach außen Partei ergreifen sollte.

Aufgrund der Psychopathologie: Bericht wäre erforderlich. Die Studie ist schlecht geplant (Versuchsplanung, Testauswahl etc). Mit den Tests können nur wenige Störungsbilder testtheoretisch sinnvoll erfasst werden. Dieser Fall gehört nicht dazu

aufgrund des depressiven Zustandsbildes nicht in der Lage

aufgrund eines Trauerfalls in der Familie schien es unangemessen, TK-Modell zeitnah anzusprechen

Aufwand: Therapieort, zukünftiger Wohnsitz, ungeklärte soz. Situation

Augenblicklich im Vordergrund Panikstörung. Patientin aufgrund alter Erfahrung sehr misstrauisch. Gefahr einer weiteren psychischen Dekompensation zu groß

Bei Pat. Liegt starker Verdacht auf Alkoholabhängigkeit vor z.Zt. Keine affektive Störung. In den störungsspezifischen Tests bzw. Störungstyp fehlen mir die Suchterkrankungen!

Borderline Patientin

Bulimie-Patientin. 17 3/4 Jahre alt. Ist zur Therapie nicht motiviert. Es haben Gespräche mit der noch erziehungsberechtigten Mutter stattgefunden. Diese ist auch sehr belastet und wird jetzt Therapie machen. HE137-04

dependente Persönlichkeit. Legt alles der Mutter vor, bzw. lässt diese entscheiden. Pat. Hat Therapie nach 5 Sitzungen beendet.

Der Ehemann der Patientin befindet sich in meiner Behandlung. Das Ehepaar befindet sich in Trennung (Scheidung steht bevor)

Der Patient hat sich die Finger der rechten Hand abgetrennt und war nicht in der Lage, Stift/Maus zu führen. Zwischenzeitlich Reha.

---

Der Patient ist bei mir bereits in Psychotherapie gewesen. Ein paar kurze Gespräche und Antidepressivum waren ausreichend

Die gesamte Familie ist mir aus langen, zurückliegenden Behandlungen bekannt.

Die Hauptsymptomatik (ADHS) kann mit den FB der TK-Studie nicht erfasst werden

Die Hauptsymptomatik (ADHS) kann mit den FB der TK-Studie nicht erfasst werden

Die Hauptsymptomatik (F60.6; F60.7; F40.1) kann mit den FB der TK-Studie nicht erfasst werden

Die Patientin ist blind. Aktuell habe ich keinen Praktikanten, der mit der Patientin die Fragebögen bearbeitet. Mir ist der Aufwand -offen gesagt- zu groß.

Die Patientin hatte von 2004 - 2006 bereits Kontakt zu mir (frühere psychotherapeutische Behandlung) durch längere Krankheit des Therapeuten hatte ich den Pat. Verloren

ein Fall für den Psychiater

Erkrankungsbild Alkoholabhängigkeit fehlt in dem Programm

es nicht ersichtlich ist, ob die Therapie bis zum Ende durchgeführt werden kann. Compliance und Motivation nicht so stark

Es waren am Anfang zu viele formale Hürden, so dass die Therapie endlich begonnen werden musste ohne weitere Formulare (fehlender Konsiliarbericht, keine Überweisung etc)

Fall für den Psychiater

Fremdmotivation, Grfa. Introspektionsfähigkeit

Fühlt sich im Moment in allem überfordert

habe Patientin Teilnahmemöglichkeit vorgestellt. Aufgrund schneller Besserung der symptomatik nach 4 Prob. Stunden keine Therapieantragsstellung

Hat bereits Therapeutin (fragl. Ambivalenz, fragl. Agierter Konflikt, fragl. Dissoziative Verarbeitung

Hat in einer früheren Psychotherapie Fragebögen (zu Beginn jeder Sitzung) als störend erlebt, möchte sich ausschließlich auf Gespräch konzentrieren und sieht das bei Teilnahme nicht gewährleistet.

Hinweis auf narzißtische Persönlichkeitsstörung mit schizoidem Modus

Ich nehme regelmäßig an einer interdisziplinären Schmerzkonferenz teil. Ich habe den Patienten nur für 3 prob. Sitzungen gesehen, um eine psychologische Diagnostik vorzunehmen, weil der Kollege (Facharzt f. Schmerztherapie) mich darum gebeten hat. Der Patient bedarf zwar einer psychotherapeutischen Behandlung, will diese aber erst nach Ausschöpfung anderer Behandlungsangebote wie Akupunktur, medikamentöse Behandlung, angehen.

In der 2. prob. Sitzung berichtet die Pat., dass sie ihren Mangel an sexuellem Verlangen mit einer weiblichen Therapeutin besprechen, bzw. behandeln möchte.

insgesamt 4 stationäre und 3 ambulante Psychotherapien => Psychotherapie wahrscheinlich ohne Aussicht auf ausreichenden Erfolg

Kein Problembewußtsein

Keine krankheitswertige Störung, eher Partnerschaftskonflikte

klärendes Gespräch (1x) Erstgespräch

Klientin hat die Therapie nach der 2. probatorischen Sitzung abgebrochen, bevor ich ihr die Fragebögen aushändigen konnte

Krankheitsbild zu schwer

KZT vorauss. Ausreichend (15-20 Sitzungen) Frequenz 1x monatlich, die Therapiedauer dürfte wohl zu lang sein

---

KZT wahrscheinlich ausreichend. Beantwortung nur formal und ohne TK-Unterlagen und ohne Gutachterverfahren einfach schneller.

lediglich Krisenintervention notwendig gewesen

mangelnde Zuverlässigkeit

neue Indikation für Psychotherapie

Neuroprobatorische Sitzungen erforderlich

Notwendige Intentionen zu Therapiebeginn würden gestört / verzögert

nur 3 probat. Sitzungen, d. h. keine Therapie

Nur eine prob. Sitzung kam zustande. Keine Therapie, weil Patientin eine Arbeitsstelle an einem anderen Ort bekam und deshalb einen anderen Therapieplatz ausgesucht hat.

Paartherapie empfohlen

paranoide Struktur: Misstrauen gegenüber Datenschutz etc.

Pat. Bereits in Psychotherapie bei einer anderen Therapeutin

Pat. Ging in stationäre Aufnahme (Klinik).

Pat. Hatte vor 18 Jahren eine Therapie bei mir. Kommt jetzt für einige Stunden im Rahmen eines KZT.

Pat. ist Ausländerin, Schwierigkeiten mit der Schriftsprache

Pat. Kam auf Empfehlung der Arbeitsämter, Ambivalenz bezügl. Der eigenen Beschwerden, Ambivalenz in Veränderung der Lebenssituation

Pat. Kam in absoluter Krisensituation, Sitzungen sehr unregelmäßig aufgrund berufl. Belastung. Pat. Ist extrem unruhig, nervös, angespannt. Anamneseerhebung sehr erschwert.

Pat. Kam mit ambivalenter Haltung

Pat. Kam zu einer Erstsitzung, die nachfolgende absagend.

Pat. Kommt nach abgeschlossener Therapie (2005) erneut zur Behandlung.

Pat. Lehnte eine weitere Behandlung (nach 2 prob. Sitzungen) ab, da sie das Ausfallhonorar nicht zahlen wolle.

Pat. Plant event. Umzug ins Ausland in 1 - 2 Jahren. Würde damit u. U. das Modellvorhaben nicht komplett durchhalten.

Pat. Unentschlossen, ob er den vorgegebenen Weg gehen wolle. Wollte sich noch einen anderen Th. Ansehen

Pat. Wüsste nur Behandlung im Rahmen einer Kurzzeittherapie

Pat. Wurde rekrutiert, die Therapie aber nach 3 Sitzungen durch mich abgebrochen wegen Gegenübertragungsphänomenen

Patient befindet sich im Finalstadium einer Krebserkrankung

Patient hat nach probat. Sitzung vor 2 Jahren zunächst Behandlung abgebrochen - traumatisierte Patientin

Patient nahm bereits als Partner an Sitzungen mit inzwischen verstorbener Ehefrau teil

Patient war schon früher bei mir in therapeutischer Behandlung. Laut Auskunft der TK sollen nur NEUE Patienten in die Studie einbezogen werden.

Patientin "testet" Therapeuten und war nur zu einem Gespräch in der Praxis.

Patientin ist gebürtige Chinesin und sprachlich den Fragebögen nicht ganz gewachsen

---

Patientin kam im Rahmen eines Rentenbegehrens => Therapie hat wenig Aussicht auf Erfolg.

Patientin mit seit Jahren bestehender Alkoholabhängigkeit Zunächst sollen probatorische Sitzungen genutzt werden, um Patientin auf dem Weg zu einer stationären Behandlung zu begleiten und zu stützen, sowie zur Krisenintervention. Es ist schwer einzuschätzen, wie lange es dauert, bis die Patientin die stationäre Behandlung abgeschlossen hat und eine ambulante Nachsorge / Therapie beginnen kann.

Patientin wird nur vorübergehend (ca. 2 - 3 Stunden) begleitet - stat. Therapie geplant

Patientin wollte nur Informationsgespräch. Kam auf Drängen der Mutter.

Patientin wurde zwar informiert, aber es wurde dann keine Therapie begonnen. Therapeut vermisst die notwendige Motivation bzw. Retrospektionsfähigkeit.

Sprachliche Überforderung, da schlechte Deutschkenntnisse

Störungsbild erfordert alternative diagnostische Instrumente und Patientenschulung; alles zusammen würde zu viel werden.

Therapiekontakt kam nicht zustande

Therapiemotivation fehlte. Schon im 1. Kontakt wurde deutlich, dass keine Therapie folgen würde.

Unstimmigkeiten bezüglich des Evaluationsplanes

voraussehbar eher Beratungsgespräche

VT-Empfehlung

VT-Empfehlung , übende Behandlung der Krankheitsbewältigung - Krebs

Wahrscheinlich wird es nur eine kleinere Anzahl von Krisengesprächen geben

Wegen Ortswechsel wird eine längere Behandlung nicht möglich sein

weil die Patientin vor ihrem Umzug nach Dortmund direkt eine Verhaltenstherapie gemacht hatte und ich nun quasi eine Verlängerung / Weiterführung der Therapie durchführe, was die Daten verfälschen würde.

weil die zugesagten Schulungen für Teilnehmer der KG nicht durchgeführt werden

Weil Therapie-Motivation und Compliance noch fraglich sind

wiederholte Krisenintervention bei lange bekanntem Pat.

wiederholte Terminabsage

Zu Beginn war nicht deutlich, ob der Patient aufgrund des Zustandes einen stationären Aufenthalt beginnen sollte; nach Absprache mit dem Hausarzt verzögerte ambulante Therapie

zukünftiges Wohnen nicht geklärt (=> unklare Behandlungsvoraussetzungen)

Zweifel an der Zuverlässigkeit des Patienten

---

### **Sonstige Gründe für Nichtteilnahme aus Patientensicht**

"Das ist sehr vertraulich." "Da gebe ich mich ja vollständig preis." "Ich habe hiervon keine Vorteile."

"Muss ich daran teilnehmen ?" Pat. Bat um Nichtteilnahme nach Ausfüllen der Fragebögen - wurde ihm zu heiß ? D. h. Beantwortung überforderte ihn.

1) da die Patientin extrem belastet und erschöpft ist, lehnt sie die Möglichkeit ab

2) fürchtet sie, dass die Kasse darüber Informationen erhält, die später negativ gehandelt werden  
aktuelle Problematik, Belastbarkeit

akute Situation (lebensbedrohende Krankheit)

Akute Trauerreaktion überfordert die diagnostische Teilnahme

akute Trauerreaktion, fühlt sich zu belastet, scheut zusätzlichen Aufwand

aufgrund ihrer Problematik braucht sie mehr Zeit zum Vertrauensaufbau

Aufgrund von Prüfungssituation derzeit keine zeitlichen Kapazitäten

Aufgrund bisher schlechter Therapieerfahrung mit männlichen Therapeuten sehr misstrauisch. Hintergrund sexuelle Traumatisierung

Befürchtet Kontrolle und negative Einflussnahme von draußen

befürchtet Verstärkung der Symptome

Bereits Therapie bei mir in 1999/2000 gemacht

Burn out, völlig überfordert im Alltag, Fragebögen sind ihr jetzt zu viel

Dem Patienten die Teilnahme offeriert. Er versprach, sich nach Klärung seiner Lebenssituation wieder zu melden (vor 6 Wochen). Bisher keine Reaktion

Deutschkenntnisse reichen nicht aus

Die berufliche Überlastung des Patienten ist so hoch, dass er sich zu stark belastet fühlt, auch noch an einer Studie teilzunehmen. Er gab die Unterlagen zurück

Die Menge der Fragen überforderten die P., die sehr erschöpft am Rande eines depressiven Syndroms war

Die Pat. War durch die akute Situation (Verkehrsunfall mit Todesfolge) so belastet, dass sich nicht in der Lage sah.

