

## 5. Ergebnisse und Diskussion

### 5.1. Beschreibung der Stichprobe

Im Erhebungszeitraum wurden von Februar 1998 bis Januar 2000 nach dem unter Kapitel 3. beschriebenen Verfahren insgesamt 84 Patienten in die Untersuchung aufgenommen. 17 Patienten mussten bei Überprüfung der Daten ausgeschlossen werden, davon 16 wegen ungenügender Qualität bzw. fehlender Datensätze, bei einem Patienten, welcher der Experimentalgruppe zugeordnet war, erfolgte der Ausschluss wegen diagnostischer Unklarheiten.

Es verblieb eine Gesamtstichprobe von  $n = 67$  Patienten mit den ICD-10 Hauptdiagnosen Schizophrenie (F20) oder schizoaffektiver Störung (F25).

Die Teilstichprobe der Experimentalgruppe bestand aus 15 Frauen (42,9%) und 20 Männern (57,1%), die Teilstichprobe der Kontrollgruppe aus 13 Frauen (40,6%) und 19 Männern (59,4%). Bei Eintritt in die Studie variierte das Alter in der Gesamtstichprobe zwischen 18 und 57 Jahren mit einem Mittelwert von  $M = 32,9$ .

Beide Teilstichproben erwiesen sich sowohl hinsichtlich des Geschlechterverhältnisses als auch bezüglich der Altersverteilung als ausreichend vergleichbar (siehe Tab. 1).

Geschlechtsverteilung und Alter zum Zeitpunkt U1 in Experimental- und Kontrollgruppe										
	Geschlecht				Alter zum Zeitpunkt U1					
	weiblich		männlich		Exp.	Mittelwert	N	s	Min	Max
	Anzahl	%	Anzahl	%						
Exp.	15	42,9	20	57,1	Exp.	32,92	35	7,36	20,30	54,78
Kontr.	13	40,6	19	59,4	Kontr.	32,85	32	8,30	18,59	57,13
					Gesamt	32,89	67	7,76	18,59	57,13

**Tab. 1: Geschlechtsverteilung und Alter zum Zeitpunkt U1 (Eintritt in die Studie)**

N: Anzahl der gewerteten Fälle; s: Standardabweichung.

Darüber hinaus scheint durch das matched-pair-Verfahren (siehe Punkt 3.1.2.) bei einem Eingangs-CGI von im Mittel 5,8 für die Experimentalgruppe respektive 6,1 für die Kontrollgruppe eine hohe Vergleichbarkeit hinsichtlich des Schweregrads der Erkrankung (6 = Patient ist deutlich krank) erreicht worden zu sein. Daten hierzu siehe Abschnitt 5.4..

## 5.2. Beobachtungszeitraum und Behandlungsabbrüche

Ein gefordertes Kriterium für die Aufnahme in die Studie war eine zu erwartende Behandlungsdauer von etwa drei Monaten (siehe Abschnitt 3.1.1.). Die tatsächlichen Beobachtungszeiträume (U1 bis U2) innerhalb der Stichprobe wiesen jedoch mit 7 bis 165 Tagen eine außerordentlich hohe Spannweite auf (siehe Tab. 2). Der Mittelwert lag mit 78 Tagen knapp unter dem Idealwert von 3 Monaten.

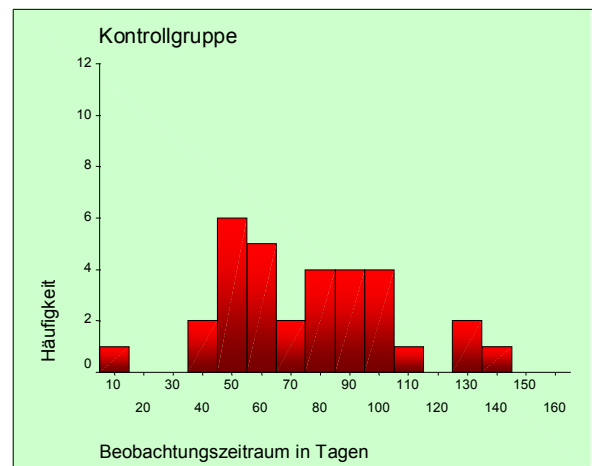
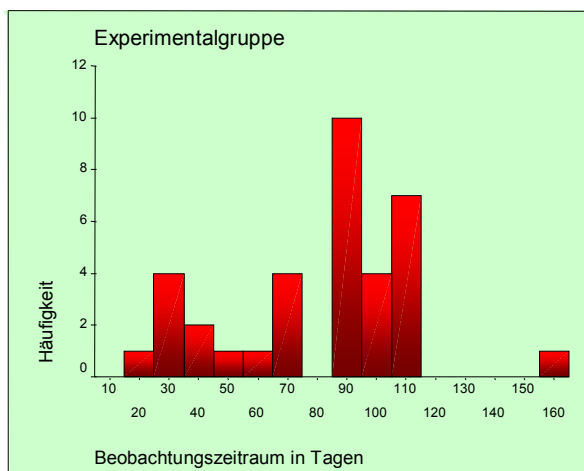
Die kürzeren Verläufe lassen sich durch vorfristige Entlassungen, durch notwendige Verlegungen und Behandlungsabbrüche erklären, die drei sehr langen Verläufe beruhen wahrscheinlich auf Versäumnissen einer zeitgerechten Nachuntersuchung in einzelnen Fällen.

