**metheval**report
4(3) 2002

Christof Nachtigall & Ute Suhl

# Warum kompliziert, wenn es auch einfach geht?

# Teil 1: Zur Analyse intraindividueller Veränderung

metheval
psychologische methodenlehre
und evaluationsforschung jena

# Impressum

**metheval**report
erscheint seit 1999
in unregelmäßigen Abständen
als „graue" Schriftenreihe des Lehrstuhls für
Psychologische Methodenlehre und Evaluationsforschung
am Institut für Psychologie der Friedrich-Schiller-Universität Jena

**Herausgeber:**
Prof. Dr. Rolf Steyer
Sekr.: +49 (3641) 945 230
Durchwahl: +49 (3641) 945 231
Fax: +49 (3641) 945 232

rolf.steyer@uni-jena.de

**Redaktion:**
Dipl.Psych. Friedrich Funke
sff@uni-jena.de

**Typographie:**
cand.psych. Silke Zachariae
zachariae@web.de

**Standort:**
Thüringer Universitäts- und Landesbibliothek
Lesesaal Zweigstelle Psychologie

**Internet**
http://www.uni-jena.de/svw/metheval/report/

**Bestellungen:**
Methodenlehre und Evaluationsforschung
Institut für Psychologie
Steiger 3 Haus 1
D-07743 Jena
Deutschland

# Warum kompliziert, wenn es auch einfach geht?

## Teil 1: Zur Analyse intraindividueller Veränderung

**Christof Nachtigall und Ute Suhl**

### 1. Einleitung

In diesem Beitrag beschäftigen wir uns mit der Analyse intraindividueller Veränderung. Diesem Thema kommt in der klinischen und pädagogischen Praxis große Bedeutung zu, wenn es um die Beurteilung der Veränderung von Personen im Zuge einer Intervention geht. Für den klinischen Psychologen stellt sich beispielsweise die Frage, ob und wie stark ein Patient sich im Verlauf der Therapie hinsichtlich relevanter klinischer Skalen verändert. Ein Problem stellt dabei die in aller Regel nicht perfekte Reliabilität der Messinstrumente dar. Erhebt man die Befindlichkeit der Patienten vor und nach einer Therapie, so stellt sich die Frage, ob unterschiedliche Messwerte auch eine tatsächliche Veränderung des Patienten bedeuten oder wohlmöglich nur aufgrund von Messfehlern zustande gekommen sind.

In diesem Beitrag werden wir exemplarisch zwei verschiedene Kennwerte zur Beschreibung intraindividueller Veränderung vorstellen und ihre Vor- und Nachteile diskutieren. Dabei beginnen wir mit der manifesten Differenz $Y$-$X$ zweier Messungen und vergleichen sie mit alternativen Kennwerten, die von Steyer, Hannöver, Telser & Kriebel (1997) vorgeschlagenen wurden.

### 2. Kennwerte zur Beschreibung intraindividueller Veränderung

In der klinisch-psychologischen Praxis steht man häufig vor der Situation, zu beurteilen, ob sich der Zustand eines Klienten durch eine therapeutische Intervention verändert hat[1]. Hierzu vergleicht man z. B. die Befindlichkeit $X$ dieser Person, erfasst durch einen geeigneten Fragebogen, zu Beginn einer Behandlung mit dem Messwert $Y$ derselben Person nach der Durchführung der Behandlung. In der Regel sind die Messwerte, die man mit einem psychologischen Messinstrument (z. B. klinischen Tests) erhebt, messfehlerbehaftet. Gemäß der klassischen Testtheorie (z. B. Lienert, 1961) setzt sich ein Messwert $X$ bzw. $Y$ additiv zusammen aus einem wahren Wert (true score) und einem Messfehleranteil:

$$\begin{aligned} X &= \tau_X + \varepsilon_X \\ Y &= \tau_Y + \varepsilon_Y \end{aligned} \qquad (1)$$

Dabei werden in der klassischen Testtheorie Annahmen über die Fehler ε gemacht:

i. der Erwartungswert der Fehler ist Null,

ii. die Fehler sind unkorreliert,

iii. die Fehlervarianzen sind gleich.

