

# **Ergebnisbericht QuintBox-Studie**

**Mai 2009**

Reinhold Jagsch

## **Ziel und Fragestellung**

Ziel der Studie war es zu überprüfen, ob gesunde Probanden nach vierwöchigem Tragen der QuintBox3 Veränderungen der Befindlichkeit aufweisen: Zeigen sich Veränderungen der Stimmungslage durch Einsatz der QuintBox3 nach 28 Tagen Tragedauer?

## **Untersuchungsdesign**

Die Wirksamkeit der QuintBox3 wurde mittels placebokontrollierter Doppelblindstudie mit randomisierter Zuweisung zu Verum- und Placebogruppe und durch Vergleich mit einer unbehandelten Kontrollgruppe im prä/post-Design geprüft (prä-Zeitpunkt: vor Vergabe der QuintBox3, post-Zeitpunkt: nach 28 Tage Tragedauer). Vor Vergabe der Geräte wurde ein kurzes Screeninginterview (Strukturiertes Klinisches Interview für DSM-IV, SKID) durchgeführt, um vorliegende psychische Störungen ausschließen zu können, die die Studie negativ beeinflussen könnten. Auf Basis des Body-Mass-Index und einiger gezielter Fragen wurden die Personen dem hyperergen oder normo- bzw. hypoergen Typus zugeordnet, für die die QuintBox3 differente Programme anbietet.

Zur Erhebung des Konstrukts Befindlichkeit wurden die beiden Fragebögen Beck-Depressions-Inventar (BDI) und der Fragebogen zum Gesundheitszustand (Short-Form Health Survey, SF-36; siehe Erhebungsinstrumente) eingesetzt. Die Zuordnung der Studienteilnehmer zu Placebo- und Verumgruppe erfolgte randomisiert und doppelblind. Placebo- und Verumgeräte waren äußerlich identisch, und somit war nicht erkennbar, welche Person welche Art der Quintbox3 (Verum- vs. Placeboggerät) erhielt. Während der vier Wochen Studienphase wurden die Teilnehmer aufgefordert, das Gerät täglich mindestens zehn Stunden lang eingeschaltet an einem Band um den Hals zu tragen, die Tragedauer täglich in eine vorgegebene Tabelle einzutragen sowie mögliche kritische Lebensereignisse mittels Social Readjustment Rating Scale einmal in der Woche zu prüfen und rückzumelden. Nach 28 Tagen wurden die Geräte, die Fragebögen, die Tabellen und Listen abgegeben, die Probanden bekamen abschließend für die Absolvierung der kompletten Studiendauer ein Honorar von 30 Euro ausgezahlt.

Nach Abschluss der Erhebung von Verum- und Placebogruppe wurden einer nach Alter und Geschlechterproportion parallelisierten Kontrollgruppe die beiden Fragebögen in einem zeitlichen Abstand von vier Wochen ebenfalls zweimal vorgegeben. Nachdem die Daten aller 120 Versuchspersonen erhoben waren, wurde der Randomisierungscode geöffnet, und die Probanden konnten der Verum- und Placebogruppe zugeordnet werden.

## **Erhebungsinstrumente**

### **Beck-Depressions-Inventar (BDI)**

Das BDI ist eines der weltweit am häufigsten eingesetzten Selbstbeurteilungsinstrumente. Es handelt sich dabei um ein Screeningverfahren, das eine depressive Symptomatik durch

Selbstbeurteilung zu erfassen versucht. Jeweils vier Aussagen werden zu 21 depressiven Symptomen vorgegeben. Diese spiegeln unterschiedliche Schweregrade des Symptoms wider. Die Items sollen die DSM-IV-Kriterien für eine Depression so gut wie möglich abdecken.

Die Items beziehen sich auf folgende Symptome:

- |                                     |  |
|-------------------------------------|--|
| 1) Traurige Stimmung                | 12) Interesselosigkeit                       |
| 2) Pessimismus                      | 13) Entschlussunfähigkeit                    |
| 3) Frühere Misserfolge              | 14) Wertlosigkeit                            |
| 4) Verlust von Freude               | 15) Verlust an Energie                       |
| 5) Schuldgefühle                    | 16) Veränderung der<br>Schlafgewohnheiten    |
| 6) Gefühle, bestraft zu werden      | 17) Reizbarkeit                              |
| 7) Abneigung gegen sich selbst      | 18) Veränderungen des Appetits               |
| 8) Selbstvorwürfe                   | 19) Konzentrationsschwierigkeiten            |
| 9) Selbstmordgedanken oder -wünsche | 20) Müdigkeit                                |
| 10) Weinen                          | 21) Verlust des Interesses an der Sexualität |
| 11) Unruhe                          |  |

Jedem Item kann je nach Stärke der Ausprägung ein Score von 0 bis 3 Punkten zugewiesen werden. Diese Punktwerte können zu einem Gesamtscore addiert werden. Von 0 bis 13 Punkte kann auf keine bzw. eine minimale Ausprägung von Depression geschlossen werden. Punkte von 14 bis 19 weisen auf eine leichte Ausprägung und Punkte von 20 bis 28 auf eine mittelschwere Depression hin. Ab 29 Punkten lässt sich auf eine schwere Depression schließen.

Die Durchführung bzw. das Ausfüllen des Fragebogens erfordert etwa fünf bis zehn Minuten Zeit. Zahlreiche Studien zeigen, dass das Verfahren zuverlässig, valide und änderungssensitiv ist. Die innere Konsistenz liegt bei Werten von über 0,89. In der klinischen Praxis wird es als wertvolles Instrument betrachtet und daher in großem Ausmaß eingesetzt.

### Short-Form Health Survey (SF-36)

Die SF-36 ist ein generisches, d.h. krankheitsübergreifendes und international häufig verwendetes Instrument zur Erfassung der gesundheitsbezogenen Lebensqualität. Der Fragebogen kann sowohl bei kranken als auch bei gesunden Populationen eingesetzt werden. Bei der SF-36 geben die Personen selbst Auskunft über ihr Befinden und ihre Funktionsfähigkeit im Sinne einer Selbstbewertung. Insgesamt besteht der Fragebogen aus 36 Items mit zwei- bis sechsfach abgestuften Antwortmöglichkeiten. Die Items können folgenden Subskalen zugeordnet werden (Beschreibungen aus dem Manual zur SF-36):

- 1. Körperliche Funktionsfähigkeit:** Diese Skala besteht aus zehn Items und erfasst das Ausmaß, in dem der Gesundheitszustand körperliche Aktivitäten (wie z.B. Selbstversorgung, Gehen, Bücken, anstrengende Tätigkeiten) beeinträchtigt. Höhere Werte deuten auf eine bessere Funktionsfähigkeit der Person hin.
- 2. Die Skala Körperliche Rollenfunktion,** die durch vier Items erhoben wird, spiegelt das Ausmaß wider, in dem der Gesundheitszustand die Arbeit oder tägliche Aktivitäten beeinträchtigt (z.B. weniger schaffen als gewöhnlich, Einschränkungen in der Art der Aktivitäten, Schwierigkeiten, bestimmte Tätigkeiten auszuführen). Durch hohe Werte zeigt sich gute körperliche Rollenfunktion.
- 3. Körperliche Schmerzen:** Diese Subskala setzt sich aus zwei Items zusammen und erfasst das Ausmaß, in dem Schmerzen Einfluss auf die Ausübung der normalen Arbeit sowohl im Haus als

auch außerhalb des Hauses nimmt. Ein geringes Ausmaß an Schmerzen wird durch einen hohen Wert in dieser Skala charakterisiert.

4. **Allgemeine Gesundheitswahrnehmung:** Fünf Items dienen der Erfassung der persönlichen Beurteilung der Gesundheit, einschließlich des aktuellen Gesundheitszustands wie auch zukünftiger Erwartungen etc. Ein hoher Wert zeigt eine gute allgemeine Gesundheitswahrnehmung.