Um die Aussagekraft der Studie nicht durch zu kleine Teilstichproben zu gefährden, war von einem Ausschluss abgesehen worden.

Beobachtungszeiträume in Tagen (Exp. / Kontr.-gruppe)					
	Mittelwert	N	s	Minimum	Maximum
Exp.	81,43	35	32,07	16	165
Kontr.	74,59	32	29,47	7	141
Gesamt	78,16	67	30,81	7	165

**Tab. 2: Beobachtungszeiträume in Experimental- und Kontrollgruppe – Mittelwerte**

N: Anzahl der gewerteten Fälle; s: Standardabweichung.



**Abb. 10 a + b: Beobachtungszeiträume – Häufigkeitsdiagramme**

N: Anzahl der gewerteten Fälle; s: Standardabweichung.

Die Auswertung ergibt für die Patienten der Experimentalgruppe im Mittel um nahezu 7 Tage längere Beobachtungszeiträume (Experimentalgruppe M = 81,4, Kontrollgruppe M = 74,6 Tage). Diese Tatsache wird durch zwei synergistische Phänomene bedingt: Die Häufigkeitsverteilungen (Abb. 10 a + b) zeigen – ausgehend vom erwünschten Beobachtungszeitraum von etwa drei Monaten – in der Experimentalgruppe zum einen häufiger längere andererseits seltener kürzere Beobachtungszeiträume als in der Kontrollgruppe an. Warum die Nachuntersuchung (besonders in der Experimentalgruppe) so oft erst deutlich nach den anvisierten drei Monaten Beobachtungszeitraum stattfand, ist aus den vorliegenden Daten nicht nachvollziehbar. Dagegen lassen sich die kürzeren Beobachtungszeiträume durch den folgenden Blick auf Verlegungen und Behandlungsabbrüche zumindest teilweise erklären (siehe Tab. 3 und Abb. 11 a + b): In der Experimentalgruppe (n = 35) kam es zu einer definitiven Verlegung sowie einem Therapieabbruch, zusammen 5,8% des Kollektivs (eine kurze zwischenzeitliche Verlegung wurde hierbei nicht berücksichtigt), gegenüber 2 definitiven Verlegungen und 4 Behandlungsabbrüchen, zusammen 18,8%, auf Seiten der Kontrollgruppe (n = 32). Die Signifikanzbestimmung mittels Vierfeldertafel ergibt hier ein Chi-Quadrat von 2,7; die Irrtumswahrscheinlichkeit liegt damit bei 5 % (einseitige Fragestellung).

Verlauf im Beobachtungszeitraum – Häufigkeiten in Exp.- und Kontr.-gruppe						
			Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Exp.	Gültig	durchgehend tagesklinisch	32	91,4	91,4	91,4
		zwischenzeitliche Verlegung	1	2,9	2,9	94,3
		Verlegung	1	2,9	2,9	97,1
		Abbruch	1	2,9	2,9	100,0
		Gesamt	35	100,0	100,0	
	Gesamt	35	100,0			
Kontr.	Gültig	durchgehend tagesklinisch	25	78,1	80,6	80,6
		Verlegung	2	6,3	6,5	87,1
		Abbruch	4	12,5	12,9	100,0
		Gesamt	31	96,9	100,0	
	Gesamt	32	100,0			
	Fehlend	fehlende Daten	1	3,1		
	Gesamt	1	3,1			

**Tab. 3: Therapieabbrüche im Beobachtungszeitraum - Häufigkeiten**

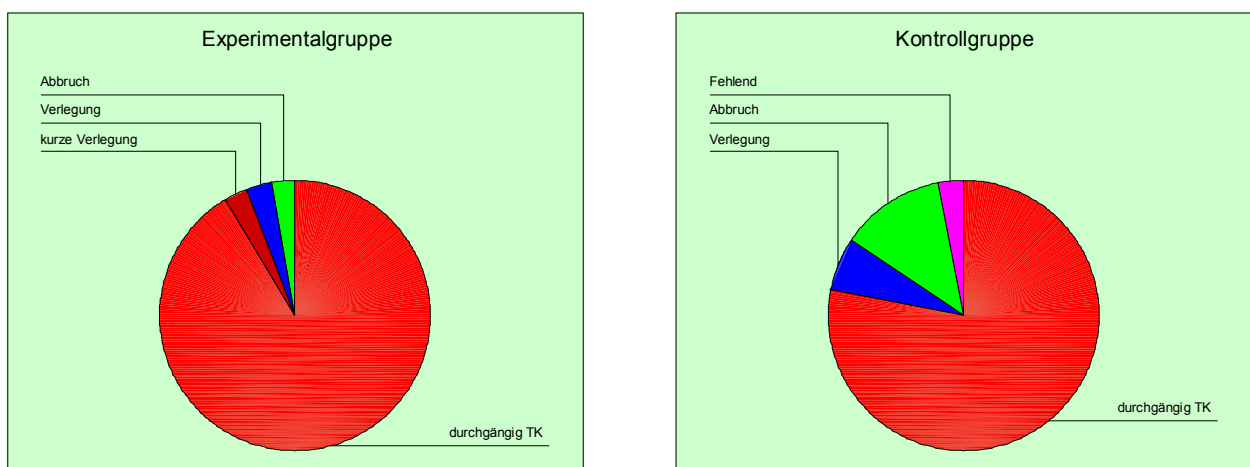
Im Vergleich mit in der Literatur berichteten Therapieabbruchraten tagesklinischer Behandlung von 28,5 und 54% [siehe 17, 46, 60, 90, 102] scheint die Rate ungeplanter Entlassungen besonders in der Experimentalgruppe mit 5,8% außerordentlich gering. Auch wenn alle 42 in dieser Studie für die Experimentalgruppe rekrutierten Probanden berücksichtigt werden, ergibt sich nach Durchsicht der Datensätze eine relativ niedrige Rate ungeplanter Entlassungen zwischen 9,5 und 21,4%, wobei der exakte Prozentsatz aufgrund fehlender Daten nicht bestimmbar war.