Die Gültigkeit von i, ii und iii wird in den folgenden Überlegungen immer vorausgesetzt.

[1] Für einen Überblick über Veränderungsmessung in der klinischen Psychologie vgl. Lambert und Hill (1994).

 

Manchmal wird der wahre Wert auch als Erwartungswert der intraindividuellen Verteilung einer Person definiert (z. B. Steyer und Eid, 2002), in diesem Fall ergibt sich i. als Folgerung aus dieser Definition und muss automatisch gelten[2].

Das Ziel des klinischen Psychologen, der beurteilen soll, ob und wie stark sich Patienten verändert haben, besteht in der Ermittlung der Differenz $\tau_Y$-$\tau_X$. Ein naheliegender Ansatz zur Beschreibung der intraindividuellen Veränderung besteht in der Betrachtung der Differenz

$$D = Y - X \tag{2}$$

Diese Messwertedifferenz als Kennwert für intraindividuelle Veränderung hat einige offensichtliche Vorteile: Sie ist einfach zu berechnen, was vor allem von Praktikern begrüßt wird. Darüber hinaus ist sie erwartungstreu für jede Person, d. h. der Erwartungswert von $D$ bei einem Patienten ist die gesuchte wahre Differenz $\tau_Y$-$\tau_X$ dieses Patienten. Es gilt in der Terminologie von Steyer und Eid (2002):

$$\begin{aligned} E(D|P_U{=}\mathrm{u}) &= E(Y\text{-}X|P_U{=}\mathrm{u}) = E(Y|P_U{=}\mathrm{u}) - E(X|P_U{=}\mathrm{u}) \\ &= \tau_Y(\mathrm{u}) - \tau_X(\mathrm{u}). \end{aligned} \tag{3}$$

$P_U$=u gibt in dieser Formel die untersuchte Person (Unit) an.

Ist ein Messinstrument nicht allen Praktikern geläufig, so fällt es schwer, eine Differenz von z. B. $D$=5 hinsichtlich ihrer Größe zu interpretieren: Sind 5 Punkte viel oder wenig? Für diesen Fall empfehlen Grawe und Braun (1994) eine Standardisierung anhand der Standardabweichung des Messinstrumentes. Dieses standardisierte Differenzenmaß $D_{st}$ berechnet sich durch

$$D_{st} := \frac{Y - X}{Std(X)} \tag{4}$$

Ein solches Vorgehen ist bei der Beurteilung von Effektstärken üblich (vgl. z.B. Cohen, 1988), um abschätzen zu können, ob die beobachtete Veränderung groß oder klein ist gemessen an der Grundvariabilität des Merkmals.

Auf der Grundlage dieser Größe kann man allerdings nicht sicher beurteilen, ob sich die True Scores der beiden Messwerte voneinander unterscheiden: Da $Y$ und $X$ beide messfehlerbehaftet sind, kann auch die Differenz $D$ bzw. $D_{st}$ allein aufgrund unterschiedlicher Messfehler von Null verschieden sein, auch wenn die wahren Werte bei beiden Messungen gleich sind. Um zu prüfen, ob $\tau_Y = \tau_X$ gilt, empfiehlt sich daher ein verändertes Vorgehen: Man setzt die beobachtete Differenz $Y$-$X$ ins Verhältnis zur Standardabweichung dieser Differenz und erhält damit die als *Reliable Change Index (RCI)* (Jacobson and Truax, 1991) bekanntgewordene Prüfgröße:

$$RCI := \frac{Y - X}{Std(Y - X)} = \frac{Y - X}{Std(X) \cdot \sqrt{2(1 - \mathrm{Rel}(X))}} \tag{5}$$