5. **Vitalität:** Die vier Items der Skala erfassen die Bandbreite zwischen den Polen sich energiegeladener und voller Schwung fühlen vs. sich müde und erschöpft fühlen. Ein hoher Wert deutet auf ein hohes Ausmaß an Vitalität hin.

6. **Soziale Funktionsfähigkeit:** Diese Skala besteht aus insgesamt zwei Items und erfasst das Ausmaß, in dem die körperliche Gesundheit oder emotionale Probleme normale soziale Aktivitäten beeinflussen. Gute soziale Funktionsfähigkeit wird durch einen hohen Wert in dieser Subskala angezeigt.

7. Die Subskala **Emotionale Rollenfunktion** besteht aus drei Items. Sie beschreibt das Ausmaß, in dem emotionale Probleme die Arbeit bzw. tägliche Aktivitäten beeinträchtigen (z.B. weniger schaffen, nicht so sorgfältig wie üblich arbeiten). Hohe Werte in dieser Skala deuten auf gute emotionale Rollenfunktion hin.

8. **Psychisches Wohlbefinden:** Die Skala mit fünf Items dient der Beurteilung der allgemeinen psychischen Gesundheit. Hohe Werte in dieser Skala sind Ausdruck einer allgemeinen positiven Gemütsstimmung.

Mit einem zusätzlichen Item kann eine Beurteilung des aktuellen Gesundheitszustandes im Vergleich zum vergangenen Jahr vorgenommen werden.

Die ersten vier Subskalen können zu einer körperlichen und die letzten vier zu einer psychosozialen Summenskala zusammengefasst werden. Jede Skala kann Werte von 0–100 annehmen, wobei hohe Werte Ausdruck einer hohen gesundheitsbezogenen Lebensqualität sind. Die Bearbeitungsdauer des Fragebogens liegt bei etwa zehn Minuten, Durchführungs- und Auswertungsobjektivität des Verfahrens ist gewährleistet. Die internen Konsistenzen der Normstichprobe (N=2914) liegen zwischen 0,74 und 0,94 und können somit als zufriedenstellend bewertet werden. Die SF-36 verfügt über inhaltlich-logische Gültigkeit, die konvergente Validität wurde über Korrelationen mit dem Nottingham Health Profile (NHP) bestätigt. Dabei ergaben sich hohe Korrelationen zwischen jeweils inhaltlich vergleichbaren Subskalen (beide Instrumente siehe Anhang).

## Statistische Verfahren

Für die allgemeine Stichprobenbeschreibung kamen deskriptivstatistische Kenngrößen (Mittelwert, Standardabweichung (SD), Spannweite (Range als Differenz zwischen Maximum minus Minimum) bzw. absolute und relative Häufigkeiten zur Anwendung. Zur Überprüfung der Homogenität der drei Stichproben bezüglich der soziodemographischen Variablen kamen inferenzstatistische Methoden zum Einsatz, der Chi-Quadrat-Test für kategoriale Daten sowie im Falle kontinuierlicher Daten einfache Varianzanalyse (ANOVA) für den Mittelwertsvergleich für mehr als zwei Gruppen bzw. der t-Test für unabhängige Stichproben für den Mittelwertsvergleich von zwei Stichproben (z.B. im Falle des Body-Mass-Index).

Die eigentliche Hypothesenprüfung erfolgte mittels General Linear Model (GLM) for Repeated

Measures. Dabei handelt es sich um eine Methode zur Überprüfung von Mittelwertsunterschieden zwischen Gruppen, kombiniert mit einem zeitlichen Wiederholungsfaktor (prä/post-Design). Mit Hilfe des GLM for Repeated Measures lassen sich Gruppeneffekt, Zeiteffekt sowie Wechselwirkung Gruppe x Zeit simultan prüfen. Voraussetzung für die Durchführung eines GLM for Repeated Measures sind Intervallskalenniveau der Daten, Homogenität der Kovarianzmatrizen (Box's-Test), Homogenität der Varianzen (Levene-Test) sowie multiple Normalverteilung der Daten (Kolmogorov-Smirnov-Test bei Teilstichprobengrößen  $< 30$ ). Zeigte sich eine dieser Voraussetzungsannahmen verletzt, konnten Gruppen- und Zeiteffekte ohne Wechselwirkung nur getrennt erhoben werden. Für die Prüfung von Gruppeneffekten kamen Einfache Varianzanalysen (ANOVA) zum Einsatz, post-hoc-Vergleiche erfolgten parametrisch unter Verwendung von Bonferroni-Alphafehlerkorrektur. Bei Verletzung der Varianzenhomogenität wurde die parameterfreie Prüfung mittels Rangvarianzanalyse nach Kruskal und Wallis angewandt. Zur Prüfung von Zeiteffekten je Teilstichprobe wurden t-Tests für abhängige Stichproben durchgeführt. Die Hypothesenprüfung erfolgte zweiseitig, als Signifikanzniveau wurde ein Alpha von 0,05 festgelegt.

Zusätzlich wurden Effektstärken mittels Cohen's d erhoben, die die inhaltliche Relevanz der zeitlichen Veränderungen darstellen sollten. Grundsätzlich gelten Effektstärken bis 0,20 als vernachlässigbar, zwischen 0,20 und 0,50 als moderat, zwischen 0,50 bis 0,80 als mittelstark und ab 0,80 als stark.

## **Stichprobenbeschreibung**

Über einen Zeitraum von etwa drei Monaten wurden insgesamt 120 Studienteilnehmer rekrutiert. Nach Durchführung des Screeninginterviews aus dem SKID sowie Aufklärung über das Procedere und Unterzeichnung der Einverständniserklärung zur Studienteilnahme wurden 80 Probanden mittels Randomisierung den beiden Gruppen (Verum, Placebo) zugeteilt. Bei den 40 Studienteilnehmern der Kontrollgruppe erfolgte kein Screening und keine randomisierte Zuweisung. Die Kontrollgruppe wurde nach Alter und Geschlecht parallelisiert, um eine möglichst homogene Gesamtstichprobe zu erhalten. Kein Proband erfüllte irgendwelche Kriterien einer psychischen Störung, die Studienteilnehmer waren bis auf einige alterstypische Symptome (etwa Gelenksschmerzen) körperlich gesund.

## **Beschreibung der Gesamtstichprobe**

Insgesamt wurden 120 Personen (60 Frauen, 60 Männer) in die Studie aufgenommen, da sie alle Einschlusskriterien (Screening hinsichtlich psychischer Störungen, keine schweren gesundheitlichen Probleme bzw. Krankheiten) erfüllten. Das mittlere Alter der gesamten Stichprobe betrug 38,73 Jahre ( $SD=11,51$ ), wobei der jüngste Studienteilnehmer 19 Jahre, der älteste 58 Jahre alt war (Range 39 Jahre), somit die ganze Bandbreite der erwachsenen Bevölkerung bis 60 Jahre abgedeckt war.

Die Studienteilnehmer wurden vier Altersgruppen mit etwa gleich großen Altersranges zugeteilt (18–29, 30–39, 40–49 und 50–60 Jahre). 32 Personen (26,7%) waren zwischen 18 und 29 Jahre alt, 28 Personen (23,3%) zwischen 30 und 39 Jahre, 30 Probanden zwischen 40 und 49 (25,0%) und 30 (25,0%) zwischen 50 und 60 Jahre alt. Abbildung 1 zeigt die Häufigkeiten gesamt, sowie für Männer und Frauen getrennt.

Bezüglich Bildungsstatus (höchste abgeschlossene Schulbildung) hatten fünf Personen (4,2%)

lediglich Pflichtschulabschluss, 76 Studienteilnehmer (63,3%) eine Lehre absolviert, 33 Personen (27,5%) Matura und sechs Probanden (5,0%) einen Studienabschluss.