Bei den Kontrollpatienten (n = 42) lag die Rate ungeplanter Entlassungen zwischen 16,7 (dokumentierte Behandlungsabbrüche) und 38,0% (mögliche Behandlungsabbrüche unter Berücksichtigung fehlender Daten), über alle in die Untersuchung aufgenommenen Patienten (n = 84) zwischen 13,1 und 29,8%.

In diesen Ergebnissen ist jedoch zu berücksichtigen, dass der Rekrutierung der Probanden eine Selektion in dem Sinne vorausgegangen war, dass diese eine Behandlungsdauer von etwa drei Monaten erwarten lassen sollten.

Kritisch anzumerken ist in diesem Zusammenhang, dass in der durchgesehenen Literatur kein Konsens darüber besteht, wann von einem Therapieabbruch zu sprechen ist [vgl. z.B. 102].

**Fazit:** Die dargestellten Ergebnisse zeigen einen geringeren Anteil an ungeplanten Entlassungen aus der tagesklinischen Therapie in der Experimentalgruppe.



**Abb. 11 a + b: Verlauf im Beobachtungszeitraum – Kreisdiagramme**

### 5.3. Verweildauer

Der Aufenthalts- oder Verweildauer wird besonders seitens der Kostenträger größte Aufmerksamkeit geschenkt. Auf die Problematik der Verwendung der Aufenthaltsdauer als Indikator für Ergebnisqualität ist bereits in Abschnitt 1.1.1.2. hingewiesen worden. Auch in dieser Arbeit ergeben sich erhebliche Schwierigkeiten in der Deutung der Ergebnisse.

Daten zum Aufnahme- und Entlassungstag lagen für 62 Patienten der Gesamtstichprobe (92,5%) vor.

Tab. 4a zeigt die Ergebnisse zur Gesamtaufenthaltsdauer im Vergleich von Experimental- zur Kontrollgruppe, Tab. 4b die entsprechende Auswertung im Vergleich der Projekthälften über die Gesamtstichprobe.

<b>Aufenthaltsdauer in Tagen (Exp. / Kontr.-gruppe)</b>					
	Mittelwert	N	s	Minimum	Maximum
Experimentalgruppe	175,27	33	123,97	37	499
Kontrollgruppe	167,28	29	135,93	18	732
Gesamt	171,53	62	128,68	18	732

<b>Aufenthaltsdauer in Tagen (1. und 2. Projekthälfte)</b>					
	Mittelwert	N	s	Minimum	Maximum
1. Projekthälfte	161,13	32	97,34	37	499
2. Projekthälfte	182,63	30	156,38	18	732
Gesamt	171,53	62	128,68	18	732

**Tab. 4 a + b: Aufenthaltsdauer - Mittelwerte**

N: Anzahl der gewerteten Fälle; s: Standardabweichung.

Es war angenommen worden, dass sich die Aufenthaltsdauer durch das Projekt verkürzt (siehe 4.2.). Dies konnte in der Auswertung nicht bestätigt werden: Der Vergleich der Teilstichproben ergab eine im Mittel für die Experimentalgruppe um fast 8 Tage längere Verweildauer, im Vergleich der Projekthälften lag die gemittelte Gesamtaufenthaltsdauer in der zweiten Projekthälfte sogar um 21½ Tage höher.

Folgende Faktoren sind in der Bewertung dieser Ergebnisse zu berücksichtigen:

- Die Probanden befanden sich teilweise schon geraume Zeit in Behandlung bevor sie in das Projekt aufgenommen wurden. Mit 0 bis 527 Tagen ist die Spannweite dieser Zeiträume außerordentlich hoch. 28,5 % der Patienten befanden sich zu U1 schon über 3 Monate, 6 % über 6 Monate in Therapie. Diese Zeiten wurden in den gezeigten Ergebnissen nicht gesondert berücksichtigt.
- Über den Behandlungszeitraum außerhalb des Beobachtungszeitraums liegen keine Daten vor (z.B. auch über zwischenzeitliche Verlegungen).
- Der Anteil ungeplanter Entlassungen geht in die Ergebnisse ein, jedoch können Therapieabbrüche (siehe oben) kaum als therapeutisch erfolgreich gelten.
- Als Aufnahmekriterium in die Studie war eine zu erwartende Behandlungsdauer von etwa drei Monaten gefordert worden. Dies könnte dazu geführt haben, eher Patienten mit erwartet langwierigen Behandlungsverläufen zur Rekrutierung in die Studie vorzuschlagen, um so Frustrationen durch Behandlungsabbrüche zu vermeiden.
- Es wurden nur wenige soziodemographischer Variablen mitgeführt und es liegen keine Informationen zu Programmen, Mitarbeiterzahl, Milieutyp der einzelnen Tageskliniken etc. vor. Eine weitere Ursachenklärung der Behandlungsdauern wird dadurch erheblich erschwert.
- Wie zahlreiche Autoren betonen, bedarf die Implementierung von QS-Programmen Zeit, um Veränderungen anzuschieben und vor allem auch der Akzeptanz aller Mitarbeiter. Die Auswertung der Fragebögen für die Mitarbeiter der besuchten Tagesklinik (siehe Abschnitt 5.8.) spiegelt jedoch gerade hinsichtlich einer Verkürzung der Behandlungsdauer eine außerordentliche Skepsis der behandelnden Teams wider.

In Konsequenz des oben Genannten scheint es nicht sinnvoll, die Verweildauer als Erfolgskriterium in dieser Studie heranzuziehen.

#### **5.4. Behandlungsergebnisse – Experimental- versus Kontrollgruppe**

Eingangs war die Hypothese aufgestellt worden, die Patienten der Experimentalgruppe würden gegenüber denen der Kontrollgruppe aus peer review und Aufstellen individueller Therapieziele einen besonderen Nutzen ziehen oder – anders ausgedrückt – über den Untersuchungszeitraum „gesünder“ werden (siehe Abschnitt 4.2.).

### 5.4.1. Methode

Die Operationalisierung erfolgte über die Verlaufsentwicklung des Schweregrads der Erkrankung – CGI und der Einschätzung zum globalen Funktionsniveau – GAF (siehe Abschnitte 4.1.2. und 4.1.3.) im Vergleich der Teilstichproben.

Die vorliegenden, mit diesen Skalen (cgi 1, cgi 2a, gaf 1 und gaf 2; siehe Legende zu Tab. 5) erfassten Daten stellen jeweils Statusbeurteilungen zu den Untersuchungszeitpunkten U1 bzw. U2 dar. Um die im Sinne einer solchen indirekten Veränderungsmessung [siehe 110] gewonnenen Informationen insbesondere graphisch anschaulich darstellen zu können, wurden zwei weitere Parameter durch Differenzbildung beider Statusbeurteilungen konstruiert (diff cgi 2a-1 und diff gaf 2-1; siehe Legende zu Tab. 5); zur methodischen Problematik von Differenzwerten siehe Cronbach & Furby [23].

Bei dem Item Zustandsänderung der CGI (cgi 2b), das ebenfalls in die Operationalisierung einging, handelt es sich per definitionem [siehe 110] um eine direkte Veränderungserfassung über den Beobachtungszeitraum (siehe 4.1.2.).

		cgi 1	cgi 2a	diff cgi 2a-1	cgi 2b	gaf 1	gaf 2	diff gaf 2-1
Exp.	Mittelwert	5,83	5,00	-,83	3,74	44,43	54,86	10,43
	N	35	35	35	35	35	35	35
	s	,66	,80	,79	1,04	10,59	12,32	12,25
	Minimum	4	4	-2	2	25	30	-10
	Maximum	7	7	1	7	65	80	40
Kontr.	Mittelwert	6,13	5,53	-,59	3,94	41,78	47,31	5,53
	N	32	32	32	32	32	32	32
	s	,71	,80	,76	1,01	10,28	13,86	11,29
	Minimum	5	4	-2	2	25	5	-20
	Maximum	7	7	1	7	60	65	35
Gesamt	Mittelwert	5,97	5,25	-,72	3,84	43,16	51,25	8,09
	N	67	67	67	67	67	67	67
	s	,70	,84	,78	1,02	10,45	13,53	11,97
	Minimum	4	4	-2	2	25	5	-20
	Maximum	7	7	1	7	65	80	40

**Tab. 5: Mittelwertbericht für CGI und GAF**

N: Anzahl der gewerteten Fälle; s: Standardabweichung.

cgi 1 : Schweregrad der Erkrankung zum Zeitpunkt U1

cgi 2a : Schweregrad der Erkrankung zum Zeitpunkt U2

diff cgi 2a-1 : cgi 2a (Schweregrad der Erkrankung zum Zeitpunkt U2) - cgi 1 (Schweregrad zum Zeitpunkt U1)

cgi 2b : Zustandsänderung des Krankheitsbildes seit Aufnahme in die Untersuchung, eingeschätzt zum Zeitpunkt U2