Die Patientin kam wegen unerfüllten Kinderwunsches. Nach zwei Sitzungen meldete sich das Jugendamt wegen eines Adoptionswunsches. Die Patientin hatte daraufhin Angst, die Therapie könnte ihr als Unfähigkeit ausgelegt werden. Daher kein Antrag.

Einschätzung des Therapeuten: die Patientin hat Lebenskonflikte, ist psychisch aber weitgehend gesund und durch zwei probatorische Sitzungen vor dem 26.11.2005 schon gestärkt. Ein Therapieantrag kommt wegen der "Gesundheit" nicht in Frage

erlebt im Moment alles als Überforderung

erlebte die Erstberatung so erfolgreich, dass sie keine weitere Therapie will

Es sei ihr zu viel Arbeit, zu viel zusätzliche Zeit

extreme zeitliche Beanspruchung / Überforderungssymptomatik

Findet das jetzige System gut und möchte keine Änderung

findet diese Art der Befragung nicht zweckdienlich

---

Findet Fragebögen zu undifferenziert. Hat schlechte Erfahrungen mit der TKK gemacht. Warum sollte sie der Kasse dann entgegenkommen ?

Fragebögen sind nicht geeignet Therapienotwendigkeit und Veränderungsprozesse in der Therapie zu erfassen. Sie eignen sich nicht die Qualität eines Behandlers zu erfassen - hier ist auch weiterhin die Einschätzung des Patienten notwendig.

Fühlt sich überfordert, möchte die Sitzungen nur für sich nutzen

fürchtet, dass die Daten doch für andere Zwecke der Kasse missbraucht werden könnten, z.B. Ablehnung der Rente

Geht`s schon wieder besser

geplanter Ortswechsel

ging 1/4 Jahr später ins Ausland und wollte sich damit nicht zusätzlich belasten

Habe neue Fragebögen erst vor einer Woche erhalten, da war die Therapie schon genehmigt.

habe so viel um die Ohren

hält Fragen für indiskret, insbes. Fragen der Basisdoku nach Einkommen, hält das Ganze für "Krankenkassenschnüffelei" und möchte, da er kaum sonst noch seine Arbeit schafft, sich nicht noch mit Fragebögen beschäftigen

Hat 100%gen Pflegefall (Ehemann) zuhause (Tumor im Kopf + halbseitige Lähmung) und gleichzeitig 8 Monate altes Baby zu versorgen. Könne sich daher nicht auch noch auf den Fragebogen geistig und emotional einlassen.

Hat bereits während der Reha viele Fragebögen ausfüllen müssen

hat prinzipielle Einwände, wie z.B.: "Ich will mich nicht in eine Norm pressen lassen." "Ich will mich nicht an der Ökonomisierung im Gesundheitswesen beteiligen." u.a.

Hat schlechte Erfahrungen mit so etwas gemacht.

hat sich gegen eine Psychotherapie bei mir entschieden

Hat sich nach Vorschlag, in die TK-Studie aufgenommen zu werden, nicht mehr gemeldet.

Hat zu viel mit sich selbst zu tun, fühlt sich überfordert

Hatte zunächst zugesagt, später Termin abgesagt. Wird Therapie nicht machen.

Hohe Zweifel, überhaupt eine Psychotherapie in Anspruch zu nehmen.

Ich habe lange nichts mehr vom Projekt gehört, was zur Folge hat, dass die eigene Motivation Patienten zu rekrutieren, deutlich sinkt.

Ich hatte die Teilnahme offeriert. Es blieb aber bei den probatorischen Sitzungen.

Ich kenne als Geschäftsführer den sicheren Umgang mit Daten: Nein!

Ist dem Patienten z. Zt. Zu viel

Ist hoch schwanger, Entbindung am 22.2.

Ist nur einmal zum Gespräch erschienen, keine Therapiefortsetzung.

Ist was Neues - kommt zu plötzlich... ist alles zu viel auf einmal

ist zu sehr in ihren Gedanken gefangen, meint, sie könne sich nicht konzentrieren, finde keine Antwort

Kann sich derzeit nicht mit dem Gedanken anfreunden, sich nach Abschluss der Therapie nochmals mit der belastenden Problematik befassen zu müssen (Traumapatientin) einfühlbar.

Kassenwechsel geplant

Kein Interesse

---

keine längere Therapie nötig

Klärung ihrer Lebenssituation; Zweifel, ob ihr eine Therapie helfen könnte.

Lehnt "Einmischung von Außen" in den Therapieverlauf ab, auch wenn sie als KG-Teilnehmerin davon zunächst nicht betroffen wäre.

Lehnt aktuelle Entwicklung im Gesundheitswesen ab. Befürchtet weitere negative Auswirkungen und Kontrolle durch Studie. Nach der 2. prob. Sitzung hat sich die Patientin gegen eine Psychotherapie bei mir entschieden.

Mir geht es schon so schlecht, wenn ich jetzt noch 150 Fragen beantworten soll, stürze ich ganz und gar ab.

Misstraut der Zielsetzung. Will nach Abschluss der Therapie nicht mehr damit behelligt werden

Möchte ihre Energie darauf verwenden gegen die Krankheit zu kämpfen, sich auf sich zu konzentrieren, ist ihr zu viel

möchte nicht "in Tabellen" eingeordnet werden

musste in der Klinik viele Fragebögen ausfüllen.

Muttersprache nicht deutsch sondern serbokroatisch

Nach erster Vorinformation über TK-Modell meldete sie sich nicht mehr.

Nach vorheriger Zusage (2. Sitzung): Absage / Abbruch vereinbarter Therapie (anderen Therapeuten gefunden)

Nicht mehr gekommen

paranoide Verarbeitung

Pat. Befürchtet eine standardisierte Behandlung zu bekommen und in Schubkästen eingeordnet zu werden

Patient wünscht Verhaltenstherapie

bereit im Erstgespräch. Ambivalenz gegenüber PSTH

keine Ablehnung des TK-Modells (möglicherweise wird sich der Pat. An die TK wenden, um VT-Therapeut zu finden)

Patient erhoffte sich Vorteile (von TK bevorzugt behandelt zu werden

Patientin fühlt sich durch die Fragebögen unter Druck gesetzt

Patientin hat zunächst Einverständniserklärung am 20.12.2005 unterschrieben, nach Aushändigung der Fragebögen aber abgelehnt und ist nach der 3. probatorischen Sitzung nicht mehr erschienen.

Patientin leidet an einer Augenkrankheit. Dachte erst, sie könne die Fragebögen mit ihrem Sichtgerät lesen. Das war dann doch zu aufwändig.

Patientin mit langjährigem Missbrauch in der Familie !

Patientin: "Ich mache solche Dinge nur, wenn es der Forschung dient - hier geht es um andere Dinge." Die vorgeschlagenen Alternativen sind sehr tendenziös. Patientin ist wissenschaftliche Angestellte der Universität!

Plant Krankenkassenwechsel

Prüfungsstress

schafft keinen zusätzlichen Aufwand

Scham wegen Symptomatik

Schwere Depression. Alle Tätigkeiten fallen schwer.

---

Schwere depressive Episode. Trotz vorh. Stat. Therapie ist Pat. Kaum in der Lage, die normalen Antragsformulare zu bearbeiten. Fühlt sich überfordert.

Sie kannte die Möglichkeit verschiedener therapeutischer Möglichkeiten nicht.

Sie entschied sich für eine VT, nicht für eine Tiefenpsychologie (wie bei mir).

Sie will keine zusätzliche Belastung

sieht für sich keinen direkten Vorteil durch die Teilnahme

Sohn ist vor 2 Monaten gestorben

Sprachprobleme (Muttersprache polnisch)

störungsbedingte Ängste

Suizidalität. PT - Beginn als Krisenintervention

Telefonische Absage ohne Begründung

Therapie wurde nicht begonnen, daher auch keine Teilnahme am TK Modellversuch.

Therapieabbruch nach 1 Behandlungsstunde, hohe Ambivalenz gegenüber Veränderungen

Therapieabbruch. Grund: bisher verschwiegener Alkoholabusus mit Dekompensation, stationärer Entzug nötig. Pat. Will PT nicht fortsetzen.

Therapieunsicherheit

traut es sich nicht zu, angst, sich zu überfordern, habe ihr Nein akzeptiert

UMSCHRIEBENS Problem

Umzug und wenig Zeit; Krise!

Vermutung der Therapeutin: Patientin fühlt sich nicht fähig

War durch die Symptomatik zu sehr belastet, sich auch noch darauf zu konzentrieren.

wechselt zur P/T - Behandlung

wird beruflich bedingt (nach Kündigung) ggf. in eine andere Stadt ziehen

Wollte eine tiefenpsychologisch orientierte Psychotherapie - Ende nach 2 probatorischen Sitzungen

wollte erst teilnehmen, hat aber dann die PT nicht aufgenommen, weil sich seine Lebenssituation wieder stabilisiert hatte

Wollte keine Psychotherapie. Die Betreuerinnen befürworten dies jedoch. Ende nach 1 probat. Sitzung

Zieht demnächst an einen entfernten Ort (Therapeutenwechsel absehbar)

Zitat: " Ich bin zu alt für sowas..."

zweifelnd, ob überhaupt Psychotherapie; ggf. VT in der Überlegung des Patienten

---

## 12 Anhang: Offene Antworten zu den Evaluationsfragen aus Patienten- und Gutachtersicht im Original-Wortlaut

### Evaluationsfragen aus Patientensicht:

#### Haben Sie weitere Anregungen, Kritik oder Fragen (zum Modellvorhaben der TK)?

Ambulante Psychotherapie finde ich besser als stationäre. Man bleibt im Leben und wird nicht isoliert. Es kommt natürlich auf den Fall an.

Anonymität ist nicht gewahrt, da ich pers. Angaben machen muss (Geburtsdatum)

[Assessment]: 1. Zu viele Fragebogen. 2. Oft widersprüchliche Fragen 3. Keine eindeutigen Fragen, oft Doppelfragen; 4. Kein Ergebnismuster; 5. Zu akademisch; 6. Wo bleibt der Patient ? Was sollen die Fragen 15 H -I?

Beantwortung erübrigt sich, da nicht benutzt.

Bei 3 Fragen war ich mir nicht sicher, ob die Antwort darauf sich positiv oder negativ bezogen hat. Ansonsten alles prima. Fand auch gut, die Ergebnisse des Tests auf dieser Metaebene zu besprechen

Bei akuten Problemen schneller Termine zu bekommen.

Bei der 1. Befragung war ich unsicher, ob die Fragebögen alle zu Beginn der Therapie ausgefüllt werden sollten.

Bei meiner Problematik waren viele Fragen überflüssig. Man müsste Fragen besser mit der seelischen Problematik, wie z.B. Panik oder Angst usw. mehr differenzieren. Teilweise waren die Fragen zu allgemein und gar nicht auf meinen Fall abgestimmt.

Bekomme ich ein Feedback?

Bekommen die Therapeuten ein Feedback? Was passiert, wenn ein Therapeut öfters nur schlecht bewertet wird und eine Behandlung nicht hilft?

Bis auf das Ausfüllen der Fragebogen hatte ich keine konkreten Vergleiche. Diese Frage (Nr. 12) kann ich nicht beantworten.

Da meine Probleme nur bei ganz bestimmten Situationen auftreten, die oft nicht die letzten 7 Tage betreffen, deckten die Fragebögen diese fast nie ab. Laut Fragebogen habe ich dann keine Probleme

Das Beantworten der Fragen zum Monitoring ist zu lange her; ich hatte meine Anregungen aber an die Therapeutin weitergegeben. Es fehlen m. E. bei einigen Fragen relevante Antwortalternativen.

Das Ende / Abschluss der Therapie kam mir so abrupt vor. Kann man diesen Ablösungsprozess nicht durch Rituale / Symbole unterstützen?

Dass über eine Verlängerung der med. Dienst entscheiden muss. Meine 80 Stunden sind aufgebraucht und jetzt stehe ich ziemlich im Regen. Dies sollte man doch den behandelnden Ärzten und Therapeuten überlassen und nicht Personen, die einen gar nicht kennen und null Ahnung von meinen persönlichen Lebensumständen haben.

Der allgemeine Fragebogen zur Persönlichkeit und zum Gesundheitszustand ist zu lang.

Der Beantwortung der Fragen sollte unbedingt noch ein Gespräch mit dem Therapeuten folgen, damit freigesetzte Emotionen noch verarbeitet werden können, d. h. Fragen zum Beginn der Therapiestunde beantworten.

Der Therapeut ist ein prima Psychologe, mir geht es seit der Therapie richtig, richtig gut.

---

Die ambulante Psychotherapie war für mich sehr hilfreich. Das Monitoring gibt zusätzliche Sicherheit, auf dem richtigen Weg zu sein und unterstützt somit das Therapieziel.

Die Antwortmöglichkeiten sind nicht spezifisch genug.