Sind $Y$ und $X$ normalverteilt mit gleichen Messfehlervarianzen und unkorrelierten Messfehlern und gilt $\tau_Y = \tau_X$, dann ist $RCI$ standardnormalverteilt (vgl. z.B. Nachtigall und Wirtz, 2002, S. 150). Die Nullhypothese $\tau_Y = \tau_X$ läßt sich also über die Standardnormalverteilung prüfen: Ist $|RCI| \geq 1.96$, dann wird die Nullhypothese gleicher wahrer Werte mit einer

[2] Dieses Vorgehen hat den Nachteil, dass der wahre Wert ungeachtet inhaltlich-psychologischer Überlegung definiert und die Möglichkeit ausgeschlossen wird, dass Messfehler unsymmetrisch um den wahren Wert verteilt sein können. Dies ist beispielsweise häufig bei der Messung von Reaktionszeiten der Fall, für die man meist unsymmetrische, links-steile Verteilungen beobachten kann.

Irrtumswahrscheinlichkeit von $\alpha$=0.05 verworfen; andernfalls wird sie beibehalten. Auf dieser Testidee basieren auch die in vielen Handanweisungen zu psychologischen Tests angegebenen kritischen Differenzen: Eine solche kritische Differenz gibt an, wie groß eine beobachtete Messwertdifferenz mindestens sein muss, damit man die Nullhypothese gleicher wahrer Werte zu einem vorgegebenen $\alpha$-Niveau verwerfen kann. Sie ergibt sich durch Umstellen der Ungleichung $|RCI| \geq 1.96$:

$$\begin{aligned} &\left|RCI\right| \geq 1.96 \\ &\Leftrightarrow \frac{Y-X}{Std(X)\cdot\sqrt{2(1-\mathrm{Rel}(X))}} \geq 1.96 \\ &\Leftrightarrow (Y-X) \geq 1.96\cdot Std(X)\cdot\sqrt{2(1-\mathrm{Rel}(X))} \end{aligned} \tag{6}$$

Die rechte Seite der letzten Ungleichung gibt gerade die kritische Differenz an.

Steyer et al. (1997) kritisieren die Verwendung der manifesten Messwertdifferenz $D$ (bzw. $D_{st}$) als Maß zur Beschreibung der intraindividuellen Veränderung, da bei dieser Größe nicht berücksichtigt werde, dass $Y$ und $X$ beide messfehlerbehaftet sind: „Eine Post-Prä-Differenz, die ausschließlich auf Messfehlern beruht, (kann) genauso groß aussehen (...), wie eine Post-Prä-Differenz, die auf perfekt reliablen Messungen beruht" (S. 292). Sie schlagen stattdessen vor, den Posttestwert $Y$ mit dem unter der Nullhypothese „keine wahre Veränderung" zu erwartenden Posttestwert zu vergleichen. Dieser wird mittels einer Regression ausgehend vom Prätestwert $X$ vorhergesagt und in ihrem Artikel mit $E_0(Y|X)$ bezeichnet. Die Nullhypothese „keine wahre Veränderung" bedeutet, dass $X$ und $Y$ die gleichen wahren Werte $\tau_X=\tau_Y$ zugrunde liegen. Außerdem wird neben den oben genannten Annahmen der klassischen Testtheorie (i, ii und iii) die Linearität der Regression $E_0(Y|X)$ vorausgesetzt. Ausgehend von diesen Annahmen gelangen sie zu den folgenden Veränderungskenngrößen:

$$V_{\text{deskript}} := \frac{Y-E_0(Y\,|\,X)}{Std(X)} = \frac{\left[Y-E(X)\right]-\mathrm{Rel}(X)\left[X-E(X)\right]}{Std(X)} \tag{7}$$

(dekriptiver Veränderungskennwert)

$$V_{\text{infer}} := \frac{Y-E_0(Y\,|\,X)}{Std(Y-E_0(Y\,|\,X))} = \frac{\left[Y-E(X)\right]-\mathrm{Rel}(X)\left[X-E(X)\right]}{Std(X)\sqrt{1-\mathrm{Rel}(X)^2}} \tag{8}$$

(inferentieller Kennwert)