gaf 1 : Globales Funktionsniveau zum Zeitpunkt U1

gaf 2 : Globales Funktionsniveau zum Zeitpunkt U2

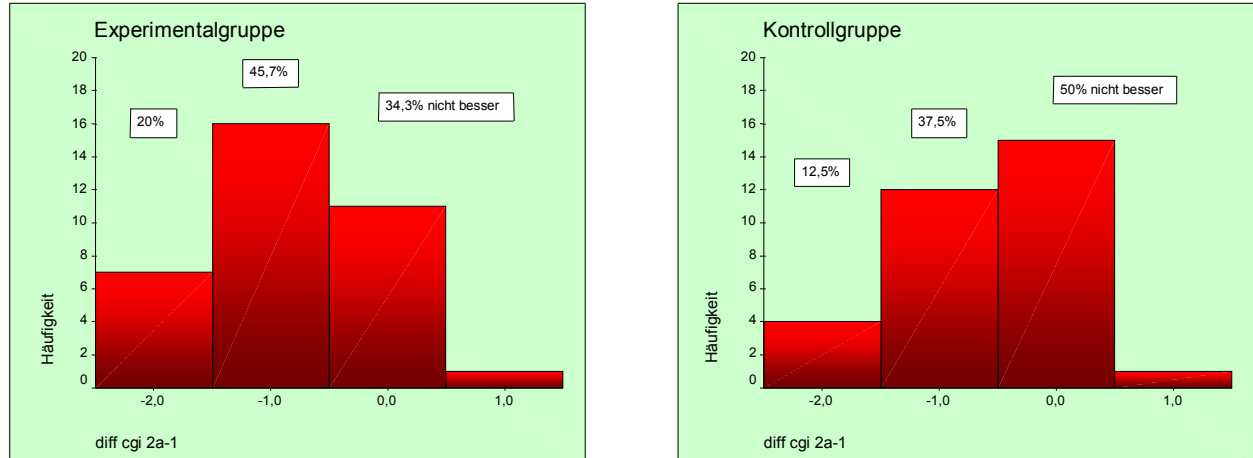
diff gaf 2-1 : gaf 2 (Globales Funktionsniveau zum Zeitpunkt U2) - gaf 1 (Globales Funktionsniveau zum Zeitpunkt U1)

## 5.4.2. Ergebnisse

### CGI – Item Schweregrad der Erkrankung

Das Niveau des Schweregrads der Erkrankung lag zum Zeitpunkt U1 in beiden Teilstichproben im Mittel bei ungefähr 6 (cgi 1 = 5,8 in der Experimentalgruppe respektive 6,1 in der Kontrollgruppe), was einer gemittelten Einschätzung der Patienten als deutlich krank nahe kommt.

In der Analyse der Mittelwerte (Tab. 5) zeigen sich für beide Gruppen Verbesserungen hinsichtlich des Items Schweregrad; ein geringerer Punktwert entspricht in der Skalenkonstruktion dabei einem geringeren Schweregrad der Erkrankung: In der Experimentalgruppe kam es im Untersuchungszeitraum zu einer Abnahme um 0,83, in der Kontrollgruppe um 0,59. Diese Veränderungen erweisen sich im T-Test für gepaarte Stichproben (Tab. 6) über den Untersuchungszeitraum als signifikant ( $p = .000$ ). Der gemittelte Schweregrad der Erkrankung zum Untersuchungsaustritt entspricht in der Experimentalgruppe einer Einschätzung als mäßig krank (cgi 2a = 5). In der Kontrollgruppe liegt dieser Wert zwischen den Einschätzungen als deutlich und mäßig krank (cgi 2a = 5,5).



**Abb. 12 a +b: Histogramme - diff cgi 2a-1**

Im Ergebnisvergleich der beiden Teilstichproben zeichnet sich tendenziell ein größerer Erfolg in der Experimentalgruppe ab: Die Darstellungen der Häufigkeitsverteilungen der einzelnen Veränderungen im Schweregrad (diff cgi 2a-1) für beide Gruppen (Abb. 12 a + b) zeigt für die Experimentalgruppe einen deutlich geringeren Anteil an nicht gebesserten (oder verschlechterten) Patienten als die Kontrollgruppe (34,3 respektive 50%), auch werden hier häufiger größere Verbesserungen erreicht.



Um zu überprüfen, ob die Gruppenzugehörigkeit einen signifikanten Effekt auf den Schweregrad der Erkrankung zum Zeitpunkt U2 (Item cgi 2a) hat, wurde eine univariate Kovarianzanalyse (basierend auf SPSS: Allgemeines lineares Modell) durchgeführt. Dabei ging neben dem festen Faktor der Gruppenzugehörigkeit der Schweregrad zum Untersuchungseintritt als Kovariate ein. Dieses Vorgehen erlaubt, den Effekt der Gruppenzugehörigkeit auf das Ergebnismaß cgi 2a nach Bereinigung um die Kontrollvariable cgi 1 zu überprüfen [Literatur zur Methode im Online-Handbuch für SPSS sowie bei Brosius, 12]. Durch den Levene-Test wurde zuvor nachgewiesen, dass sich die Fehlervarianz der abhängigen Variablen (cgi 2a) über die Gruppen hinweg nicht signifikant unterscheidet ( $F = 0,061$ ,  $p = 0,806$ ). Im Test für Zwischensubjekteffekte (Tab. 7a) konnte für cgi 2a ein signifikanter Effekt der Gruppenzugehörigkeit nachgewiesen werden ( $F = 4,216$ ,  $p = 0,044$ ).