Die Auswahlmöglichkeit an Antworten ist bei den Fragebogen zu gering

Die Auswertungen hätte ich gern schriftlich erhalten mit einer Empfehlung der TK, damit ich mehr damit anfangen kann

Die Bögen sind in wechselndem Fragestil. Da wäre Einheitlichkeit besser

Die Fragebögen des Programms haben verhindert, dass der Therapeut mit seinen eigenen Fragebögen eine Diagnose stellt und diese behandelt.

Die Fragebögen sollten für vor und nach der Therapie unterschiedliche Fragen enthalten. Z.B. BDI -> habe ich das Interesse an Mitmenschen verloren oder wie gut kann ich arbeiten Hinterher: kann ich jetzt besser arbeiten ...

Die Fragen 11a und b kann ich nicht beantworten, weil mich keine Vergleiche habe, z.B. von Behandlungsformen. Bei mir hat das Therapeuten-Patientenverhältnis von Anfang an gepasst. Für mich war es der richtige Weg.

Die Fragen betrafen häufig nicht mein Problem

Die Fragen deckten das Spektrum meiner Störung nicht ab. Zu allgemein ! Zu wenig individuell !

Die Fragen fand ich teilweise zu schwierig formuliert, bzw. ich sehe keinen eindeutigen Unterschied

Die Fragen im ersten Bogen BDI finde ich viel zu einseitig und haben wenig mit meinem Denken gemeinsam, mein Kopf ist viel komplizierter. Ich denke, Krebs(oder ähnliches) Patienten denken oftmals anders!

Die Fragen sind teilweise unkonkret, z.B. habe ich mich trotz psychischer Probleme nie "krank" gefühlt.

Die Fragen sollten immer für einen Zeitraum von einer Woche beantwortet werden. Zu kurz gefasst. Was ist bei Krankheit, schlechter Woche? Verfälscht das Ergebnis. Therapieerfolg, Besserung der Situation ist ja auch trotz "Erkältung + Unwohlsein" möglich.

Die Fragen sollten konkreter formuliert und der situative Rahmen sollte berücksichtigt werden (z.B. emotionale Stresssituationen wie Jobwechsel, Trauerfälle etc.) da diese die Ergebnisse z. T. erheblich beeinflussen.

Die Fragen überschneiden sich zum Teil sehr oft, sind insgesamt zu zahlreich und die Antworten decken oft nicht alle Möglichkeiten ab.

Die Maßnahme ist zeitlich begrenzt, was ist mit den Menschen, die das Programm ansprechend finden, sich aber nach Abschluss der Maßnahme wieder mit ihren Problemen überfordert, bzw. alleingelassen fühlen, wo bekommen sie Hilfe, wie geht es weiter ? Wo, wie bekommen sie Hilfe ?

Die Möglichkeit zu haben, etwa 3 - 4 Sitzungen nach Beendigung der Therapie über einen längeren Zeitraum hinweg wahrnehmen zu können. Dies würde den Therapieerfolg festigen (Zeitfenster alle 3 Monate eine Sitzung). In Bezug auf Nachhaltigkeit eine sinnvolle Maßnahme.

Die Nachsorge / Perspektive ist für mich unklar, bzw. nicht ausreichend aufgrund der Beschränkungen der TK

Die offenen Fragen kann ich nicht beantworten. Dies war meine erste Therapie und ich bin sozusagen wissenlos hineingegangen. Meine Gedanken kann ich nicht in einem Kreuz zum Ausdruck bringen. Da wäre ein offenes Gespräch, vielleicht auch mit dem Therapeuten, sinnvoller.

Die Wartezeiten sind oft zu lange, um einen Platz bei einem guten Therapeuten zu bekommen.

Die Wirkung der Therapie hat schnell nachgelassen. Diese Methode ist wirklich NUR für den akuten Fall.

---

Die zeitliche Begrenzung von 7 Tagen bei der Feststellung der emotionalen Befindlichkeit halte ich für sehr gewagt, da prozesshafte Veränderungen oft mit starken Schwankungen einhergehen und reale Empfinden und Veränderungen nicht wirklich verdeutlichen können.

Diese psychotherapeutische Behandlung hätte ich schon vor 35 Jahren und bei diesem Therapeuten machen sollen. Dann wäre mir viel Seelenschmerz und gesundheitlicher Schaden erspart geblieben.

Durch das Modellvorhaben wurde mir in einer schwierigen Lebensphase gut und vor allen Dingen schnell geholfen. Vielen Dank!

Durch die Befragungen an dem Computer gingen mir einige Stunden verloren. Ich fände es besser, wenn es nicht meine Stunden ersetzen würde.

Durch noch mehr Kontrolle (z.B. durch Anwesenheitsliste) könnte die Wahrscheinlichkeit einer positiven Entwicklung der Therapie vielleicht erhöhen.

Ein Mensch fühlt sich selten jeden Tag gleich. Die Antworten gelten nur für den Moment wo ich sie ankreuze.

Eine Psychotherapie über Fragebögen zu messen ist schwierig, weil in diesem Bereich auch das Zwischenmenschliche eine Rolle spielt. Was für den einen Patienten gut ist, ist für einen anderen nicht so vorteilhaft. Ich bin an dem Ergebnis der Studie interessiert. Leider erfährt man darüber nicht so viel.

Anmerkung: Therapie bei der Therapeutin nach 6 Sitzungen abgebrochen. Fortsetzung der Therapie bei einer anderen Therapeutin.

Einige Fragen sind zu allgemein gehalten und deshalb schwer zu beantworten

Ergänzend zu den Fragebögen sollten die Veränderungen im Zuge der Therapie mündlich erörtert werden (jedenfalls nicht in Form einer Besprechung des Fragebogens) 2. Einige Fragen, zu bewertende Punkte sind diffus, vom Wortlaut mehrdeutig. Akute Lebenssituationen sollten Berücksichtigung finden (z.B. bin ich seit kurzem Vater, was den Stress-, aber auch Wohlfühllevel deutlich beeinflusst)

Es kommt oft vor, dass Menschen mit psychischen Problemen zur Appetitlosigkeit neigen, aber es gibt auch Menschen, die das Essen erst richtig anfangen und diese Richtung kommt in den Fragen gar nicht vor.

Es sollte zwischendurch auch noch per Fragebogen der Zustand und Erfolg überprüft werden.

Es sollten ruhig auch mal Gespräche mit TK-Experten und Therapeut / Patient stattfinden.

Es wäre gut, wenn von Ihnen alle drei Monate ein Fragebogen käme

Es wäre sehr gut, wenn dieses Projekt weiterlaufen könnte, damit so vielen Patienten wie möglich Hilfe zur Selbsthilfe angeboten werden könnte.

Es wäre sehr sinnvoll, auch bei den positiven Antworten abgestuft wählen zu können. Es gibt oft nur eine positive und diverse negative Aussagen zur Auswahl

Es wäre zur Besserung (Stabilisierung) des Patienten gut, in der Folgezeit hin und wieder den Therapeuten aufsuchen zu können.

Es wurden nicht alle Teile meiner Probleme durch die Fragen erfasst. Die Fragen waren zu oberflächlich.

Frage 10 bräuchte auch die Möglichkeit: sehr viel besser als Antwort z.B. bei mir Frage B. Frage 11 G: Spontan können nicht alle Fragen beantwortet werden. Z.B. Frage nach Anweisung von Vorgesetzten, da ich Vorgesetzte habe, deren Anweisungen z.B. gegen die Arbeitszeitordnung verstoßen

Frage 10 dieses Bogens mehr aufteilen. Wenn man gerade an einer schweren Krankheit leidet, kann man trotzdem zufrieden sein.

Fragebogen sollte in jedem Fall während der Therapie ausgefüllt werden - nicht am Ende. Zielvereinbarungen mit dem Therapeuten wurden nicht angesprochen.

---

Fragebögen zu Gesundheitszustand / Befinden sollten Rubrik enthalten, in denen man besondere Lebensumstände angeben kann, und Frage, ob man an Tod / Sterben denkt, sollte explizit sagen, ob man allgemein daran denkt oder auf sich selbst bezogen

Fragen ausweiten und stärker differenzieren, da teilweise zu allgemein gehalten und zu hoher neg. interpretatorischer Gehalt

Fragen werden m. E. abhängig von täglichen Gefühlsschwankungen beantwortet. Hätte an einem anderen Tag evtl. anders angekreuzt.

Fragestellung auf dem BSQ könnte mit einem Skalenpunkt "Empfindung habe / kenne ich nicht" versehen sein

Gesundheitsstörungen, die keine organischen Ursachen haben, müssen nicht zwangsläufig psychische Ursachen haben

Gruppentherapie war mir eine große Hilfe

Habe ich den Test gemacht wenn ich eine schlechte Woche hatte, hat sich das Ergebnis nicht verbessert, sondern war schlecht. Hätte ich den Test an einem guten Tag gemacht, wäre er anders ausgefallen. Der Test hätte auch mehr auf die jeweilige Erkrankung eingehen müssen. Und nur, weil der Test schlecht ausfällt, heißt es nicht, dass mein Therapeut mir nicht hilft!!! Und mir dann keine weiteren Stunden zu bewilligen hat mir mehr geschadet als mir zu helfen.

Ich bedanke mich für Ihre Mühen

Ich bekam die Fragebögen ohne Kommentar und Information. Wusste nichts von einer Teilnahme. Welchen Nutzen habe ich dadurch?

Ich bin beruflich mit Kundenfeedback beschäftigt und halte auch deshalb Ihre Aktion für erforderlich.

Ich bin froh, dass ich daran teilnehmen konnte

Ich bin verärgert, dass für den Zeitaufwand die Bogen auszufüllen Therapiezeit ersatzlos verloren geht. Und zwar auf meine Kosten.

Ich fand es ärgerlich, zigmal nach Gewaltphantasien gefragt zu werden, aber kaum nach Zwangshandlungen aus meinem Problemkreis. Folglich wirkte ich gesünder als ich war.

Ich finde die Abstände zwischen den Fragen zu lang, da sie immer nach entsprechender Tagesform des Patienten ausgefüllt werden. Sie spiegeln nicht den Gesundheitszustand der kompletten Zeit wider. Die Befragung selber empfand ich als zu lang. Wenn man alle Fragen genau liest, nimmt sie 30-45 Min. in Anspruch.

Ich finde die Idee mit dem Projekt super!

Ich finde es schwierig, aussagen anzukreuzen, wenn das vorhandene gute oder schlechte Gefühl nicht unbedingt in das vorgegebene Zeitraster fallen, das Ankreuzverfahren oft starr ist und wenig flexibel.

Ich finde es wichtig für mich und gut, die Möglichkeit für Gespräche mit meiner Therapeutin in größeren Abständen nutzen zu können.

Ich hätte die Auswertung gerne für meine Akten nach Hause mitgenommen.

Ich hätte gern die Chance, die Auswertung der Bögen mit meinem Therapeuten zu besprechen.

Ich kann mir nicht vorstellen, dass ein solcher Fragebogen eine Auswirkung hat.

Ich konnte mich gegenüber meiner Therapeutin nicht öffnen und dadurch nicht mehr erreichen. Habe Therapeuten gewechselt

Ich leide an einer bipolaren Störung Typ I, die gerade durch ein emotionales Auf und Ab gekennzeichnet ist. Psychotherapie ist nur ein Ansatz neben medikamentöser Therapie und Psychoedukation oder Selbsthilfe. Außerdem leide ich an Asthma und Herzkrankheit. Ihre Fragen sind mir deswegen teilweise zu indifferent.

---

Ich war mit meiner Therapeutin sehr zufrieden und hatte alle Probleme ansprechen können. Einzig meine sexuellen Bedürfnisse, darüber konnte ich nicht ganz offen reden. Den gesundheitlichen Zustand von vor einer Woche in den Fragebögen zu beschreiben finde ich zu wenig. Kurze Tiefs können so das Gesamtergebnis verfälschen.

Ich ziehe voraussichtlich nach Hamburg, würde gerne mit der Therapie weitermachen. Geht das in HH - Bergedorf?

Immer mehr Menschen benötigen eine Psychotherapie. Ich hoffe, dass die TK auch weiterhin die Kosten trägt, das beziehe ich auf die Allgemeinheit.

In den Tests geht es um die letzten 7 Tage, ich hatte immer das Gefühl, es ist eine zu kurze Zeitspanne. Es war schwierig für mich, nur auf die letzten 7 Tage zu achten.

In manchen Bereichen "Fragestellung" war es mir zu intim, auch wenn diese Fragen anonym behandelt werden.

In meinem Fall wäre es eine große Hilfe gewesen, wenn es bei der TK einen Topf gäbe, aus dem vielleicht noch 4 Therapien hätten bezahlt werden können, ohne ein Gutachten zu erstellen. Weitere 20 Therapiestunden wären zu viel gewesen, aber 4 hätten das Ergebnis vielleicht noch verbessert.

Insbesondere den BDI fand ich schwierig zu beantworten: 1. Es fehlt "Manchmal" zwischen 1 u. 2

Bei D fehlt: "Ich kann die Dinge besser genießen" Vielleicht konnte ich sie ja noch nie genießen.