Insgesamt 14 Personen, i.e. 17,5% der 80 Teilnehmer der Verum- und Placebogruppe (vier Frauen, zehn Männer), wurden dem hyperergen Typus zugeordnet, die übrigen 66 Studienteilnehmer waren von normoergem bzw. hypoergem Typus.

## Vergleich der Studiengruppen

Ein Studienteilnehmer musste nach zwei Wochen aus der Studie ausgeschlossen werden, da ein kritisches Lebensereignis eingetreten war („Verletzung oder Krankheit“). Drei weitere Probanden berichteten Probleme mit dem Ein-/Ausschalten der QuintBox3 und wurden deshalb aus der Analyse ausgeschlossen, da nicht gewährleistet war, dass die Geräte täglich mindestens zehn Stunden in Betrieb waren. Alle anderen Teilnehmer der Verum- bzw. der Placebogruppe beendeten die Studie und wurden in der statistischen Analyse berücksichtigt. Damit bestand die Studienpopulation schlussendlich aus insgesamt 38 Personen der Verum-, 38 Personen der Placebo- und 40 Personen der Kontrollgruppe.

Bezüglich des Alters der Studienteilnehmer zeigten sich keine Unterschiede zwischen den drei Gruppen ( $F=0,524$ ,  $df=2$ ,  $p=0,593$ ). In der Verumgruppe betrug das mittlere Alter 37,00 Jahre ( $SD=11,05$ ), in der Placebogruppe 39,29 Jahre ( $SD=11,74$ ) und in der Kontrollgruppe 39,35 Jahre ( $SD=11,56$ ). Auch hinsichtlich der Verteilung der Geschlechterproportionen sowie des Bildungsstandes zeigte sich zwischen den Gruppen kein signifikanter Unterschied (Geschlecht:  $\chi^2=0,211$ ,  $df=2$ ,  $p=0,900$ ; Bildung:  $\chi^2=0,825$ ,  $df=4$ ,  $p=0,935$ ). Der durchschnittliche Body-Mass-Index in der Verumgruppe betrug 25,23 ( $SD=4,37$ ) und in der Placebogruppe 24,87 ( $SD=5,17$ ). Der Mittelwertsvergleich erwies sich als nicht signifikant ( $t=0,329$ ,  $df=74$ ,  $p=0,743$ ).

Sieben der 13 hyperergen Personen (ein aus der Analyse ausgeschlossener Proband war hypererg) wurden der Verum- (53,8%), sechs (46,2%) der Placebogruppe zugeteilt, das Ergebnis ist nicht signifikant ( $\chi^2=0,093$ ,  $df=1$ ,  $p=0,761$ ). Hypererge Personen wiesen mit 32,59 ( $SD=4,01$ ) einen signifikant höheren BMI als die normoergen Studienteilnehmer mit 23,49 ( $SD=3,14$ ) auf ( $t=9,049$ ,  $df=74$ ,  $p<0,001$ ).

Die durchschnittliche tägliche Tragedauer der Geräte betrug in der Verumgruppe 11,20 Stunden ( $SD=1,12$ ) und in der Placebogruppe 10,95 Stunden ( $SD=0,89$ ), die Differenz war nicht signifikant ( $t=1,064$ ,  $df=74$ ,  $p=0,291$ ).

## Ergebnisse der Hypothesenprüfung

### Ergebnisse zum BDI

Da sich die Kovarianzmatrizen als nicht homogen erwiesen, konnte die Auswertung bezüglich des Summenscores des BDI nicht mittels GLM for Repeated Measures erfolgen. Weder zum ersten ( $F=1,692$ ,  $df=2$ ,  $p=0,189$ ) noch zum zweiten Testzeitpunkt ( $\chi^2=1,501$ ,  $df=2$ ,  $p=0,472$ ) konnten Gruppenunterschiede erhoben werden. Im Zeitvergleich zeigte sich nur für die Verumgruppe eine signifikante Reduktion der BMI-Scores ( $t=2,231$ ,  $df=37$ ,  $p=0,032$ ;  $d=0,40$ ), während weder für Placebo- ( $t=0,792$ ,  $df=37$ ,  $p=0,434$ ;  $d=0,08$ ) noch für die Kontrollgruppe ( $t=1,267$ ,  $df=39$ ,  $p=0,213$ ;  $d=0,19$ ) Veränderungen zu verzeichnen waren (siehe Abbildung 2).

## Ergebnisse zur SF-36

### Physische Subskalen

Die erste Subskala, die **Körperliche Funktionsfähigkeit**, misst das Ausmaß, in dem der Gesundheitszustand körperliche Aktivitäten (wie z.B. Selbstversorgung, Gehen, Bücken, anstrengende Tätigkeiten) beeinträchtigt.

Aufgrund der Verletzung der Voraussetzung der Kovarianzenhomogenität konnte die Anwendung des GLM for Repeated Measures nicht erfolgen. Weder zum prä- ( $F=1,977$ ,  $df=2$ ,  $p=0,143$ ) noch zum post-Zeitpunkt der Erhebung ( $F=0,890$ ,  $df=2$ ,  $p=0,414$ ) konnten Gruppenunterschiede festgestellt werden. Auch die Zeitvergleiche erwiesen sich für diese Subskala für die Verum- ( $t=0,374$ ,  $df=37$ ,  $p=0,711$ ;  $d=-0,03$ ) und für die Placebogruppe ( $t=-1,676$ ,  $df=37$ ,  $p=0,102$ ;  $d=0,10$ ) als insignifikant. Für die Kontrollgruppe zeigte sich zwar ein signifikanter Zeiteffekt ( $t=-3,122$ ,  $df=39$ ,  $p=0,003$ ), mit einer Effektstärke von  $d=0,12$  kann jedoch nicht von einer inhaltlich relevanten Veränderung gesprochen werden (siehe Abbildung 3).

Die zweite Subskala, die **Körperliche Rollenfunktion**, spiegelt das Ausmaß wider, in dem der Gesundheitszustand die Arbeit oder tägliche Aktivitäten beeinträchtigt (z.B. weniger schaffen als gewöhnlich, Einschränkungen in der Art der Aktivitäten, Schwierigkeiten, bestimmte Tätigkeiten auszuführen).

Auch hier konnte aufgrund der Verletzung der Voraussetzung der Kovarianzenhomogenität die Anwendung des GLM for Repeated Measures nicht erfolgen. Weder zum prä- ( $\chi^2=0,251$ ,  $df=2$ ,  $p=0,882$ ) noch zum post-Zeitpunkt der Erhebung ( $F=0,944$ ,  $df=2$ ,  $p=0,392$ ) konnten Gruppenunterschiede festgestellt werden. Auch die Zeitvergleiche zeigten für diese Subskala kein signifikantes Resultat, weder für die Verum- ( $t=0,941$ ,  $df=37$ ,  $p=0,353$ ;  $d=-0,22$ ) noch für die Placebo- ( $t=-0,298$ ,  $df=37$ ,  $p=0,767$ ;  $d=0,03$ ) oder die Kontrollgruppe ( $t=1,482$ ,  $df=39$ ,  $p=0,146$ ;  $d=-0,25$ ) konnten Veränderungen über die Zeit festgestellt werden (siehe Abbildung 4).

Die dritte Subskala **Körperliche Schmerzen** erfasst das Ausmaß, in dem Schmerzen Einfluss auf die Ausübung der normalen Arbeit sowohl im Haus als auch außerhalb des Hauses nimmt.