### CGI – Item Zustandsänderung

In der Beurteilung der Zustandsänderung zeigt sich im Vergleich zu den Beurteilungen des Schweregrads zu beiden Untersuchungszeitpunkten ein geringerer prozentualer Anteil an im Zustand nicht gebesserter Patienten (Skalenwert 5 = „unverändert“, 6 = „etwas schlechter“ und 7 = viel schlechter“) von 14,3 in der Experimental- versus 21,9% in der Kontrollgruppe (Abb. 13 a + b). Auch in diesem Item werden bei den Patienten der Experimentalgruppe häufiger größere Verbesserungen erreicht. Bei einem Behandlungsergebnis von im Mittel 3,7 in der Experimentalgruppe gegenüber 3,9 in der Kontrollgruppe (Tab. 5) ergibt sich im T-Test für unabhängige Stichproben über cgi 2b kein signifikanter Unterschied in den Teilstichproben (Daten nicht gezeigt). Dabei entspricht der Wert 3 der Einschätzung „viel besser“, 4 der Einschätzung „nur wenig besser“ im Vergleich zum Zustand bei Eintritt in die Studie.

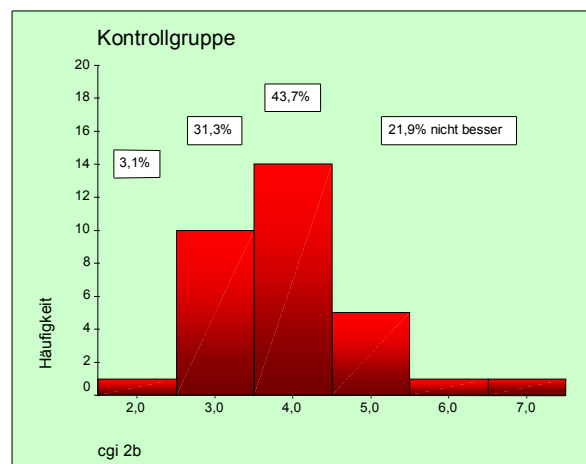
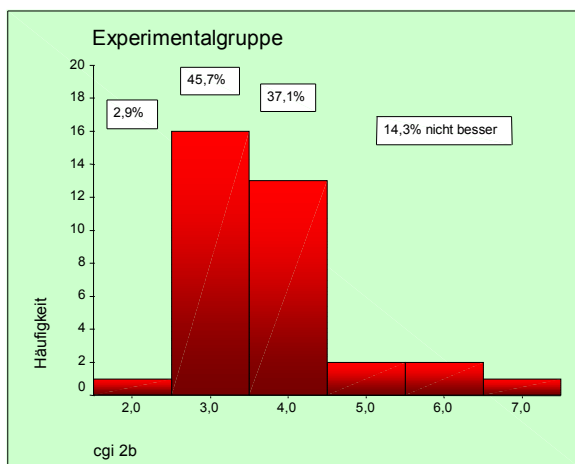


Abb. 13 a + b: Histogramme - cgi 2b

### GAF – Globales Funktionsniveau

Die Analyse der Mittelwerte (Tab. 5) zeigt auch hier im Ausgangsniveau eine gute Vergleichbarkeit beider Teilstichproben (gaf 1 = 44,4 in der Experimentalgruppe respektive 41,8 in der Kontrollgruppe). Das Globale Funktionsniveau ist im Untersuchungszeitraum nach den Ergebnissen der durchgeführten Prä-/Postmessung in beiden Gruppen signifikant gestiegen (Tab. 6).

Beschrieben durch die Differenz beider Messungen (diff gaf 2-1) ist in der Experimentalgruppe mit einem Wert von 10,4 ein fast doppelt so hoher Zuwachs im Vergleich zur Kontrollgruppe (5,5) in diesem Merkmal zu verzeichnen.

Um zu einer anschaulichen graphischen Darstellung der Häufigkeitsverteilungen von Veränderungen im Globalen Funktionsniveau beider Teilstichproben zu gelangen, musste der Informationsgehalt reduziert werden (Abb. 14 a + b): Die einzelnen Veränderungen (Werte von diff gaf 2-1) wurden in Gruppen mit einer Spannweite von 5 Punkten auf der GAF-Skala willkürlich zusammengefasst. Es zeigt sich, dass der Anteil im Funktionsniveau nicht wesentlich gebesserter und verschlechterter Patienten in der Kontrollgruppe mit 43,8% deutlich höher als in der Experimentalgruppe (25,7%) ausfällt.

Unterzieht man die Daten analog den Ausführungen zur cgi einer univariaten Kovarianzanalyse (Tab. 7b) ergibt sich auch für das Ergebnismaß globales Funktionsniveau ein signifikanter Unterschied hinsichtlich der Gruppenzugehörigkeit ( $F = 4,39$ ,  $p = 0,04$ ). Der zuvor durchgeführte Levene-Test ergab keine signifikanten Unterschiede der Fehlervarianz für gaf 2 über die Teilstichproben ( $F = 0,183$ ,  $p = 0,670$ ).