Insgesamt hat man das Gefühl, einige Kreuzchen erklären zu müssen.

Können Sie mir Zugang zu den Ergebnissen der Befragung verschaffen ?

Leider beschränkt sich die Behandlung auf 50 Sitzungen. Ich persönlich hätte gerne noch weiter mit meinem Therapeuten gearbeitet, da ich schon besser zurechtkam, aber trotzdem noch Hilfe benötigt hätte.

Man müsste eine Auswertung dieses Modells als Teilnehmer erhalten. Der Weg zu einer ambulanten Psychotherapie ist aufgrund der langen Wartezeiten (halbes bis ein Jahr) viel zu lang.

Man sollte verstärkt darauf achten, dass sich der bürokratische Aufwand sowohl für den Patienten als auch den Therapeuten in vertretbarem Rahmen hält.

Man weiß nicht, in welcher Richtung das Modellvorhaben interpretiert wird.

Manche Fragen - zu allgemein

Manche Menschen sind depressiv und nehmen zu und nicht nur ab.

Mehr auf gesundheitliche Probleme eingehen.

Mehr zugelassene Therapeuten = schnelle Behandlung = viel früher wieder gesund. Bedeutet insgesamt weniger Kosten.

Mein Problem war die Akzeptanz einer plötzlich diagnostizierten chronischen Krankheit. Die Fragen trafen auf mich nicht immer zu.

Meiner Meinung nach lässt sich der Bereich Therapie/Psychotherapie nur sehr unbefriedigend anhand quantitativer Forschungsmethoden einschätzen und / oder beurteilen

Momentan keine, aber eine Information über die Ergebnisse d. Qualitätsmonitoring (bspw. Wie viel Prozent d. Therapien verliefen zufriedenstellend, welche Therapieeinrichtungen erscheinen nach diesen Ergebnissen besonders wirksam) würde mich nach Ende des Modellvorhabens interessieren. Daher hoffe ich, dass entsprechende Informationen im Internet zu finden sein werden.

Neben dem subjektiven Empfinden konnte ich mit den Auswertungen sehen, wie sich auch nach objektiven Kriterien mein persönliches Befinden zum Besseren wendet. => Sollte im Rahmen einer jeden Behandlung durchgeführt werden.

Online-Assessment entwickeln

---

Patientenbefragung auch nach Ende der Therapie mehrmals durchführen (z.B. nach 6 Monaten und 1 Jahr)

Qualitätsmonitoring ist wichtig, aber stört den Prozess der Therapie.

Qualitätssicherung in der ambulanten Psychotherapie sollte über den Modellstatus hinaus institutionalisiert werden.

Sie sollten auch die Therapieformen untersuchen, die Sie als Kasse nicht bezuschussen (z.B. bioenergetische Therapie...)

Sind die Ergebnisse öffentlich?

Sollten in anderen Bereichen auch durchgeführt werden (behandelnder Arzt, Medikamente, Therapien)

Um einen Termin zu bekommen habe ich zu lange warten müssen. Es sollte mehr gute Therapeuten mit Terminen geben.

Verbesserung der Fragebögen. Eindeutigere Fragen (in Bezug auf Zeitraum & Symptome)

Verfolgen Sie weiter die Wichtigkeit einer umfangreichen Depressionsbehandlung. Viele Menschen brauchen das. Therapeuten und Psychiater sollten schneller Termine zur Verfügung stellen können.

Viel zu lange Wartezeit, wenn man Hilfe braucht in der Not kann man unmöglich 1 Jahr oder länger warten, die Probleme sind jetzt da .... (für andere)...

Viele Antwortmöglichkeiten würde ich nicht so formulieren, z.B. BDI, wie vorgegeben, sie sind nur negativ formuliert

Viele Fragen sind eigentlich doppelt.

Vielen Dank !

Warum bin ich wieder am Anfang meiner Probleme ?

Weiter so !

Weiter so ! Ich finde es wichtig, dass sich die Krankenkasse für..... ?

Welche Möglichkeiten bestehen noch, die ich zur Gesundheitserhaltung unternehmen kann, damit ich meine Arbeit weiterhin mit Motivation und Engagement leisten kann ?

Werden die Ergebnisse veröffentlicht und wenn ja, wo?

z. Zt. Nicht

zu 12: Auf jeden Fall begrüße ich jedes Qualitätsmonitoring. Zu einer Beurteilung fehlt mir leider die Möglichkeit

Zu allgemeine Fragen, die z. T. auch widersprüchlich beantwortet werden. Trotzdem stimmen die Antworten. Die Fragebögen sollten evtl. öfter ausgefüllt werden oder möglicherweise nach jeder Sitzung / via Internet z.B. Ich glaube nicht, dass dadurch die Qualität eines Therapeuten eingeschätzt werden kann; das Verhältnis zwischen Patient und Therapeut ist zu speziell und intim; entweder stimmt die Richtung gleich am Anfang, die eingeschlagen wird, oder es bleibt erfolglos.

Zu kurze Zeit, zu wenig Stunden

Zu Punkt 2: Die Kurärztin empfahl mir die EMDR-Methode zur Behandlung. Sämtliche Therapeuten der Umgebung hatten keine freien Termine, so wandte ich mich an die TK die mit Therapeuten für Traumaverarbeitung und freiem Terminkalender vermittelte.

---

### **Evaluationsfragen aus Patientensicht:**

#### **Haben Sie zu den (grafischen) Rückmeldungen weitere Anregungen, Kritik oder Fragen?**

Da ich keine erhalten oder gesehen habe, kann ich das nicht beurteilen.

Da ich zu den Fragebögen keine Rückmeldung während meiner Therapie bekommen habe, kann ich zu diesen Punkten nichts sagen. Eventuell werde ich mich nochmal an ihn wenden.

Da meine Behandlung nicht von langer Dauer war, gab es während der Therapie wahrscheinlich keine Rückmeldung von Ihrer Seite.

Da meine Gesprächszeit an dem Tag dann verkürzt war und auch nicht meine aktuelle Situation spiegeln konnte, waren die Ergebnisse für mich uninteressant, von mir nicht besonders geschätzt

Darstellung und Bezeichnung schien mir eher für Therapeuten verständlich.

Das Wichtigste zu den Rückmeldungen: die Infos lassen einen zum Mitwisser über einen selbst werden und das Selbstbild mit der Realität vergleichen. Als aktiver Mitwisser kann man sich daher auch selbst angehen und bewegen, auch hinsichtlich der grafischen Ausschläge über die Limitgrenze.

Dazu kann ich keine Stellung beziehen, weil diese Rückmeldungen nicht mit mir persönlich besprochen wurden.

Der grafische Verlauf der Genesung sollte den Patienten als Kopie zur Verfügung gestellt werden. Ich habe nach wie vor großes Interesse daran. Schließlich geht es ja um mich selbst und nicht um eine dritte Person. Danke, dass ich an diesem Modellprojekt teilnehmen konnte.

Der Test müsste in kürzeren Abständen durchgeführt werden, um eine wirkliche Entwicklung oder einen Krankheitsverlauf zu beschreiben.

Die Besprechung des Ergebnisses hat mir rein gar nichts gebracht, weil ich ja weiß, wie es mir geht. Aber die Tests könnten zum Beispiel als Vorabinformationen für einen Vertrauensarzt dienen, aber nur in Verbindung mit einem persönlichen Gespräch. Und man müsste die Therapiedauer, bzw. den Zeitraum jedem Patienten anpassen. Meine Stunden sind vorbei und eine wichtige Stütze fällt weg, nämlich die Gespräche mit meinem Therapeuten.

Der Zeitpunkt der Beantwortung ist teils in schlechte Befindlichkeitsphasen gefallen, von daher nicht 100%ig aussagefähig.

Die Grafiken könnten noch übersichtlicher ausfallen, mit laienverständlichen Skalierungen, dickeren Linien oder überhaupt eher als Balkendiagramme gestaltet. Die Legenden könnten etwas ausführlicher sein.

Die Grafiken waren teilweise recht unübersichtlich und ich hätte sie oft ohne die Erläuterungen des Therapeuten nicht verstanden.

Die Grafiken zeigten die bereits mit dem Therapeuten zuvor besprochenen Veränderungen und bestätigten dessen außerordentlich gute Arbeit. Ich persönlich brauchte die Rückmeldungen nicht.

Die grafischen Rückmeldungen hatten mit der wirklichen Situation überhaupt keine Ähnlichkeit und sind meiner Meinung nach blödsinnig !

Die in der Rückmeldung formulierten negativen Entwicklungen würde ich nicht als negativ bezeichnen. Es sind Ängste und Probleme in der Therapie aufgetaucht, intensive Gefühle, die ich als Prozess in der Therapie begriffen habe, die mich auch belastet haben - die aber auch zu mir gehören, genauso wie positive Gefühle - durch die Fragebögen kommen solche Entwicklungen nicht zum Ausdruck.

Die Rückmeldungen sind nur in Zusammenarbeit mit dem Therapeuten verständlich.

Die verschiedenen Linien waren sehr schwer auseinanderzuhalten.

eigentlich nicht

---

Eine Konfliktsituation zwischen Weiterbewilligung einer Therapie und Beantwortung des Fragenkatalogs wird es immer geben.

Ergebnisse können verfälscht werden durch andere Erkrankungen, die sich auf den Gesundheitszustand auswirken (z.B. Infekt). Da man die letzten Wochen, bzw. die vergangene Woche bei manchen Fragen beurteilen sollte, gab es Verzerrungen im Ergebnis. Die Abstände zwischen den Messungen waren m. E. zu groß. Ansonsten fand ich es gut, noch eine Kontrolle der Therapie bzw. des Verlaufs in Form der Fragebögen und der Auswertung zu haben.

Es fehlen Musterbewertungen und Erklärungen. Es fehlen Verhaltensempfehlungen aus den grafischen Ausschlägen.

Es kamen wenig Rückmeldungen seitens des Therapeuten. Daher war ich teilweise verunsichert, was er wirklich denkt oder wie er die Lage beurteilt.

Für mich waren die wichtigsten Fragen enthalten.

Grafische Auswertung aller Einzelfrage. Nicht nur kumuliert

Grafische Rückmeldungen habe ich keine, mein Allgemeinzustand ist schlechter denn am Anfang. Habe versucht, mit dem was ich gelernt habe zu leben. Aber es hat nicht gereicht.

Hatte den Eindruck, dass die Auswertung negativer war als das Ausfüllen.

Ich bin mir nicht mehr sicher, ob ich die Ergebnisse gesehen habe. Wenn ja, kann ich mich nicht an die jeweilige Aussagekraft erinnern und somit auch nicht die Fragen 15 und 16 beantworten.

Ich bin mir nicht sicher, ob ich die graphischen Rückmeldungen und Reports je gesehen habe, daher habe ich überall in der Mitte das Kreuz gesetzt.

Ich finde diese statistischen Auswertungen psychischer Belange wenig sinnvoll,; wichtiger wäre, dass einzelne Aspekte, die auf Probleme hinweisen gezielt angesprochen werden.

Ich habe diese nicht gesehen.

Ich möchte mir die Option offen halten, mein Kontingent an Sitzungen innerhalb der Frist wenn es notwendig sein sollte, zu nutzen.

Ich würde mir wünschen, nach Abschluss der Auswertungen die Unterlagen zugeschickt zu bekommen, wenn das möglich wäre...?

Kann mich an Gespräche über grafische Rückmeldung nicht erinnern und daher keine Aussage darüber machen.

Keine grafische Rückmeldung erfolgt.

Kurze schriftliche Gesamtentwicklung

Leider sind die Fragen nicht für jeden anwendbar. Sie gehen von einem sehr passiven Patienten aus und sehr partielle Probleme werden nicht abgefragt - also überhaupt nicht bewertet.

Leider wurden mir diese Rückmeldungen nicht vom Therapeuten vorgelegt.

Min. & Max nicht erreichbar

Nach meiner Meinung sind sie nicht für jeden Patienten verständlich. Mein Therapeut hat mir die Grafiken immer erklärt, da ich sie sonst nicht immer verstanden hätte.

Nein. Zur Zeit bin ich voll zufrieden und hoffe, falls nötig meinen Therapeuten aufsuchen zu können (klingt wie Rückversicherung, hilft aber).

Nein, war für mich nicht so wichtig.

Sie sollten dem Patienten zur Verfügung gestellt werden.

Sie spiegeln den Erfolg der Therapie nicht wider.

Sie waren für die eigentliche Therapie völlig irrelevant.

---

Sollte Teil einer jeden Therapie werden.

Die Arbeitsbeziehung war sehr gut und wurde durch die graphischen Auswertungen unterstützt. Die Frage G sollte umbenannt werden. Z. Bsp. Ersetze "... Zum Positiven verändert" durch "... unterstützt."

Telefonisch wurde mir mitgeteilt, dass ich eine Schwäche im sozialen Bereich hätte und hier ein Therapiebedarf bestehen kann. Diesen kann ich jederzeit in Anspruch nehmen bzw. im sozialen Bereich könnten Probleme auftreten (Auch in der Zukunft) => Ist dies die grafische Rückmeldung ?