Auch hier konnte aufgrund der Verletzung der Voraussetzung der Kovarianzenhomogenität die Anwendung des GLM for Repeated Measures nicht erfolgen. Weder zum prä- ( $F=0,776$ ,  $df=2$ ,  $p=0,463$ ) noch zum post-Zeitpunkt der Erhebung ( $F=0,002$ ,  $df=2$ ,  $p=0,998$ ) konnten Gruppenunterschiede festgestellt werden. Auch die Zeitvergleiche erwiesen sich für diese Subskala für die Verum- ( $t=0,251$ ,  $df=37$ ,  $p=0,803$ ;  $d=-0,03$ ) und für die Kontrollgruppe ( $t=-1,176$ ,  $df=39$ ,  $p=0,247$ ;  $d=0,16$ ) als insignifikant. Für die Placebogruppe zeigte sich zwar ein signifikanter Zeiteffekt ( $t=-2,420$ ,  $df=37$ ,  $p=0,021$ ), mit einer Effektstärke von  $d=0,27$  kann jedoch nicht von einer inhaltlich relevanten Veränderung gesprochen werden (siehe Abbildung 5).

Die vierte Subskala **Allgemeine Gesundheitswahrnehmung** der Erfassung der persönlichen Beurteilung der Gesundheit einschließlich des aktuellen Gesundheitszustands wie auch zukünftiger Erwartungen.

Bei der Berechnung dieser Subskala waren alle Voraussetzungen für die Durchführung des GLM for Repeated Measures gegeben. Die Analyse erbrachte einen signifikanten Gruppeneffekt

( $F=4,064$ ,  $df=2$ ,  $p=0,020$ ), einen signifikanten Zeiteffekt ( $F=5,584$ ,  $df=1$ ,  $p=0,020$ ), die Wechselwirkung Gruppe x Zeit war jedoch nicht statistisch signifikant ( $F=0,682$ ,  $df=2$ ,  $p=0,508$ ). Die Detailanalyse zeigte signifikante Gruppenunterschiede zu beiden Erhebungszeitpunkten, sowohl zum prä- ( $F=3,520$ ,  $df=2$ ,  $p=0,033$ ) als auch zum post-Zeitpunkt ( $F=3,474$ ,  $df=2$ ,  $p=0,034$ ), wobei beim ersten Zeitpunkt die Verumgruppe signifikant bessere Wert als die Kontrollgruppe aufweist ( $p=0,033$ ) und beim zweiten Zeitpunkt tendenziell bessere Werte sowohl als die Kontrollgruppe ( $p=0,090$ ) als auch als die Placebogruppe ( $p=0,059$ ) aufweist.

Die Detailanalyse des Zeiteffektes hingegen weist nur für die Kontrollgruppe eine signifikante Verbesserung über die Zeit aus ( $t=-2,912$ ,  $df=39$ ,  $p=0,006$ ), jedoch zeigt ein  $d=0,29$  lediglich eine moderate Effektstärke an. Die Veränderungen sowohl für die Verum- ( $t=-1,311$ ,  $df=37$ ,  $p=0,198$ ;  $d=0,20$ ) als auch für die Placebogruppe ( $t=-0,456$ ,  $df=37$ ,  $p=0,651$ ;  $d=0,07$ ) erreichten nicht Signifikanzniveau (siehe Abbildung 6).

### Physische Summenskala

Auch bei der Berechnung der Physischen Summenskala waren alle Voraussetzungen für die Durchführung des GLM for Repeated Measures gegeben. Die Analyse erbrachte weder einen signifikanten Gruppeneffekt ( $F=1685$ ,  $df=2$ ,  $p=0,190$ ) noch einen signifikanten Zeiteffekt ( $F=1,828$ ,  $df=1$ ,  $p=0,179$ ) noch eine signifikante Wechselwirkung Gruppe x Zeit ( $F=0,941$ ,  $df=2$ ,  $p=0,393$ ). Die einzelnen Gruppenmittelwerte zu den beiden Zeitpunkten sind in Abbildung 7 zu ersehen.

### Psychosoziale Subskalen

Die fünfte Subskala **Vitalität** die Bandbreite zwischen den Polen „sich energiegelad und voller Schwung fühlen“ vs. „sich müde und erschöpft fühlen“. Ein hoher Wert weist auf ein hohes Ausmaß an Vitalität hin.

Bei der Berechnung dieser Subskala waren alle Voraussetzungen für die Durchführung des GLM for Repeated Measures gegeben. Die Analyse erbrachte einen tendenziell signifikanten Gruppeneffekt ( $F=2,456$ ,  $df=2$ ,  $p=0,090$ ), einen hoch signifikanten Zeiteffekt ( $p=14,219$ ,  $df=1$ ,  $p<0,001$ ), die Wechselwirkung Gruppe x Zeit war jedoch nicht statistisch signifikant ( $F=0,799$ ,  $df=2$ ,  $p=0,452$ ). Die Detailanalyse zeigte keine signifikante Gruppenunterschiede zum prä- ( $F=1,154$ ,  $df=2$ ,  $p=0,319$ ), sehr wohl aber zum post-Zeitpunkt ( $F=3,539$ ,  $df=2$ ,  $p=0,032$ ), wobei die Verumgruppe signifikant bessere Wert als die Placebogruppe aufweist ( $p=0,043$ ), der Vergleich mit der Kontrollgruppe im post-hoc-Test nur knapp nicht Tendenzniveau ( $p=0,121$ ) erreicht.

Die Detailanalyse des Zeiteffektes weist nur für die Verumgruppe eine signifikante Verbesserung über die Zeit auf ( $t=-3,325$ ,  $df=37$ ,  $p=0,002$ ;  $d=0,38$ ). Weder die Veränderung für die Placebo- ( $t=-1,428$ ,  $df=37$ ,  $p=0,162$ ;  $d=0,16$ ) noch für die Kontrollgruppe ( $t=-1,863$ ,  $df=39$ ,  $p=0,070$ ;  $d=0,25$ ) erreicht Signifikanzniveau (siehe Abbildung 8).

Die sechste Subskala **Soziale Funktionsfähigkeit** erfasst das Ausmaß, in dem die körperliche Gesundheit oder emotionale Probleme normale soziale Aktivitäten beeinflussen.

Bei der Berechnung dieser Subskala waren alle Voraussetzungen für die Durchführung des GLM for Repeated Measures gegeben. Die Analyse erbrachte keinen signifikanten Gruppeneffekt ( $F=0,018$ ,  $df=2$ ,  $p=0,983$ ), einen signifikanten Zeiteffekt ( $p=8,216$ ,  $df=1$ ,  $p=0,005$ ), die Wechselwirkung Gruppe x Zeit war nicht statistisch signifikant ( $F=0,138$ ,  $df=2$ ,  $p=0,871$ ).



Die Detailanalyse des Zeiteffektes zeigt, dass für keine der drei Gruppen Signifikanzniveau erreicht werden kann, weder für die Verumgruppe ( $t=-1,356$ ,  $df=37$ ,  $p=0,183$ ;  $d=0,20$ ) noch für die Placebo- ( $t=-1,968$ ,  $df=37$ ,  $p=0,057$ ;  $d=0,25$ ) noch für die Kontrollgruppe ( $t=-1,729$ ,  $df=39$ ,  $p=0,092$ ;  $d=0,28$ ). Abbildung 9 zeigt die parallelen Zeitverläufe für die drei Gruppen.

Die siebente Subskala **Emotionale Rollenfunktion** beschreibt das Ausmaß, in dem emotionale Probleme die Arbeit bzw. tägliche Aktivitäten beeinträchtigen (z.B. weniger schaffen, nicht so sorgfältig wie üblich arbeiten).

Aufgrund der Verletzung der Voraussetzung der Kovarianzenhomogenität konnte die Anwendung des GLM for Repeated Measures nicht erfolgen. Weder zum prä- ( $F=0,416$ ,  $df=2$ ,  $p=0,661$ ) noch zum post-Zeitpunkt der Erhebung ( $\chi^2=2,115$ ,  $df=2$ ,  $p=0,347$ ) konnten Gruppenunterschiede festgestellt werden.