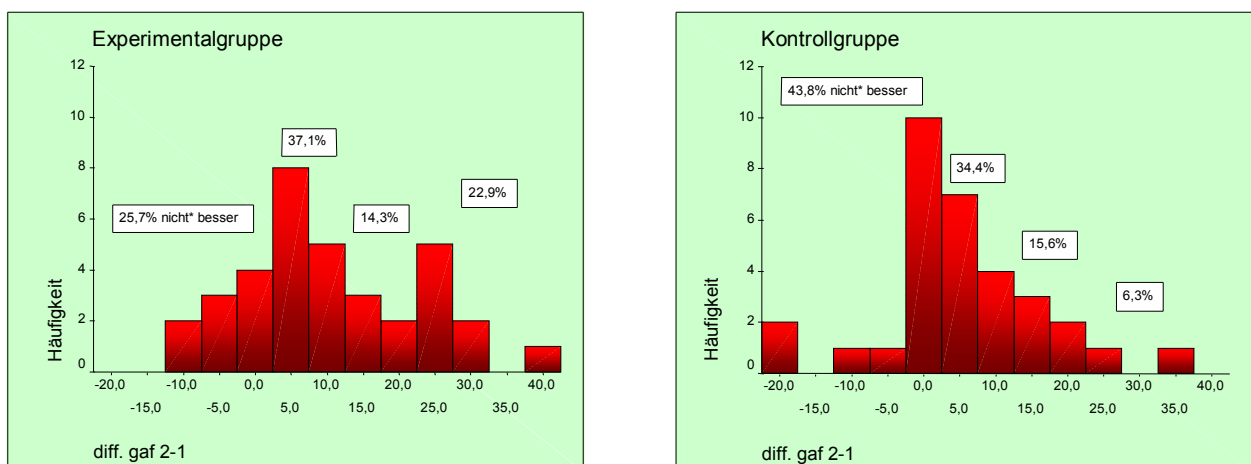


Abb. 14 a + b: Histogramme – diff gaf 2-1

Test bei gepaarten Stichproben									
	Paaren	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	Gepaarte Differenzen		T	df	Sig. (2-seitig)
					95% Konfidenzintervall der Differenz				
					Untere	Obere			
Exp.	cgi 1 – cgi 2a	,8286	,7854	,1328	,5588	1,0984	6,242	34	,000
	gaf 1 – gaf 2	-10,4286	12,2505	2,0707	-14,6368	-6,2204	-5,036	34	,000
Kontr.	cgi 1 – cgi 2a	,5938	,7560	,1336	,3212	,8663	4,443	31	,000
	gaf 1 – gaf 2	-5,5313	11,2880	1,9955	-9,6010	-1,4615	-2,772	31	,009

**Tab. 6: T-Test bei gepaarten Stichproben**

Kovarianzanalyse: Test der Zwischensubjekteffekte – abhängige Variable: cgi 2a						
Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz	
Korrigiertes Modell	14,063 <sup>a</sup>	2	7,032	13,794	,000	
Konstanter Term	3,221	1	3,221	6,320	,014	
cgi 1	9,345	1	9,345	18,333	,000	
Gruppe	2,149	1	2,149	4,216	,044	
Fehler	32,623	64	,510			
Gesamt	1896,000	67				
Korrigierte Gesamtvariation	46,687	66				

a R-Quadrat = ,301 (korrigiertes R-Quadrat = ,279)

Kovarianzanalyse: Test der Zwischensubjekteffekte – abhängige Variable: gaf 2						
Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz	
Korrigiertes Modell	3909,691 <sup>a</sup>	2	1954,846	15,326	,000	
Konstanter Term	1951,197	1	1951,197	15,298	,000	
gaf 1	2958,165	1	2958,165	23,193	,000	
Gruppe	559,887	1	559,878	4,390	,040	
Fehler	8162,996	64	127,547			
Gesamt	188078,000	67				
Korrigierte Gesamtvariation	12072,687	66				

a R-Quadrat = ,324 (korrigiertes R-Quadrat = ,303)

**Tab. 7 a + b: Kovarianzanalysen**

### **5.4.3. Allgemeine Überlegungen zu CGI und GAF – Diskussion**

Hinsichtlich Schwere der Erkrankung und psychosozialem Funktionsniveau ergeben sich bei Durchsicht der Literatur [18, 31, 38, 43, 44, 47, 53, 58, 94, 99, 118] keine eindeutigen Zusammenhänge zu bestimmten Behandlungsformen (ambulant / stationär / teilstationär). So erreichen z.B. die nicht-stationären Patienten der EFESO-Studie [44] im Mittel ähnliche Einschätzungen im globalen Funktionsniveau („ernste Symptome oder Beeinträchtigungen“) wie stationäre [58, 118] oder teilstationäre psychiatrische Patienten [31 sowie in dieser Studie].

Auf der anderen Seite konnten in einzelnen Untersuchungen Unterschiede hinsichtlich CGI und GAF in vergleichsweise nah beieinander liegenden Kategorien allgemeinspsychiatrisch-stationärer Patienten aufgezeigt werden: So fanden Richter et al. [94] nach Einstufung von 170 Patienten (davon ca. 60% mit einer Schizophrenie) in die Behandlungskategorien A1 und A2 der Psych-PV [14], einen signifikant höheren Schweregrad der Erkrankung sowie ein signifikant niedrigeres Funktionsniveau bei Patienten der Intensiv-Behandlungskategorie.