Toll, weiter so !!

Von meinem Therapeuten bekam ich diese Ergebnisse nicht. Meine überweisende Psychiaterin hat sie mit mir besprochen. Leider stimmte die Einstellung und Zusammenarbeit beider nicht überein.

Was hat das jetzt außer Zeit und Arbeit gebracht ?

zu 15 G: Die Atmosphäre war schon vorher gut

---

## **Evaluationsfragen aus Gutachtersicht:**

### **Sonstige Anmerkungen und Hinweise**

Am hilfreichsten waren die Selbstaussagen des Pat.

Anderer Vorgutachter? Unvollständige Unterlagen. Bericht bietet keine Entscheidungsgrundlage.

Angaben der Th. Handschriftlich und kaum lesbar.

Gleiches gilt für den Bericht des Gutachters (Anm. AZ)

Anonymität der Pat. Aufgehoben !

Antrag wurde falsch an Prof. Jacobi geschickt. Dieser hat ihn dann mir zugesandt. Ich habe den Antrag vorerst abgelehnt, da aus den abstrakten Informationen heraus keine Beurteilung der Zweckmäßigkeit möglich war.

Antragsinformationen knapp ausreichend.

Auf Grundlage der vorliegenden Informationen ist keine methodenspezifische Behandlungsplanung erkennbar.

Ausführlicher Befund musste nachgefordert werden.

Bagatellisierung

Behandlungsbedürftige Symptomatik fraglich; Psychodiagnostik nur angedeutet; Therapeuten nicht zu beurteilen

Bei allen bisherigen Fällen waren Nachforderungen an den Therapeuten nötig (Verhaltensanalyse, Diagnostik, Basisdokumentation) Dies erhöht den Aufwand. Eine klare Anweisung an die Therapeuten bezüglich der Unterlagen für den Gutachter wäre hilfreich.

Bei allen bisherigen Fällen waren Nachforderungen an den Therapeuten nötig (Verhaltensanalyse, Diagnostik, Basisdokumentation). Dies erhöht den Aufwand. Eine klare Anweisung an die Therapeuten bezüglich der Unterlagen für den Gutachter wäre hilfreich.

Bei so starken Diskrepanzen wie hier, vertraue ich mehr auf den Kurzbericht.

Bipolare Störung unter Medikation? Manische Phasen.

Das gesamte Projekt erscheint mir fragwürdig, da dem Gutachter keine hinreichenden (unlesbares Wort) für eine verantwortliche Beurteilung zur Verfügung stehen.

Das Modellvorhaben liefert keine hinreichende Grundlage für verantwortungsvolle Gutachterentscheidungen

Daten zu medizinischer Anamnese fehlten völlig, Medikation wurde nicht benannt. Offensichtlich ist eine GAD eine (nächstes Wort völlig unleserlich) die sich hier schlecht abbilden lässt.

Der handschriftliche Bericht des Therapeuten an den Gutachter war schlichtweg nicht leserlich.

Der Kurzbericht der Th. War sehr gut. Allerdings auch auf 5 Seiten ! Der Bericht war maschinenschriftlich erstellt und z. T. umfangreich, so dass (Rest unleserlich ...)

Der stationäre Entlassungsbericht wäre noch hilfreicher gewesen

Diagnosewechsel! Wird nicht in den Tests abgebildet.

Die Anmerkungen des Gutachters waren in großen Teilen unleserlich

Die Fragebögen spiegeln den Verlauf nicht präzise wider. Der Bericht ist viel individueller und damit besser. Mir fehlt hier der ärztliche Bericht! Insbesondere bei einer solchen Störung ist eine enge Kooperation zwischen PP und Arzt sehr bedeutsam. Das wurde hier nicht abgebildet.

Die graphischen Rückmeldungen alleine würden mir keine Entscheidung ermöglichen.

---

Die handschriftlichen Ergänzungen sind sehr wichtig, aber kaum lesbar und zu kurz

Die im TK-Modell gelieferten Informationen sind keine hinreichende Datengrundlage für eine verantwortungsvolle Gutachterentscheidung.

Die psychometrischen Skalen sind nicht änderungssensibel genug oder die Therapeuteneinschätzung ist zu optimistisch

Die schriftlich mitgeteilten Informationen und Ausführungen der Therapeutin ermöglichen keine qualifizierte Beurteilung von Notwendigkeit und Zweckmäßigkeit der durchgeführten Therapie.

Die Strukturierung des Fortsetzungsantrags müsste mehr auf die Diskrepanz Test / Bericht eingehen

Die textlichen Ausführungen sind sehr knapp und grenzwertig. Die undokumentierte Testdiagnostik und der Wust von Anamnesefragebögen ist nicht zielführend.

Die Therapeutenberichte sind insgesamt doch zu "dünn"

Die Unterlagen waren nicht vollständig und auch nach Anforderung ergänzungsbedürftig (es fehlten die Basisdokumentation und Planung der weiteren Therapie)

Die Verhaltensanalyse sollte bei dem Kurzbericht obligatorisch sein

Dies war der 1. Antrag, bei dem ich aufgrund der mitgelieferten Unterlagen die Indikation für die Fortführung einer ambulanten VT über 45 Stunden hinaus relativ problemlos beurteilen konnte.

Dieser Fall ist mit diesem Verfahren überhaupt nicht beurteilbar. Eine bipolare Störung ist so nicht abbildbar und damit nicht beurteilbar!

Ein Teil der Unterlagen fehlte, mehrfache Rückfragen erforderlich. Der Kurzbericht war mehr als dürftig und lässt eine gutachterliche Stellungnahme eigentlich gar nicht zu.

Ein Wust von z. T nicht oder falsch ausgefüllten Unterlagen. Die Testergebnisse waren NICHT grafisch ausgewertet.

Erstbericht liegt nicht vor. Fortsetzungsbericht auf dieser Basis kaum zu beurteilen. Einblick in den Therapieprozess ergibt sich nicht.

Erste Betrachtung in diesem Rahmen

Es handelt sich um einen Fortführungsantrag 2 für insgesamt 80 Std. Die eingereichten Unterlagen waren völlig unzureichend, so dass eine Beurteilung ob die Indikation für eine LZT über 80 Std. gegeben war, nicht möglich war. Der Antrag wurde deshalb nicht befürwortet.

Es handelt sich um eine schwere, chronifizierte Zwangsstörung bei einer Pat. Therapieziele und Behandlungskonzept wurden völlig unzureichend dargestellt.

Es kostet viel Zeit, die unterschiedlichen Informationen zu strukturieren. Hilfreich wäre, wenn die einzelnen Dokumente (Psych. Messungen, Basisdaten usw. zusammengeklammert vorlägen, sodass nicht geordnet werden müsste.

Es wird vom Therapeuten ein kompletter und guter Antrag gemäß PTR über 3 Seiten vorgelegt. Zusätzlich noch der Wust an TK-Modell-Unterlagen. Die Befürwortung erfolgte aufgrund des schriftlichen Berichtes des Therapeuten.

Falsches Formular! Starke Diskrepanzen Selbst-Fremd. Fehlende Informationen. Diagnostik bei Störungsbild ungeeignet.

Fragebögen lagen nicht vor.

Für mich war ausschlaggebend der beigefügte Bericht der Pat. Über ihre bisherige Therapie.

Habe sehr wenig mit dem Modellprojekt zu tun! Sehr selten Beurteilungen. Vergesse dann, wie es geht. Bitte um Entschuldigung für die Verspätung, hatte es als Sonderfall aussortiert.

Handschriftlich schlicht unlesbar.

---

handschriftliche Darstellung nicht gut

Hier Behandlung als TP begonnen, ohne Begutachtung seitens KK Umwandlung in PA fehlt Fortführungsantrag. Die wesentlichen Informationen zur Beurteilung der Psychopathologie und Behandlungsplanung fehlen. Auf dieser Basis ist eine Begutachtung verantwortungsvoll kaum durchzuführen. (Anm. des Trustcenters: Die Bemerkungen des Gutachters waren im Großen und Ganzen kaum zu entziffern.)

Hier ist die Beurteilung sehr schwer, ein ärztlicher Bericht fehlt, Angaben zu (indizierten) Medikation auch störungsspezifischer Test geht hier an der Diagnose vorbei.

Hier liegt wohl eher eine organische Störung vor (Verdacht), die im alten Verfahren durch einen ärztlichen Gutachter früher entdeckt worden wäre.

Hilfreich war eine vom Th. und von der Pat. Angefertigte Problembeschreibung inkl. Einer Krankheitsverlaufsdarstellung. Zur befriedigenden Beurteilung ist mehr Klartext erforderlich. Stichworte und Fragebogenergebnisse ersetzen nicht den Nachweis einer stimmigen Analyse der Störungsbedingungen.

Ich habe bisher kaum einen korrekt ausgefüllten Antrag erhalten. Gut beurteilbare Anträge entsprechen in Art und Umfang dem bisher im Gutachterverfahren üblichen. Das TK-Modell scheint für alle Beteiligten, insbesondere für die Therapeuten zu kompliziert zu sein. Immer wieder gibt es viele formale Fehler.

In diesem Fall lagen dem Antrag keine Unterlagen für den Gutachter bei, so dass er auf Kopien zurückgreifen musste.

Informationen aus Bericht (TK) unzureichend. Nachforderung von Unterlagen notwendig, hoher Zeitaufwand.

Informationen mussten nachgefordert werden. Basisdokumentation in Originalversion schwer lesbar.

Keine Übereinstimmung zwischen Psychometrik und Klinik wegen Abwehrkonfiguration des Patienten - beantragte Verfahren unzureichend

Kurzbericht vergleichsweise gut. Lässt Indikation erkennen. Psycho???? Angedeutet. Behandlungsplan kann gerechtfertigt sein, Differentialindikation nicht bewertbar. Prognostische Aussichten aufgrund fehlender Info nicht bewertbar.

Leider waren die Tests... Rest unleserlich

Nachforderung von Unterlagen

Nachforderung von Unterlagen.

Nachforderungen nötig.

Name war angegeben: Datenschutzverletzung. Psychometr. Daten BSI, IIP-D, SF12 fehlten

Nicht alle Infos verfügbar. Falsches Formular verschickt.

Normale Berichtsform trotz TK-Modell

Papiere, die im chaotischen Zustand ankommen und geordnet werden müssen

Psychometrie stimmte nicht mit klein. Bericht überein.

Psychometrische Erfassungen und persönliche Berichte scheinen doch eher aufeinander zu???

Sehr guter und ausführlicher Kurzbericht, der den Prozess nachvollziehbar macht

Selbst- und Fremdbeurteilungen sind aufgrund der Abwehr d. Pat. sehr different, der Kurzbericht ist hinreichend hinsichtlich Psychodynamik und Therapieplanung.

Sonderfall, schwere Tumorerkrankung

Starke Diskrepanzen Selbst-Fremd, fehlende Informationen durch TK

Starke Diskrepanzen, selbst-fremd, falsches Berichtsformular, fehlende Informationen, aufwändige Nachforderung nötig

---

Starke Verleugnungstendenz. Abhängigkeit nicht erkannt.

Therapeutin reichte fehlende Unterlagen nicht nach.

Undurchsichtiges Formulargewirr, Kurzbericht des Therapeuten aus Platzgründen grenzwertig, z. T. auch Ausbesserungen. So bietet das Verfahren KEINE angemessene Bearbeitungsgrundlage.

Unterlagen / Bericht musste nachgefragt werden

Unterlagen offenbar unvollständig. Wäre besser elektronisch. Zu viel Papier. Störungsspezifische Testung fehlt.

Unterlagen wurden nachgereicht. Verlauf nicht adäquat abgebildet.

Veraltetes Formular, umständliche Nachforderung von Informationen, Handschrift schwer lesbar, Störung in Tests nicht abgebildet.

Verhaltensanalyse und Bedingungsanalyse sollten ausführlicher und konkreter dargestellt werden

Von Beginn an Zusatzinformationen des Therapeuten (Bericht) vorhanden. Daher ausnahmsweise kein Nachfragen.

Weigerung Therapeut auf nähere Angaben mit Verweis auf 30 Min. Maximum für Bericht

Widerspruch zwischen TK-Report und Gutachterangaben. Wird durch TK-Report nicht abgesichert

wir Gutachter verstehen hier nicht die Zuteilung zum Verfahren, die Testlage ist anders. Die ? der Behandlung ist zweifellos gegeben. Die Beschreibung der Ther. ist kompetent, die aktuelle Psychodynamik bleibt unklar.

zu 3 Tests: Überhaupt nicht hilfreich, da keine Werte mitgeliefert werden.

Zu den Punkten 3 F,G und I lagen mir keine Informationen vor.

Zu wenig beansprucht. Diskrepanzen Therapeuten vs. TK-Dokumentation

Zum Antrag wurden keinerlei Unterlagen hinzugefügt, so dass ich einen Antrag auf LZT von 45 Stunden auf 10 Std. gekürzt habe.