Die Detailanalyse des Zeiteffektes weist nur für die Verumgruppe eine signifikante Verbesserung über die Zeit auf ( $t=-2,135$ ,  $df=37$ ,  $p=0,039$ ;  $d=0,45$ ). Weder die Veränderung für die Placebo- ( $t=-1,404$ ,  $df=37$ ,  $p=0,169$ ;  $d=0,18$ ) noch für die Kontrollgruppe ( $t=-1,356$ ,  $df=39$ ,  $p=0,183$ ;  $d=0,17$ ) erreicht Signifikanzniveau (siehe Abbildung 10).

Die achte Subskala **Psychisches Wohlbefinden** der Beurteilung der allgemeinen psychischen Gesundheit. Hohe Werte in dieser Skala sind Ausdruck einer allgemeinen positiven Gemütsstimmung.

Aufgrund der Verletzung der Voraussetzung der Kovarianzenhomogenität konnte die Anwendung des GLM for Repeated Measures nicht erfolgen. Weder zum prä- ( $F=0,566$ ,  $df=2$ ,  $p=0,569$ ) noch zum post-Zeitpunkt der Erhebung ( $F=0,857$ ,  $df=2$ ,  $p=0,427$ ) konnten Gruppenunterschiede festgestellt werden.

Die Detailanalyse des Zeiteffektes weist nur für die Verumgruppe eine signifikante Verbesserung über die Zeit auf ( $t=-2,464$ ,  $df=37$ ,  $p=0,019$ ;  $d=0,30$ ). Weder die Veränderung für die Placebo- ( $t=-1,466$ ,  $df=37$ ,  $p=0,151$ ;  $d=0,19$ ) noch für die Kontrollgruppe ( $t=-1,803$ ,  $df=39$ ,  $p=0,079$ ;  $d=0,28$ ) erreicht Signifikanzniveau (siehe Abbildung 11).

## Psychosoziale Summenskala

Aufgrund der Verletzung der Voraussetzung der Kovarianzenhomogenität konnte die Anwendung des GLM for Repeated Measures nicht erfolgen. Weder zum prä- ( $F=0,329$ ,  $df=2$ ,  $p=0,720$ ) noch zum post-Zeitpunkt der Erhebung ( $F=1,602$ ,  $df=2$ ,  $p=0,206$ ) konnten Gruppenunterschiede festgestellt werden.

Die Detailanalyse des Zeiteffektes weist sowohl für die Verumgruppe ( $t=-3,521$ ,  $df=37$ ,  $p=0,001$ ;  $d=0,48$ ) als auch für die Placebogruppe ( $t=-2,389$ ,  $df=37$ ,  $p=0,022$ ) eine signifikante Verbesserung über die Zeit auf, wobei sich letztere mit einem  $d=0,23$  von eingeschränkter inhaltlicher Relevanz zeigt. Die Veränderung für die Kontrollgruppe ( $t=-1,943$ ,  $df=39$ ,  $p=0,059$ ;  $d=0,30$ ) hingegen erreicht nicht Signifikanzniveau (siehe Abbildung 12).

## Boden- und Deckeneffekte

Bei Betrachtung der Abbildungen der einzelnen Skalen ist zu erkennen, dass die Gruppenmittelwerte zu einem großen Teil in den untersten (BDI) bzw. obersten Regionen (Subskalen der SF-36) der Skalen angesiedelt sind. So liegen alle Teilergebnisse des BDI (mit einem Range von 0–63) im einstelligen Bereich (siehe Abbildung 2) und die Teilergebnisse der SF-36-Subskalen (mit jeweils einem Range von 0–100) übergreifend über Werten von 60, in einzelnen Skalen sogar im obersten Percentilbereich zwischen 90 und 100 (siehe Abbildungen 3–6 und 8–11). Wie aus der Literatur bekannt, zeigen beide Instrumente bei Vorgabe an „gesunden“ Stichproben je nach Polung Boden- (BDI) bzw. Deckeneffekte (SF-36), was bei Längsschnittstudien dazu führt, dass Verbesserungen nur in einem eng umgrenzten Bereich möglich sind. Eine genaue Analyse zeigt folgende Ergebnisse bezüglich der Quoten an Probanden, die zu den beiden Testzeitpunkten das jeweils bestmögliche Ergebnis in den einzelnen Skalen erzielte:

	prä	post
Körperliche Funktionsfähigkeit	60 (51,7%)	60 (51,7%)
Körperliche Rollenfunktion	107 (92,2%)	97 (83,6%)
Körperliche Schmerzen	63 (54,3%)	67 (57,8%)
Allgemeine Gesundheitswahrnehmung	4 (3,4%)	2 (1,7%)
Vitalität	2 (1,7%)	3 (2,6%)
Soziale Funktionsfähigkeit	73 (62,9%)	89 (76,7%)
Emotionale Rollenfunktion	100 (86,2%)	107 (92,2%)
Psychisches Wohlbefinden	4 (3,4%)	6 (5,2%)
BDI	25 (21,6%)	39 (33,6%)

## Zusammenfassung

Insgesamt wurden 120 Probanden in die Studie eingeschlossen, von denen 116 nach Absolvierung der vollen Studiendauer von 28 Tagen in die statistische Analyse aufgenommen werden konnten. Ein Teilnehmer beendete die Studie vorzeitig wegen des Eintretens eines kritischen Lebensereignisses, drei Teilnehmer wurden post hoc aus der Analyse ausgeschlossen, da wegen geäußerter Probleme mit dem Ein-/Ausschaltens des Gerätes nicht gewährleistet war, dass die QuintBoxen über etliche Tage hinweg die geforderte Dauer von mindestens zehn Stunden pro Tag aktiviert waren. Damit betrug die Retentionsrate der Studie über vier Wochen 96,7%, die Dropout-Rate lediglich 3,3%.

Die randomisierte Zuordnung der Probanden zu Verum- und Placebogruppe und die Parallelisierung der Kontrollgruppe kann als gelungen bezeichnet werden, da die Gruppen sich in allen erhobenen Ausgangsvariablen (Alter, Geschlechterproportion, Bildung und Body-Mass-Index) als homogen erwiesen. Auch zeigten die Ausgangswerte der Subskalen der SF-36 im Vergleich mit den Normwerten aus dem Manual des Fragebogens (Bullinger & Kirchberger, 1998; Teilstichprobe 30–40 Jahre, n=547), ausgenommen die Subskala Körperliche Rollenfunktion, keine auffälligen Abweichungen. Die durchschnittliche Tragedauer der Geräte war für Verum- und Placebogruppe nicht unterschiedlich – mit etwa elf Stunden pro Tag lagen die Werte etwa eine Stunde über der geforderten täglichen Tragedauer von zehn Stunden, woraus man auf ausreichende Grundmotivation und Compliance der Probanden schließen kann.

Die Analyse der Outcome-Variablen lässt sich zusammenfassend folgendermaßen charakterisieren: In den Fällen, in denen die Interpretation des General Linear Models for Repeated Measures aufgrund gegebener Voraussetzungen (in den drei Subskalen der SF-36 Allgemeine

Gesundheitswahrnehmung, Vitalität, Soziale Funktionsfähigkeit sowie der Physischen Summenskala) möglich war, zeigte sich in zwei Fällen ein Zeiteffekt, einmal ein Gruppeneffekt, aber nie eine Wechselwirkung von Gruppe und Zeit – Wechselwirkungen sind Ausdruck differenter Kurvencharakteristiken im Zeitvergleich und in den Abbildungen am ehesten im BDI bzw. in der SF-36-Subskala Emotionale Rollenfunktion zu erkennen).