Es scheint deshalb von erheblicher Bedeutung zu sein, in welchem Rahmen Beurteilungen auf den Globalskalen erhoben werden. Auch schreibt die Anleitung zur CGI-Skala, wie sie in der vorliegenden Studie eingesetzt wurde, dem Rater vor, bei seiner Einschätzung den „gesamten Erfahrungsschatz an dieser Art von Kranken“ in Betracht zu ziehen, was je nach Klientel (und Untersucher) eine andere Verteilung der Ratingwerte zur Folge haben kann.

Inwieweit dies die Beurteilung in den einzelnen teilnehmenden Tageskliniken dieser Studie beeinflusst haben könnte, muss offen bleiben, da bezüglich der Globalskalen auf ein Einrichtungs-übergreifendes Ratertraining – nicht zuletzt aufgrund knapper Ressourcen – verzichtet worden war.

Erheblich erschwert wird die Vergleichbarkeit mit anderen Studien darüber hinaus durch unterschiedlich skalierte Versionen dieser Erhebungsinstrumente.

Hinsichtlich des Behandlungsergebnisses psychiatrischer Patienten, gemessen mit der CGI- und GAF-Skala, möchte man gemeinhin erwarten, dass akut und schwer Erkrankte – also vornehmlich die Klientel psychiatrischer Akutstationen – im Vergleich zu Gesünderen oder gar Patienten mit chronifizierten Krankheitsverläufen, wie sie häufig in der tagesklinischen Behandlung anzutreffen sind, generell größere Verbesserungen schneller erreichen. Dies spiegelt auch die Studienlage wider [z.B. 31, 38, 43, 118].

Die Problematik der eher als gering zu erwartenden Verbesserungen im tagesklinischen Behandlungsverlauf soll an folgendem Beispiel verdeutlicht werden:

Eine kanadische (unkontrollierte) Längsschnittstudie untersuchte 1992 die Anwendbarkeit und den Nutzen der GAF-Skala im tagesklinischen Setting an 59 Patienten, wovon allerdings die größte Gruppe (33,9%) den affektiven Störungen nach DSM-III-R und nur ein vergleichsweise geringer Anteil den psychotischen Störungen zugeordnet wurde (13,6%). Der mittlere GAF-Wert bei Aufnahme lag im Bereich „serious symptoms“, nach einer durchschnittlichen Behandlungsdauer von 99,2 Tagen zum Zeitpunkt der Entlassung im Bereich „moderate symptoms“. 67,8% der Patienten wurden bei Entlassung als etwas gebessert „somewhat improved“ eingeschätzt, gefolgt von „unimproved“ 13,6 und „much improved“ 11,9%. Die Autoren resümierten hierzu:

“Patients improved from admission to discharge at a high level of statistical significance. However, on average, the improvement was fewer than 6 points, an improvement that many referral sources might not find clinically significant. A reliance on the GAF Scale as an indication of improvement, then, might not be particularly effective in educational referral sources about the utility of day hospital. The GAF Scale may lack sensitivity to changes over time. On the other hand, when viewing a day hospital population, a small change may indeed be clinically significant. Perhaps even the absence of a negative change may be noteworthy. ...” [31]

Auch eine Untersuchung an 84 Patienten (davon 20,2% mit der Diagnose einer Schizophrenie) der Tagesklinik der Hochschule Hannover [102, 2001] fand nur leichte bis moderate Verbesserungen im Behandlungsverlauf, kaum profitierten Patienten mit einer Persönlichkeitsstörung oder Schizophrenie.

#### **5.4.4. Zusammenfassung der bisherigen Ergebnisse**

In beiden Teilstichproben konnten signifikante Verbesserungen im Schweregrad der Erkrankung und im psychosozialen Funktionsniveau nachgewiesen werden. Damit wird die Wirksamkeit tagesklinischer Behandlung erneut bestätigt [vgl. z.B. 17, 32, 90]. Der Grad der Verbesserungen wie auch die niedrige Rate an ungeplanten Entlassungen kann im Studienvergleich als hoch eingeschätzt werden, jedoch dürfen dabei die besonderen Anforderungen an die Probanden dieser Studie nicht außer Acht gelassen werden.

Darüber hinaus konnte in den durchgeführten Kovarianzanalysen der statistische Nachweis erbracht werden, dass zwischen Experimental- und Kontrollgruppe auf 5%-Niveau signifikante Unterschiede hinsichtlich der um die jeweiligen Ausgangswerte bereinigten Ergebnisparameter Schweregrad der Erkrankung und globalem Funktionsniveau vorliegen. Die Hypothese, dass die Patienten der Experimentalgruppe durch peer review und Aufstellen individueller Therapieziele signifikant gesünder werden als die der Kontrollgruppe, wurde bestätigt. Es ist damit der empirische Beweis erbracht, dass Qualitätssicherungsmaßnahmen in der Psychiatrie eine Verbesserung der Gesundheit von Patienten bewirken können.

Da sich aus Studiendesign und fehlenden Daten (siehe Abschnitte 5.1.-3. und Kapitel 6.) gewisse Unschärfen ergeben haben könnten, sollten diese Ergebnisse anhand größerer Stichproben überprüft werden. Anforderungen hinsichtlich Design und Durchführung zukünftiger Studien, die eine höhere Aussagekraft ermöglichen sollten, werden in Kapitel 6. aufgeführt.