---

## 13 Anhang: Optimierung der Rückmeldungen im Projektverlauf

Zum 01.04.2007 wurden die Texte der Rückmeldungen redaktionell geändert. Die bis dahin verwendete Fassung sowie Beispiel-Auswertungen vorher-nachher sind hier abgebildet (zu der vom 01.04.2007 bis zum Ende der Feldphase verwendeten Fassung vgl. Abschnitt 6.1):

Keine Veränderung	
<b>„Bisher keine klinisch bedeutsamen Veränderungen“</b>	Die Erhebung in den verschiedenen Messinstrumenten zeigt keine eindeutige Entwicklung dieser Patientin / dieses Patienten im Sinne einer klinisch bedeutsamen Abnahme oder Verstärkung der Problematik in der subjektiven Selbsteinschätzung <i>im Vergleich zu den Werten bei Therapiebeginn</i> . Dies kann unterschiedliche Ursachen haben. Bitte überdenken Sie den bisherigen Therapieverlauf auch mit Blick auf eine eventuelle Vermeidung problematischer Verhaltensweisen, auf Verdrängung der Hintergründe der psychischen Störung oder auf Probleme in der therapeutischen Beziehung.
<b>Vormals: „Keine eindeutige Entwicklung“</b>	Bisher keine klinisch bedeutsamen Veränderungen! Die Erhebung in den verschiedenen Messinstrumenten zeigt keine eindeutige Entwicklung des Patienten gemäß einer klinisch bedeutsamen Veränderung. Die Auswirkungen Ihrer therapeutischen Strategien können daher zum jetzigen Zeitpunkt nicht genau beurteilt werden und die Weiterführung der Therapie erscheint angemessen.
Verschlechterung	
<b>„Verstärkung der Problematik“</b>	Die verschiedenen Messinstrumente zeigen <i>im Vergleich zu den Werten bei Therapiebeginn</i> bisher (noch) keine Abnahme der Problematik in der Selbsteinschätzung dieser Patientin / dieses Patienten, sondern deuten auf eine Verstärkung hin. Dies kann unterschiedliche Ursachen haben. Bitte überdenken Sie daher den bisherigen Therapieverlauf v.a. unter dem Aspekt der Angemessenheit oder eventuellen Anpassung ihrer bisherigen therapeutischen Strategien an die Belange dieser Patientin / dieses Patienten. Berücksichtigen Sie dabei vor allem mögliche Erklärungen für die im Moment vorliegende stärker ausgeprägte Problematik. Sie kann auch ein erwünschter Teil des therapeutischen Prozesses sein (z.B. im Sinne einer notwendigen Destabilisierung).
<b>Vormals: „Negative Entwicklung“</b>	Therapeutische Strategie (noch) nicht wirksam. Die Erhebung in den verschiedenen Messinstrumenten zeigt bisher keine Verbesserung des Patienten, sondern deuten eher auf eine Verstärkung der Problematik hin. Dies kann unterschiedliche Ursachen haben und mit dem therapeutischen Prozess in Verbindung stehen. Bitte überdenken Sie dennoch den bisherigen Therapieverlauf v.a. unter dem Aspekt der Angemessenheit oder eventuellen Anpassung ihrer bisherigen therapeutischen Strategien an die Belange dieser Patientin/ dieses Patienten. Berücksichtigen Sie dabei vor allem mögliche Erklärungen für die im Moment vorliegende stärker ausgeprägte Problematik bei dieser Patientin/ diesem Patienten.

Verbesserung	
<b>„Abnahme der Problematik“</b>	<p>Die Erhebung in den verschiedenen Messinstrumenten zeigt eine klinisch bedeutsame Abnahme der Problematik in der Selbsteinschätzung dieser Patientin / dieses Patienten <i>im Vergleich zu den Werten bei Therapiebeginn</i>. Dies kann unterschiedliche Ursachen haben.</p> <p>Bitte überdenken Sie daher den bisherigen Therapieverlauf v.a. unter dem Aspekt der Angemessenheit oder eventuellen Anpassung ihrer bisherigen therapeutischen Strategien an die Belange dieser Patientin / dieses Patienten. Berücksichtigen Sie dabei vor allem mögliche Erklärungen für die im Moment vorliegende positive Entwicklung. Es kann sich um eine produktive Entwicklung handeln, es kann aber auch sein, dass es sich um eine Vermeidung bzw. Verdrängung der Hintergründe der psychischen Störung handelt, durch die der verbesserte Zustand des Patienten nur vorübergehend / temporär wäre.</p>
<b>Vormals: „Gute Entwicklung“</b>	<p>Guter Verlauf! Bitte Therapieende in Betracht ziehen.</p> <p>Die Erhebung in den verschiedenen Messinstrumenten zeigt eine klinisch bedeutsame Veränderung der Patientin/ des Patienten. Das weist darauf hin, dass Ihr therapeutisches Vorgehen sehr gute Auswirkungen hat. Entsprechend sollten sie über das Therapieende nachdenken. Das bisherige Vorgehen sollten Sie dabei versuchen im Sinne Ihrer bisherigen therapeutischen Strategien fortzusetzen, um seine Wirkung bei der Patientin/ beim Patienten zu stabilisieren.</p>

Nachhaltige Verbesserung	
<b>„Nachhaltige Abnahme der Problematik“</b>	<p>Die Erhebungen in den verschiedenen Messinstrumenten zeigen zum wiederholten Male eine klinisch bedeutsame und eventuell nachhaltige Abnahme der Problematik in der Selbsteinschätzung dieser Patientin / dieses Patienten <i>im Vergleich zu den Werten bei Therapiebeginn</i>.</p> <p>Bitte überdenken Sie erneut den bisherigen Therapieverlauf v.a. unter dem Aspekt der Angemessenheit oder eventuellen Anpassung ihrer therapeutischen Strategien an die Belange dieser Patientin/ dieses Patienten. Berücksichtigen Sie dabei vor allem mögliche Erklärungen für die im Moment vorliegende nachhaltige Abnahme der Problematik. Auch diese könnte noch vorübergehend/temporär sein und würde dementsprechend eine weitere Auseinandersetzung mit der Problematik bzw. weitere therapeutische Interventionen erfordern. Reflektieren Sie nochmals eventuell weiter bestehende selbst- und fremdschädigende Anteile/Verhaltensweisen des Patienten sowie mögliche Abwehr- oder Vermeidungsstrategien, welche gegebenenfalls einer dauerhaften Besserung im Wege stehen.</p> <p>Prüfen Sie weitere therapeutische Schritte im Sinne einer Fortsetzung / Intensivierung der Therapie oder der Vorbereitung des Behandlungsabschlusses (Ablösungsprozess einleiten, Rückfallprophylaxe planen).</p>
<b>Vormals: „Stabil gute Entwicklung“</b>	<p>Erfolgreicher Verlauf! Therapieende erscheint möglich.</p> <p>Die Erhebungen in den verschiedenen Messinstrumenten zeigen zum wiederholten Male eine klinisch bedeutsame Verbesserung der Patientin/des Patienten. Ihr therapeutisches Vorgehen hat entsprechend sehr gute Auswirkungen und zeigt auch über den Verlauf der Therapie eine gute und ausreichende Stabilität. Nach den vorliegenden Erhebungen erscheint ein baldiges Therapieende möglich und Sie sollten überlegen eventuelle therapeutische Schritte, wie etwa die Planung einer Rückfallprophylaxe oder den beginnenden Ablösungsprozess zum Patienten, zu initiieren.</p>

## 13.1 Anonymisierte Beispiel-Auswertung (grafische Rückmeldung) vor Modifikation der Rückmeldungen

TKPROBE-1048 (M - 35;7 Jahre)  
Affektive Störung und Anpassungsstörung (Verlauf)

Seite 1 von 5  
Test vom 12.04.2005 11:45

### VERTRAULICHE UNTERLAGEN

### PSYCHODIAGNOSTIK - AUSWERTUNG

vom 12.04.2005, 11:45 Uhr

PCode	1E10001048
Name	Fallbeispiel 2+3
Vorname	
Alter	35;7 Jahre
Auftraggeber	
Fragestellung	

---

Serien-Nr: 8E728B8101E1  
Versions-Nr: 3.8.4  
Ausgewertet durch:

Die Interpretation von Testergebnissen muss durch eine psychologisch und psychodiagnostisch ausgebildete Person vorgenommen werden. Dabei gelten die Richtlinien des Testkuratoriums der

TKPROBE-1048 (M - 35;7 Jahre)  
Affektive Störung und Anpassungsstörung (Verlauf)

Seite 2 von 5  
Test vom 12.04.2005 11:45



**Techniker Krankenkasse**  
Gesund in die Zukunft.

## Qualitätsmonitoring in der ambulanten Psychotherapie Modellvorhaben der Techniker Krankenkasse

### Dokumentation: Verlauf

Patientenkennzeichen:	TKPROBE-1048
Geschlecht:	männlich
Alter:	35
Name Therapeutin/Therapeut:	
Untersuchung vom:	12.04.2005 (5. Therapiesitzung)

Im folgenden Bericht sind die Befunde zu Ihrer Patientin/Ihrem Patienten dargestellt, welche für die Gesamteinschätzung der Behandlung herangezogen wurden. Die Rückmeldungen sind dabei ausdrücklich als Ergänzung zu Ihrem klinischen Eindruck und Ihrer diagnostischen Abklärung zu verstehen.

Da es sich bei dieser Gesamteinschätzung um eine ergänzende Darstellung aus der Patientenperspektive handelt, bitten wir Sie, die Angaben und Handlungsempfehlungen für Ihr weiteres Vorgehen und Behandlungsangebot zu berücksichtigen.

Für die ermittelte Gesamteinschätzung wurden die allgemeine psychische Belastung, die interpersonalen Probleme und die störungsspezifischen Beschwerden herangezogen.

Als ergänzende Information finden Sie die durch die Patientin/den Patienten eingeschätzte therapeutische Arbeitsbeziehung und Angaben zur Lebensqualität.

Alle Normwerte sind in T-Werten für die Normalpopulation angegeben (Mittelwert 50, Standardabweichung 10). Weitere Informationen zur Erläuterung der einzelnen Teile der grafischen Rückmeldungen entnehmen Sie bitte der Zusatzinformation für Therapeutinnen und Therapeuten (HTS: TK-Hilfe als PDF-Dokument aufrufen).

#### Gesamteinschätzung:

1. Messung, Sitzung 2	
2. Messung, Sitzung 10	Negative Entwicklung
3. Messung, Sitzung 20	Keine eindeutige Entwicklung
4. Messung, Sitzung 45	Gute Entwicklung
5. Messung, Sitzung 55	<b>Stabil gute Entwicklung</b>

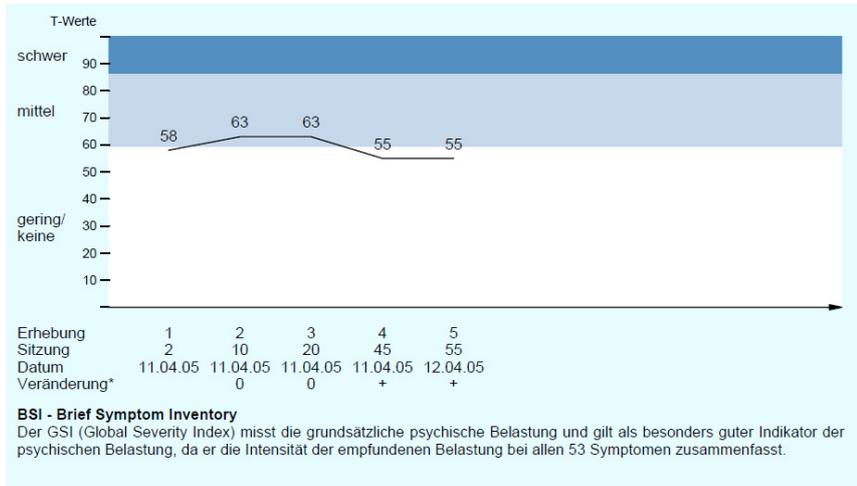
Erfolgreicher Verlauf! Therapieende erscheint möglich.

Die Erhebungen in den verschiedenen Messinstrumenten zeigen zum wiederholten Male eine klinisch bedeutsame Verbesserung der Patientin/des Patienten. Ihr therapeutisches Vorgehen hat entsprechend sehr gute Auswirkungen und zeigt auch über den Verlauf der Therapie eine gute und ausreichende Stabilität. Nach den vorliegenden Erhebungen erscheint ein baldiges Therapieende möglich und Sie sollten überlegen eventuelle therapeutische Schritte, wie etwa die Planung einer Rückfallprophylaxe oder den beginnenden Ablösungsprozess zum Patienten, zu initiieren.