Zeiteffekte konnten in der Detailanalyse in insgesamt acht von elf analysierten Skalen identifiziert werden. Aufgrund der oben erwähnten Problematik von Boden- und Deckeneffekten wurden zusätzlich Effektstärkenberechnungen in die Interpretation der Ergebnisse miteinbezogen. Zieht man Ergebnisse mit Effektstärken unter  $d=0,30$ , verbleiben lediglich x Zeiteffekte mit inhaltlicher Relevanz, die allesamt von der Verumgruppe erzielt wurden: Vitalität ( $d=0,38$ ), Emotionale Rollenfunktion ( $d=0,45$ ), Psychisches Wohlbefinden ( $d=0,30$ ), Psychosoziale Summenskala ( $d=0,48$ ) und die Summenskala des BDI ( $d=0,40$ ). Anzumerken ist aber auch hier, dass alle Effektstärken im Bereich von  $d=0,30$  bis  $d=0,48$ , also im Bereich moderater Effekte, angesiedelt sind.

Gruppeneffekte konnten lediglich in zwei der insgesamt elf analysierten Variablen beobachtet werden, in beiden Fällen zeigten sich die Werte der Verumgruppe höher als zumindest eine der beiden Vergleichsgruppen. Im Falle der SF-36-Subskala Allgemeine Gesundheitswahrnehmung wies die Verumgruppe aber bereits zum ersten Erhebungszeitpunkt bessere Werte als die Kontrollgruppe auf, zum zweiten Zeitpunkt tendenziell bessere Werte als die beiden anderen Gruppen. Die Effekte in dieser Skala können daher nicht kausal auf die QuintBox zurückgeführt werden, sondern auf das bessere Ausgangsniveau der Verumgruppe. Im zweiten Fall, der Subskala Vitalität, verhält sich die Sachlage jedoch anders: Basierend auf einer statistisch gesehen gleichen Ausgangslage kann zum Testzeitpunkt post in dieser Subskala eine Überlegenheit der Verumgruppe im Vergleich mit der Placebo-Gruppe attestiert werden und mit einem  $d=0,54$  als inhaltlich gut abgesichert gelten. Der Vergleich mit der Kontrollgruppe erreicht zwar knapp nicht Signifikanz, mit einem  $d=0,52$  kann aber von inhaltlich relevanten Unterschieden gesprochen werden.

Zusammenfassend lassen sich die Ergebnisse dahingehend interpretieren, dass die Anwendung der QuintBox3 bei einer gesunden Stichprobe zu Verbesserungen vornehmlich im Bereich der psychosozialen Subskalen der gesundheitsbezogenen Lebensqualität, erhoben mit der Short-Form-36, sowie zu einer Reduktion von Depressionssymptomen, erhoben mit dem BDI, führt, wobei zu erwähnen ist, dass unter den gegebenen Erhebungsbedingungen (Tragedauer 28 Tage) alle Zeiteffekte lediglich im moderaten Bereich (Cohen's  $d$  zwischen 0,30 und 0,48) liegen. Effekte in den physischen Subskalen konnten nicht nachgewiesen werden. Die erzielten Ergebnisse sind durch die bei gesunden Stichproben üblicherweise auftretenden Boden- (BDI) bzw. Deckeneffekte (SF-36) beeinträchtigt.

Die Teilnehmer der Studie zeigten sich allgemein motiviert, das Gerät wurde täglich durchschnittlich länger als empfohlen getragen, auch die geringe Dropoutrate weist auf hohe Studien-Compliance hin. Auch auf Ebene subjektiver Beobachtungen zeichnete sich ein durchaus positives Bild ab: Probanden der Verumgruppe berichteten nach den vier Wochen, in denen sie das Gerät getragen hatten, Auswirkungen auf deren Stimmung sowie auf deren Blutdruck. Eine Probandin der Verumgruppe, die vor Studienbeginn niedrigen Blutdruck aufwies, zeigte im Verlauf der vier Wochen eine Normalisierung des Blutdruckes. Andere Probanden beschrieben eine Besserung ihrer Stimmung während der Tragephase. Sie gaben an, gelassener gewesen zu sein bzw. weniger reizbar. Während der gesamten Studiendauer berichtete nur eine Studienteilnehmerin, dass die Wirkung der Quintbox3 für sie als unangenehm empfunden wurde (Persönliche Mitteilung durch Kerstin Pöll, 19.5.2009).

Auf Basis dieser Ergebnisse und Erfahrungen wird empfohlen, im Rahmen eines ähnlichen

Studiendesigns eine Vergleichsstudie durchzuführen. Anbieten würde sich etwa ein Symptombereich, der für die QuintBox3 indiziert erscheint (i.e. „Personen mit Überforderung ihres Energiesystems“), indem die Effekte der Behandlung mittels QuintBox mit denen einer Standardbehandlung (Treatment as usual; TAU) verglichen werden, u.U. mittels Crossover-Design, d.h. jede Person bekommt in randomisierter Reihenfolge beide Behandlungen..

### **Acknowledgement**

Dank ergeht an Kerstin Pöll, die im Zuge ihrer Diplomarbeit mit der Datenerhebung der Studie beauftragt war, und an Mag. Dorothea König für ihre Unterstützung bei der Erstellung der Einverständniserklärung für die Studie.

Abb. 1. Anzahl der Studienteilnehmer nach Altersgruppen gesamt und getrennt nach männlich und weiblich

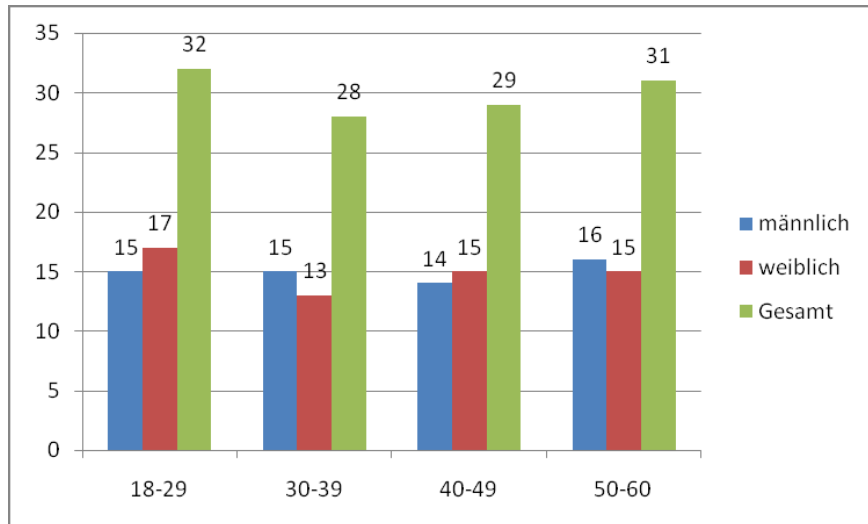


Abb. 2. Summenscore des BDI (Ausschnitt 0-10)

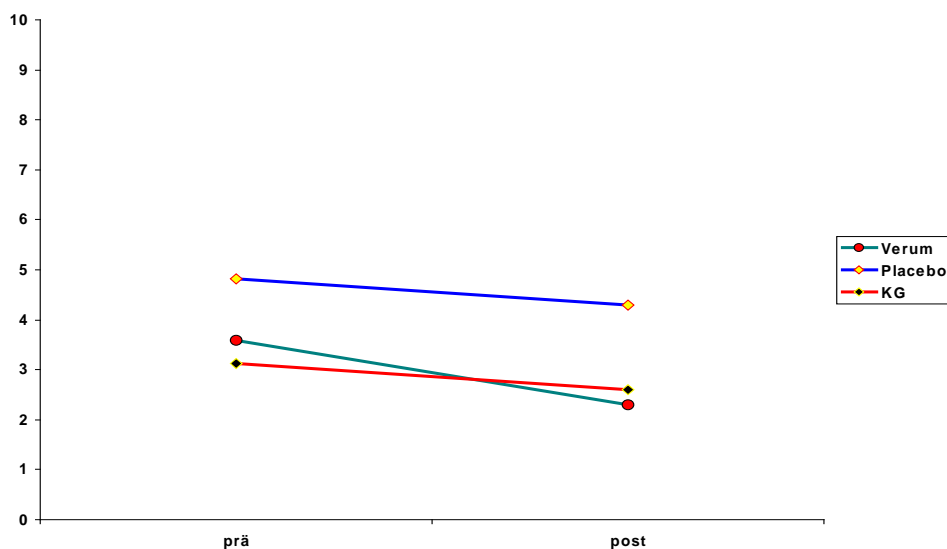


Abb. 3. Körperliche Funktionsfähigkeit (50-100)