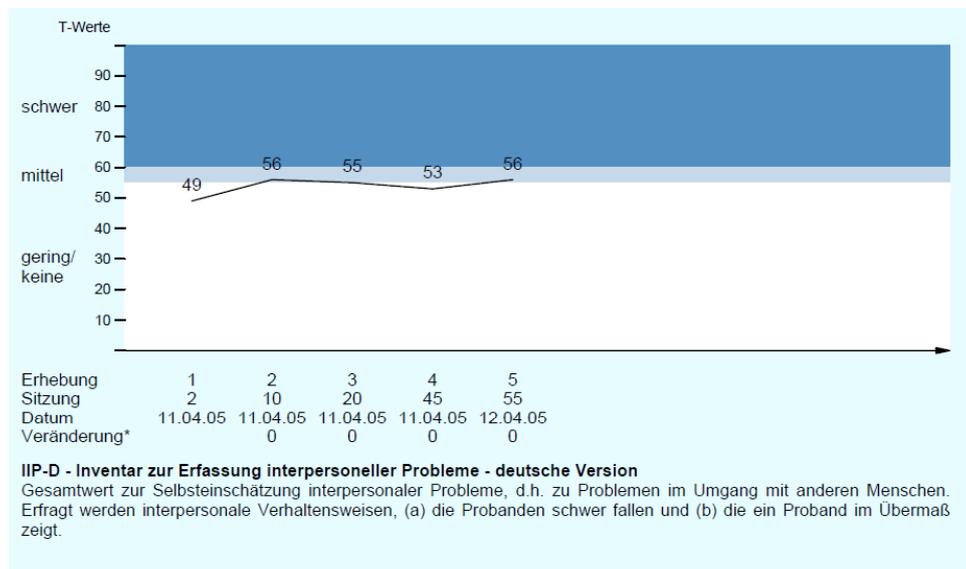
TKPROBE-1048 (M - 35;7 Jahre)  
Affektive Störung und Anpassungsstörung (Verlauf)

Seite 3 von 5  
Test vom 12.04.2005 11:45

**Wie schätzt Ihre Patientin/ Ihr Patient die allgemeine psychische Belastung ein?**



**Wie schätzt Ihre Patientin/ Ihr Patient die interpersonalen Probleme ein?**



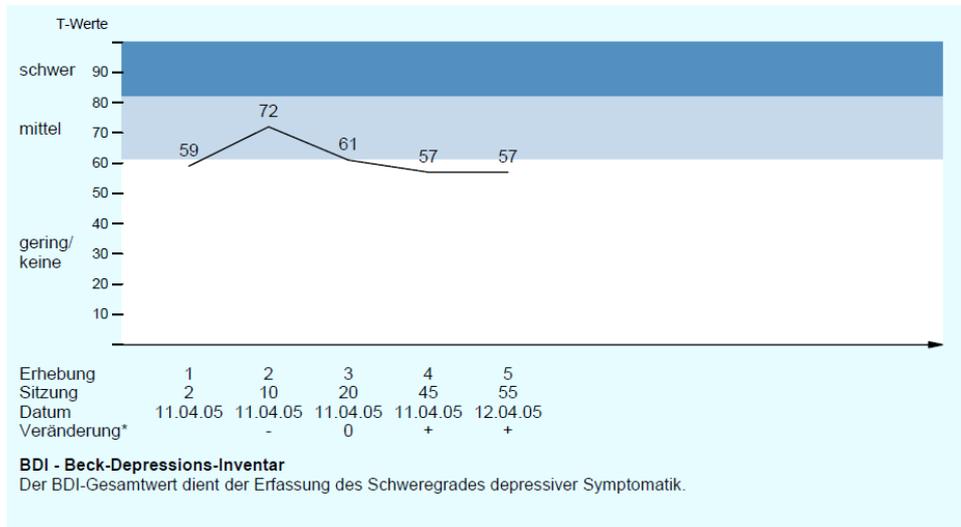
\* Veränderungsbewertung je Instrument bezogen auf die erste Messung:

- + „gute Entwicklung“
- 0 „keine eindeutige Entwicklung“
- „negative Entwicklung“

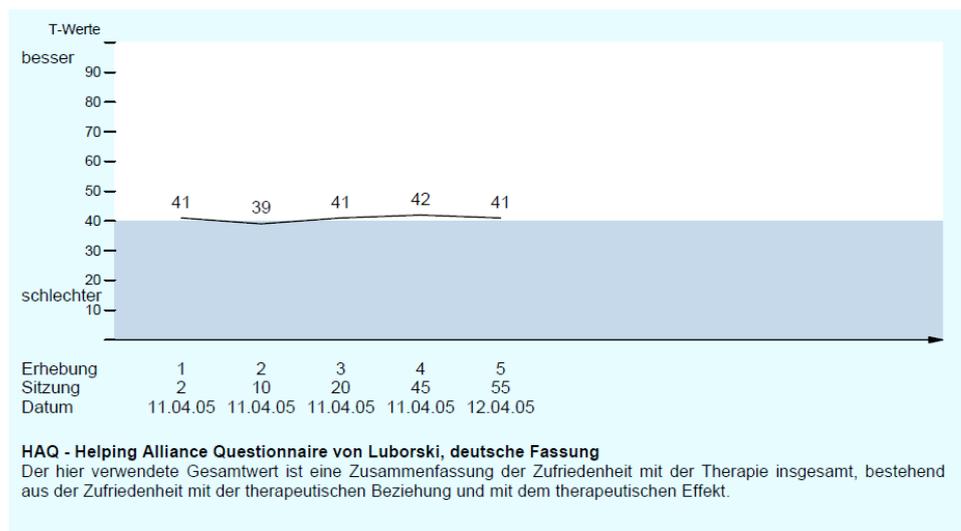
TKPROBE-1048 (M - 35;7 Jahre)  
 Affektive Störung und Anpassungsstörung (Verlauf)

Seite 4 von 5  
 Test vom 12.04.2005 11:45

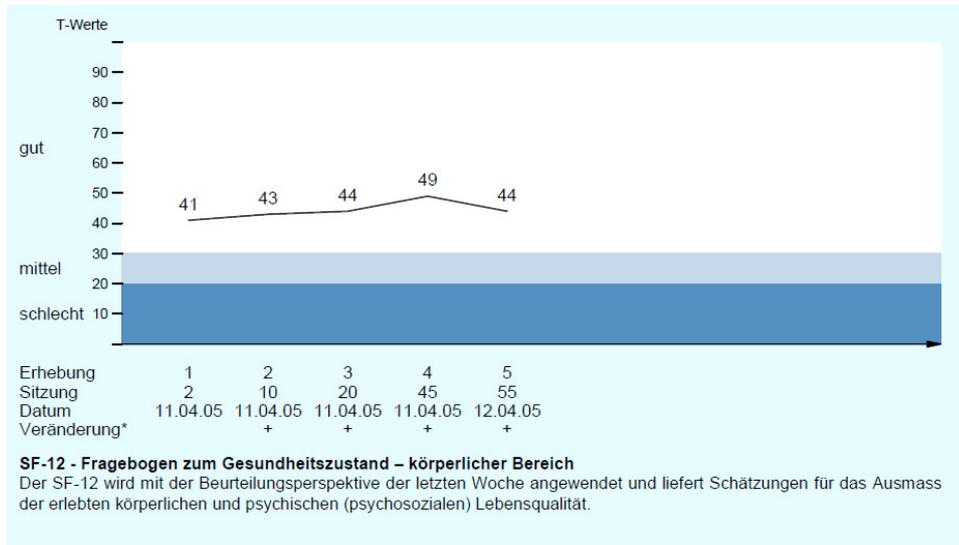
**Wie schätzt Ihre Patientin/ Ihr Patient die störungsspezifischen Beschwerden ein?**



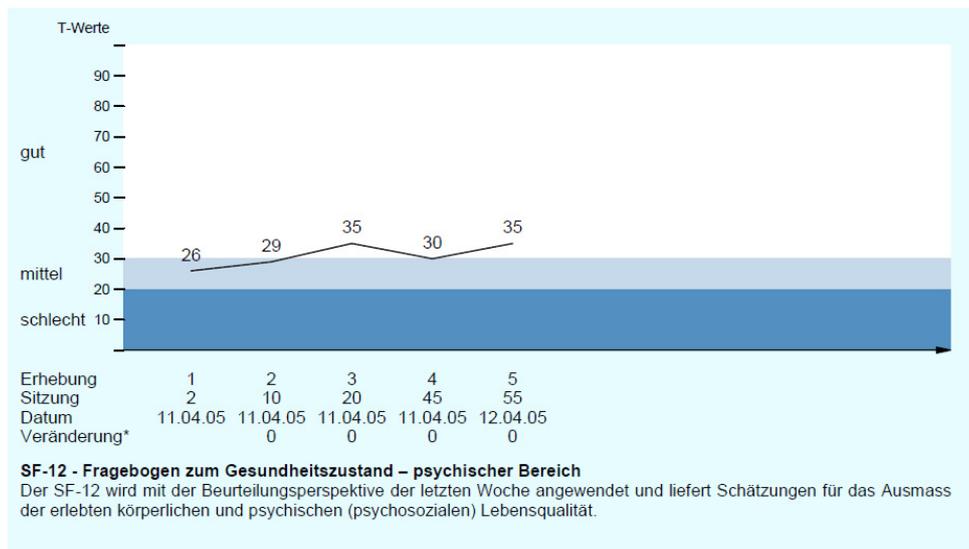
**Wie schätzt Ihre Patientin/Ihr Patient die therapeutische Arbeitsbeziehung ein?**



**Wie schätzt Ihre Patientin/ Ihr Patient die gesundheitsbezogene Lebensqualität ein?  
- Körperlicher Bereich**



**- Psychischer Bereich**



**Impressum**

Modellvorhaben der Techniker Krankenkasse nach §63 SGB V „Qualitätsmonitoring in der ambulanten Psychotherapie“  
Projektleitung: Dipl.-Soz. Andreas Nagel (Techniker Krankenkasse). Verantwortlicher Berater für Entwicklung und Umsetzung im Hogrefe TestSystem (HTS): Prof. Dr. Klaus-Dieter Hänsgen (Universität Fribourg). Entwicklung der Entscheidungsregeln und grafische Reports: Prof. Dr. Wolfgang Lutz und Dipl.-Psych. Sven Tholen (Universität Bern).

## 13.2 Anonymisierte Beispiel-Auswertung (grafische Rückmeldung) nach Modifikation der Rückmeldungen

██████████ (W - 39;6 Jahre)  
Angst und Phobie (Verlauf)

Seite 1 von 6  
Test vom 01.07.2008 14:59

### VERTRAULICHE UNTERLAGEN

#### PSYCHODIAGNOSTIK - AUSWERTUNG

vom 01.07.2008, 15:00 Uhr

PCode	██████████
Name	██████████
Vorname	
Alter	39;6 Jahre
Auftraggeber	
Fragestellung	Psychische Störung

Serien-Nr: 55FC5225C2D6  
Versions-Nr: 3.9  
Ausgewertet durch:

Die Interpretation von Testergebnissen muss durch eine psychologisch und psychodiagnostisch ausgebildete Person vorgenommen werden. Dabei gelten die Richtlinien des Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen (Psychologische Rundschau, 37, 1986, 162-165) bzw. für die Eignungsdiagnostik die DIN-Norm 33430.



W - 39;6 Jahre)  
Angst und Phobie (Verlauf)

Seite 2 von 6  
Test vom 01.07.2008 14:59



## Qualitätsmonitoring in der ambulanten Psychotherapie Modellvorhaben der Techniker Krankenkasse

### Dokumentation: Verlauf

Patientenkennzeichen: [REDACTED]  
Geschlecht: weiblich  
Alter: 39

Name Therapeutin/Therapeut: [REDACTED]  
Untersuchung vom: [REDACTED]

Im folgenden Bericht sind die Befunde zu Ihrer Patientin/Ihrem Patienten dargestellt, welche für die Gesamteinschätzung der Behandlung herangezogen wurden. Die Rückmeldungen sind dabei ausdrücklich als Ergänzung zu Ihrem klinischen Eindruck und Ihrer diagnostischen Abklärung zu verstehen. Da es sich bei dieser Gesamteinschätzung um eine ergänzende Darstellung der Patientenperspektive handelt, bitten wir Sie, die Angaben und Handlungsempfehlungen für Ihr weiteres Vorgehen und Behandlungsangebot zu berücksichtigen. Für die ermittelte Gesamteinschätzung wurden die allgemeine psychische Belastung, die interpersonalen Probleme und die störungsspezifischen Beschwerden herangezogen. Als ergänzende Information finden Sie die durch die Patientin/den Patienten eingeschätzte therapeutische Arbeitsbeziehung und Angaben zur Lebensqualität. Alle Normwerte sind in T-Werten für die Normalpopulation angegeben (Mittelwert 50, Standardabweichung 10). Weitere Informationen zur Erläuterung der einzelnen Teile der grafischen Rückmeldungen entnehmen Sie bitte der Zusatzinformation für Therapeutinnen und Therapeuten (HTS: TK-Hilfe als PDF-Dokument aufrufen).

### Gesamteinschätzung der Patientenangaben:

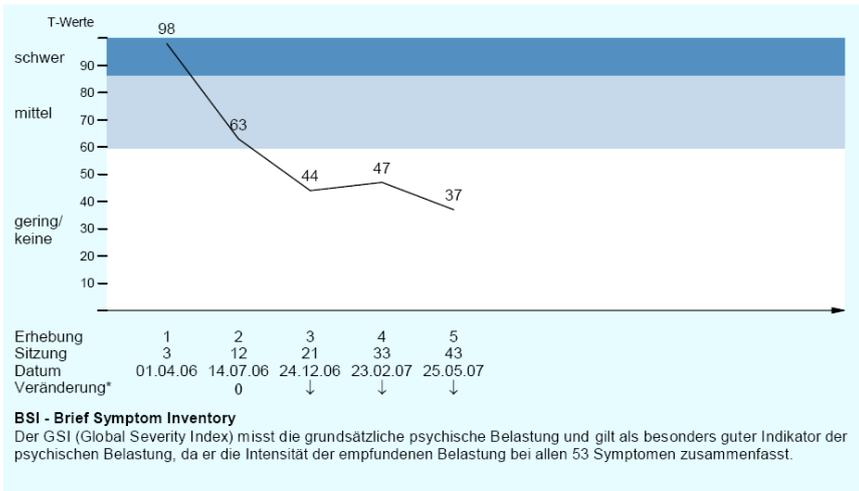
1. Messung, Sitzung 3	
2. Messung, Sitzung 12	Abnahme der Problematik
3. Messung, Sitzung 21	Nachhaltige Abnahme der Problematik
4. Messung, Sitzung 33	Nachhaltige Abnahme der Problematik
5. Messung, Sitzung 43	<b>Nachhaltige Abnahme der Problematik</b>

Die Erhebungen in den verschiedenen Messinstrumenten zeigen zum wiederholten Male eine klinisch bedeutsame und eventuell nachhaltige Abnahme der Problematik in der Selbsteinschätzung dieser Patientin / dieses Patienten im Vergleich zu den Werten bei Therapiebeginn. Bitte überdenken Sie erneut den bisherigen Therapieverlauf v.a. unter dem Aspekt der Angemessenheit oder eventuellen Anpassung ihrer therapeutischen Strategien an die Belange dieser Patientin / dieses Patienten. Berücksichtigen Sie dabei vor allem mögliche Erklärungen für die im Moment vorliegende nachhaltige Abnahme der Problematik. Auch diese könnte noch vorübergehend/temporär sein und würde dementsprechend eine weitere Auseinandersetzung mit der Problematik bzw. weitere therapeutische Interventionen erfordern. Reflektieren Sie nochmals eventuell weiter bestehende selbst- und fremdschädigende Anteile/Verhaltensweisen des Patienten sowie mögliche Abwehr- oder Vermeidungsstrategien, welche gegebenenfalls einer dauerhaften Besserung im Wege stehen. Prüfen Sie weitere therapeutische Schritte im Sinne einer Fortsetzung / Intensivierung der Therapie oder der Vorbereitung des Behandlungsabschlusses (Ablösungsprozess einleiten, Rückfallprophylaxe planen).

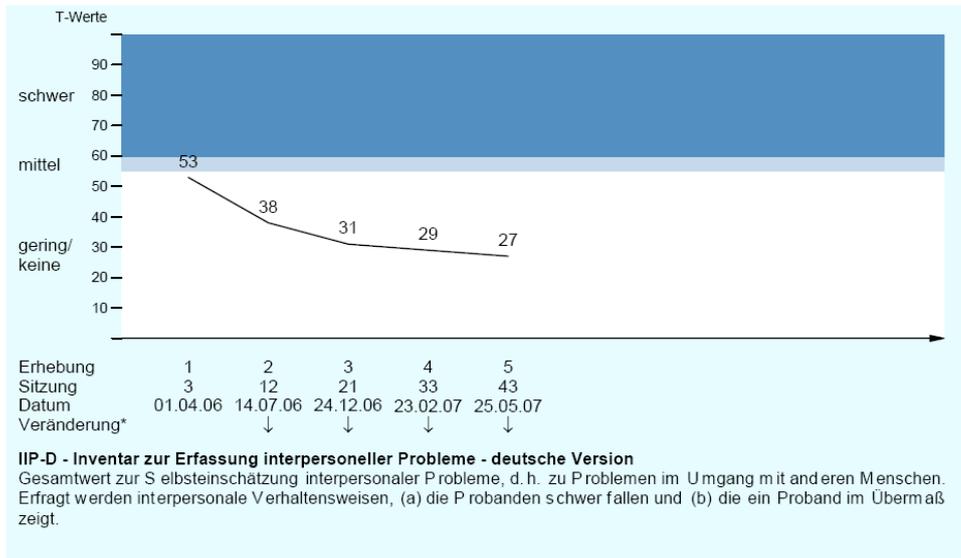
█ (W - 39,6 Jahre)  
Angst und Phobie (Verlauf)

Seite 4 von 6  
Test vom 01.07.2008 14:59

**Wie schätzt Ihre Patientin/ Ihr Patient die allgemeine psychische Belastung ein?**



**Wie schätzt Ihre Patientin/ Ihr Patient die interpersonalen Probleme ein?**



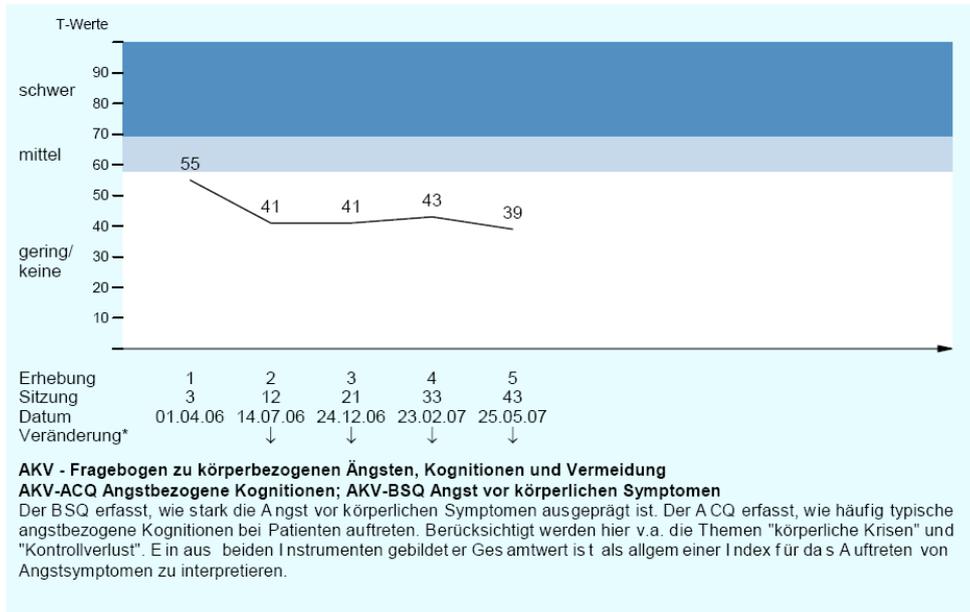
\* Veränderungsbewertung je Instrument bezogen auf die erste Messung:

- ↓ „Abnahme der Problematik“
- 0 „keine klinisch bedeutsamen Veränderungen“
- ↑ „V erstärkung der Problematik“

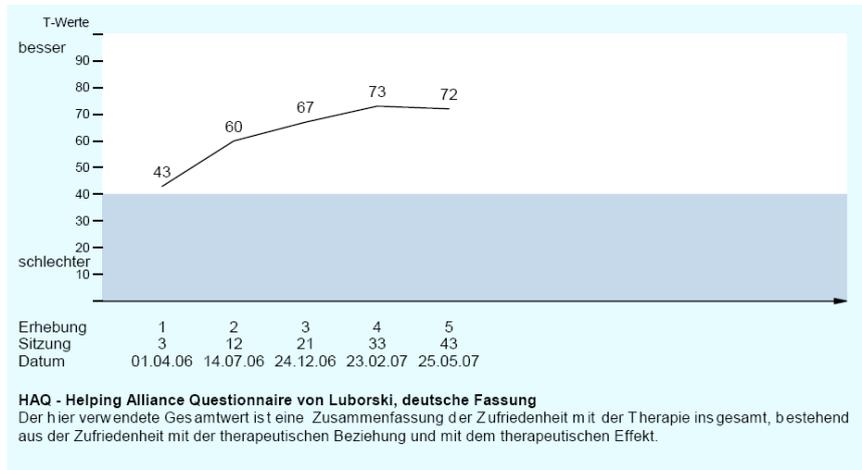
█ (W - 39;6 Jahre)  
Angst und Phobie (Verlauf)

Seite 5 von 6  
Test vom 01.07.2008 14:59

**Wie schätzt Ihre Patientin/ Ihr Patient die störungsspezifischen Beschwerden ein?**



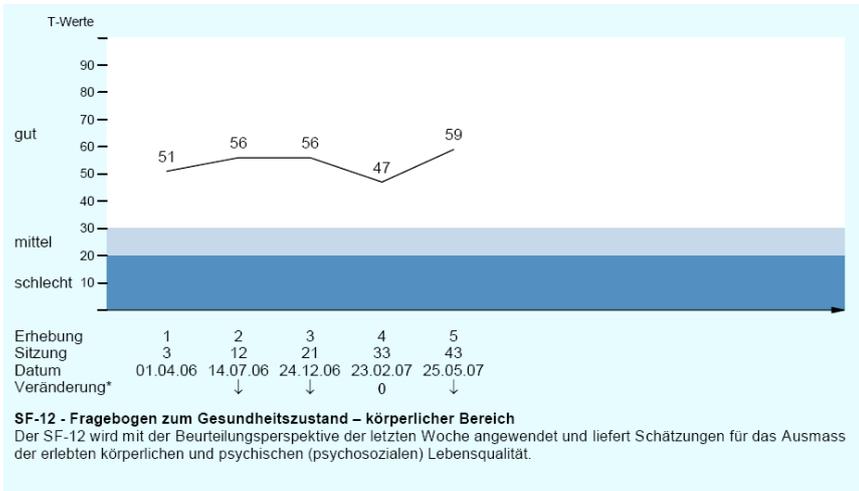
**Wie schätzt Ihre Patientin/Ihr Patient die therapeutische Arbeitsbeziehung ein?**



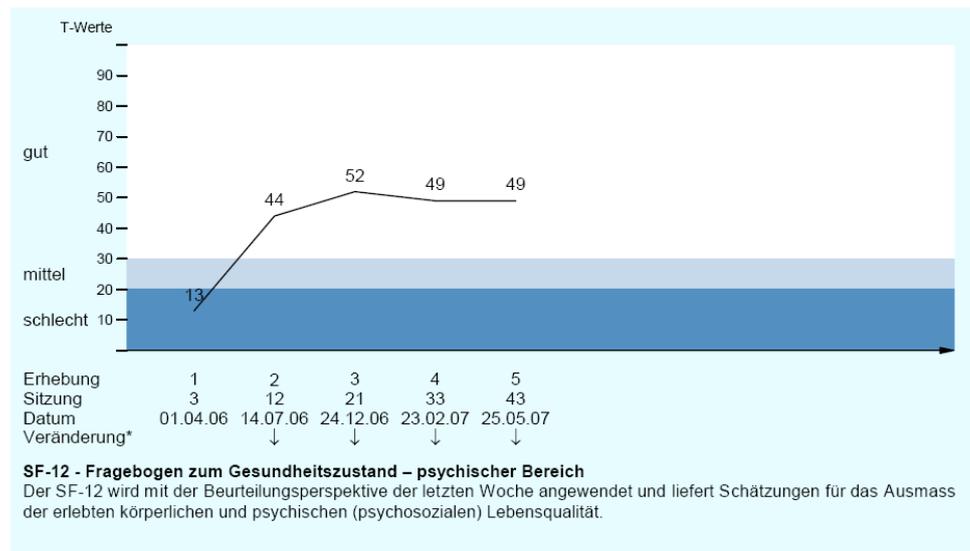
█ (W - 39;6 Jahre)  
 Angst und Phobie (Verlauf)

Seite 6 von 6  
 Test vom 01.07.2008 14:59

**Wie schätzt Ihre Patientin/ Ihr Patient die gesundheitsbezogene Lebensqualität ein?  
 - Körperlicher Bereich**



**- Psychischer Bereich**



**Impressum**

Modellvorhaben der Techniker Krankenkasse nach §63 SGB V "Qualitätsmonitoring in der ambulanten Psychotherapie"  
 Projektleitung: Dr.med. Thomas Ruprecht, Verantwortlicher Berater für Entwicklung und Umsetzung im Hogrefe TestSystem (HTS); Prof. Dr. Klaus-Dieter Hänsgen (Universität Fribourg), Entwicklung der Entscheidungsregeln und grafische Reports: Prof. Dr. Wolfgang Lutz und Dipl.-Psych. Sven Tholen (Universität Bern), Dipl.-Soz. Andreas Nagel (Techniker Krankenkasse).