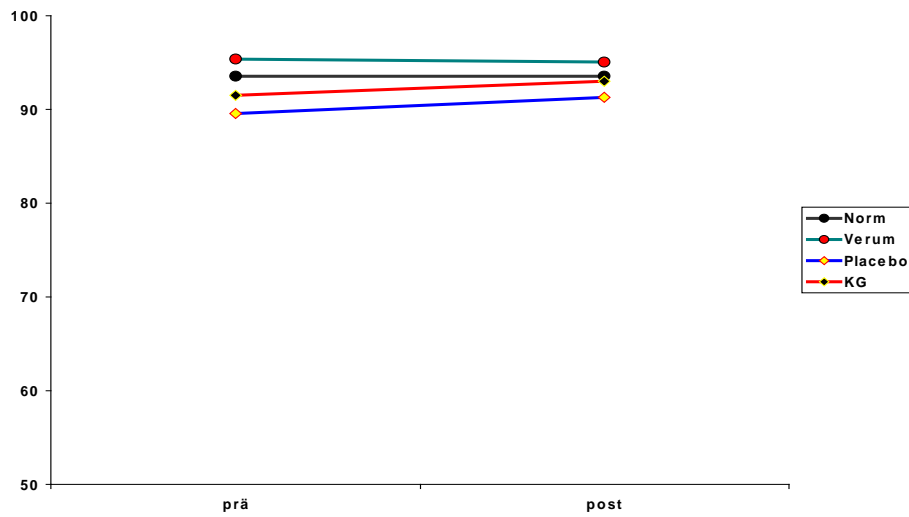


Abb. 4. Körperliche Rollenfunktion (50-100)

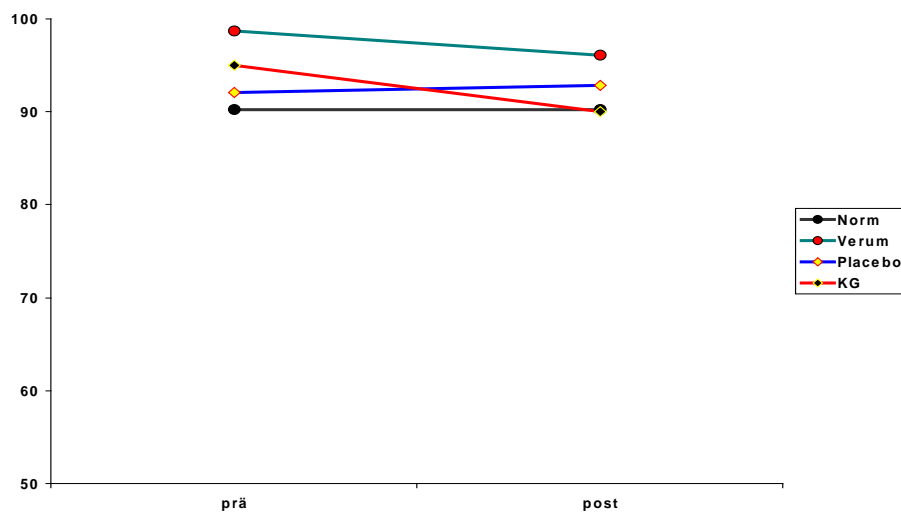


Abb. 5. Körperliche Schmerzen (50-100)

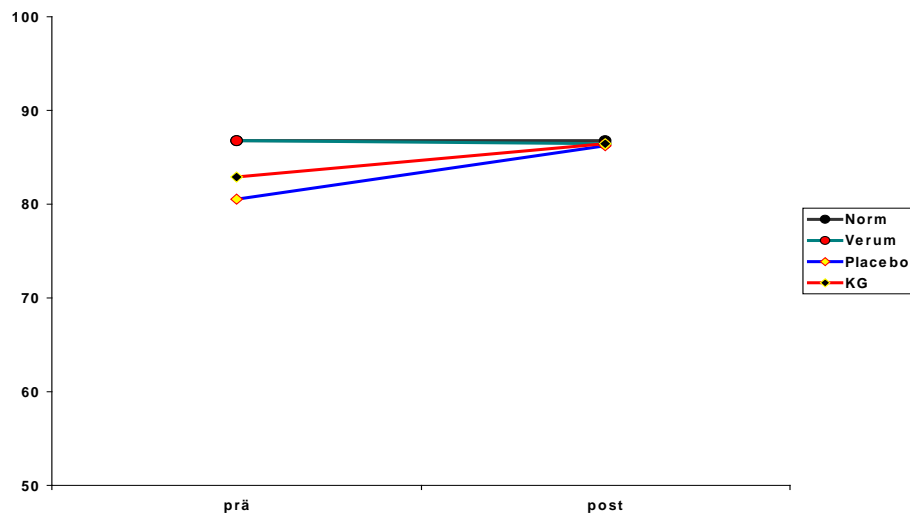


Abb. 6. Allgemeine Gesundheitswahrnehmung (50-100)

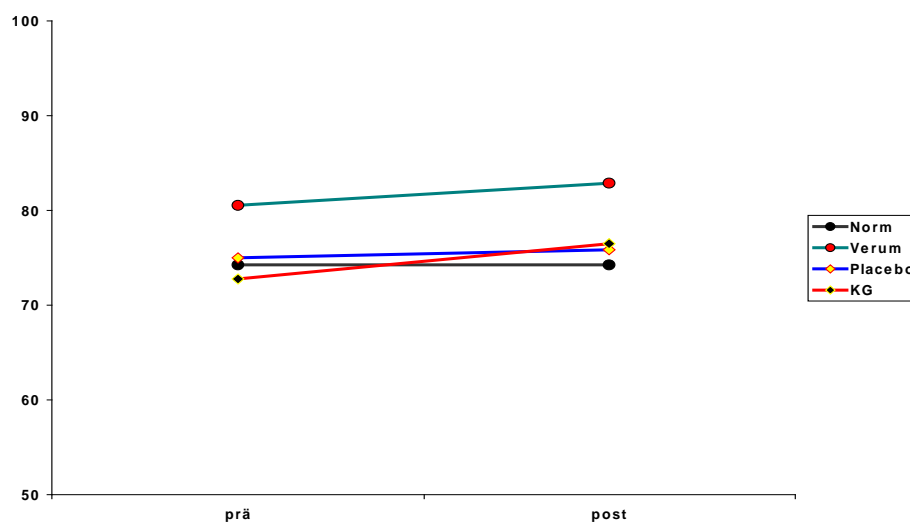


Abb. 7. Physische Summenskala (50-100)

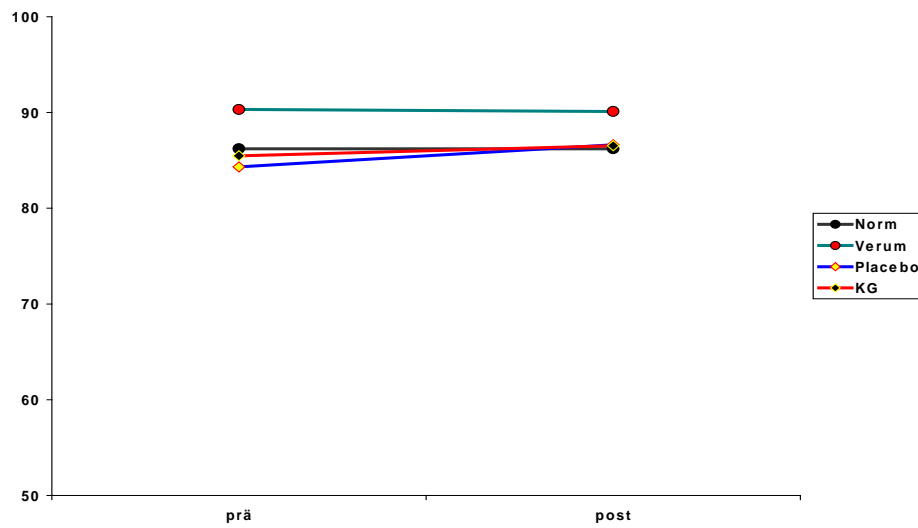


Abb. 8. Vitalität (50-100)

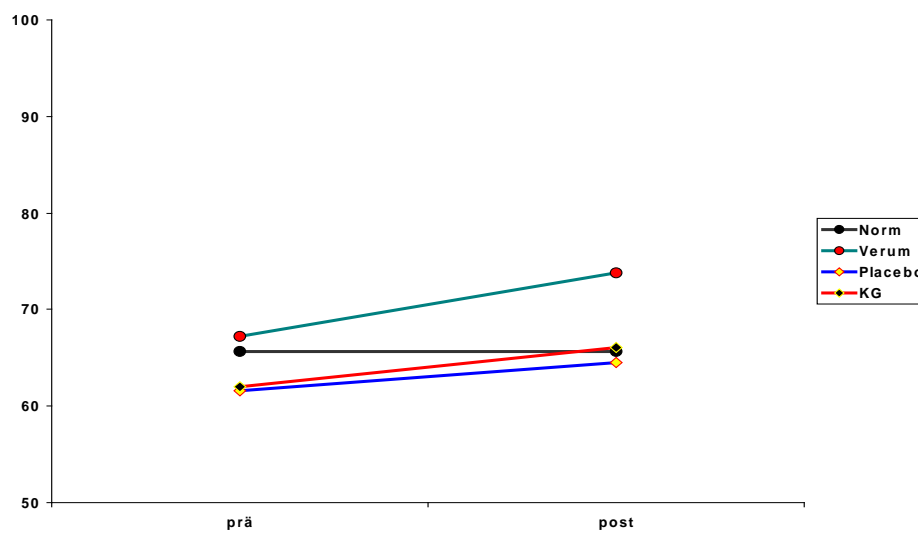




Abb.9. Soziale Funktionsfähigkeit (50-100)

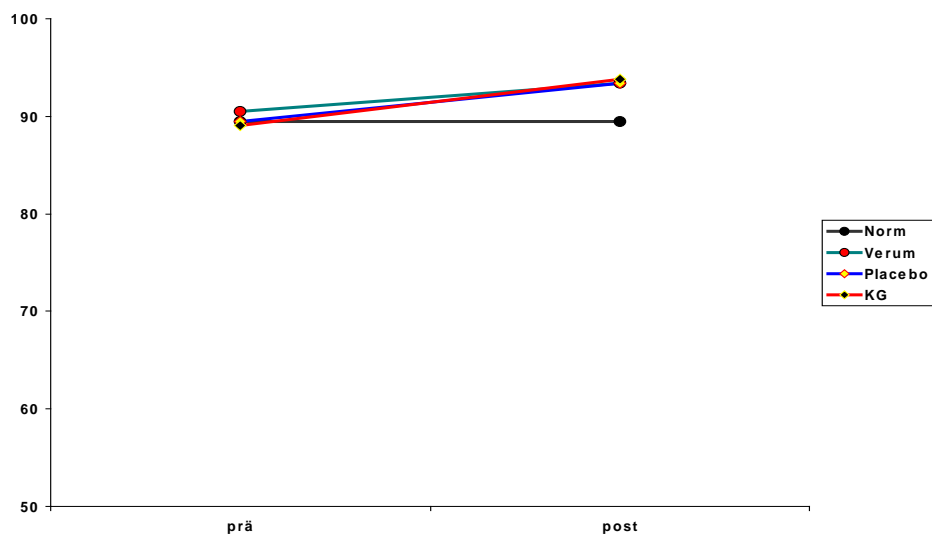


Abb. 10. Emotionale Rollenfunktion (50-100)

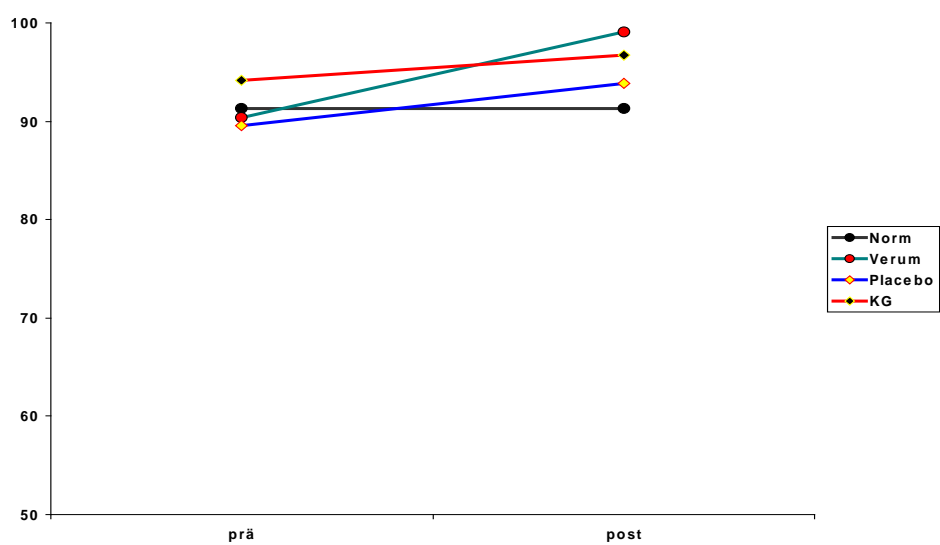


Abb. 11. Psychisches Wohlbefinden (50-100)

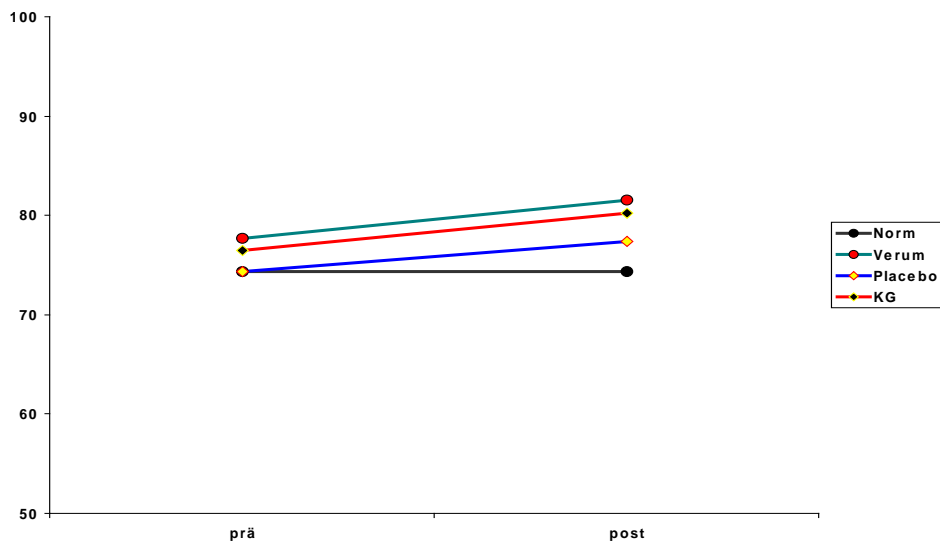


Abb. 12. Psychosoziale Summenskala (50-100)