Beide von Steyer et al. (1997) vorgeschlagenen Kennwerte sind vom Grundprinzip her ähnlich aufgebaut wie die üblicherweise verwendeten Kennwerte $D_{st}$ und $RCI$: Zur Beschreibung der Veränderung mit Hilfe von $V_{\text{deskript}}$ wird die jeweilige Differenz ins Verhältnis gesetzt zur Streuung des Prätests $X$, zur Prüfung, ob eine Veränderung des wahren Wertes vorliegt (inferentielle Veränderungskenngröße), wird diese Differenz zu ihrer eigenen Standardabweichung ins Verhältnis gesetzt. Die Kennwerte unterscheiden sich letztlich nur darin, dass bei $D_{st}$ bzw. $RCI$ der Messwert $X$ selbst als Referenz herangezogen wird, während bei den von Steyer et al. (1997) vorgeschlagenen Kennwerten $V_{\text{deskript}}$ und $V_{\text{infer}}$ ein korrigierter Prätestmesswert als Vergleichspunkt eingeht. Durch die Korrektur werden allerdings die Formeln deutlich komplizierter und recht unübersichtlich. Um die Eigenschaften dieser Kennwerte besser diskutieren zu können, werden wir im folgenden ohne Einschränkung der Allgemeinheit zur Betrachtung z-standardisierter Messwert übergehen (also $E(X)=E(Y)=0$ und $Std(X)$=$Std(Y)$=1). Die Kennwerte nach Steyer et al. vereinfachen sich dann wie folgt:

$$V_{\text{deskript}} := Y-\mathrm{Rel}(X)\cdot X \tag{9}$$

 

$$V_{\text{infer}} := \frac{Y - \text{Rel}(X) \cdot X}{\sqrt{1 - \text{Rel}(X)^2}} \tag{10}$$

Bei perfekt reliablen Messungen stimmen $D_{st}$ und $V_{\text{deskript}}$ überein. In diesem Fall erübrigt sich auch ein inferentieller Kennwert, da Unterschiede in den Messwerten automatisch unterschiedliche wahre Werte bedeuten. Wir gehen im Folgenden von der in der Praxis vorherrschenden Situation nicht perfekt reliabler Messinstrumente aus. $X$ und $Y$ sollen dabei zwei Messungen mit dem gleichen Messinstrument sein, insbesondere gelte $\text{Rel}(X) = \text{Rel}(Y)$. Eine solche Situation findet sich z.B. in der klinischen Psychologie bei Prä-Post-Messungen.

## 3. Zu den Eigenschaften der von Steyer et al. vorgeschlagenen Veränderungskennwerte

Was bewirkt die von Steyer et al. vorgeschlagene Korrektur im Vergleich zur manifesten Differenz? Die Antwort auf diese Frage ist nicht einfach zu geben, da das Ausmaß der Korrektur zum einen von der Zuverlässigkeit (Reliabilität) des Messinstrumentes abhängt, zum anderen aber auch von der Differenz $X\text{-}E(X)$, also davon, wie „extrem" die Messwerte vom Mittel abweichen. Wir konzentrieren uns in diesem Beitrag auf eine Untersuchung von $V_{\text{deskript}}$, da dieser Kennwert als Grundlage für eine inhaltliche Interpretation heranzuziehen ist. Der inferentielle Index $V_{\text{infer}}$ ist jedoch bis auf die Normierung mit $V_{\text{deskript}}$ identisch, so dass die folgenden Ergebnisse auch für $V_{\text{infer}}$ gelten[3].

Generell gilt: Im Vergleich zu den Kennwerten $D$ bzw. $D_{st}$ wird bei $V_{\text{deskript}}$ das Ergebnis der Messung $X$ in Richtung auf den Erwartungswert $E(X)$ korrigiert. Diese Korrektur ist umso stärker, je geringer die Reliabilität des Messinstrumentes ist oder je extremer $X$ ausfällt. Entsprechend vergrößert bzw. verkleinert der deskriptive Veränderungskennwert $V_{\text{deskript}}$ die beobachtete manifeste Differenz.

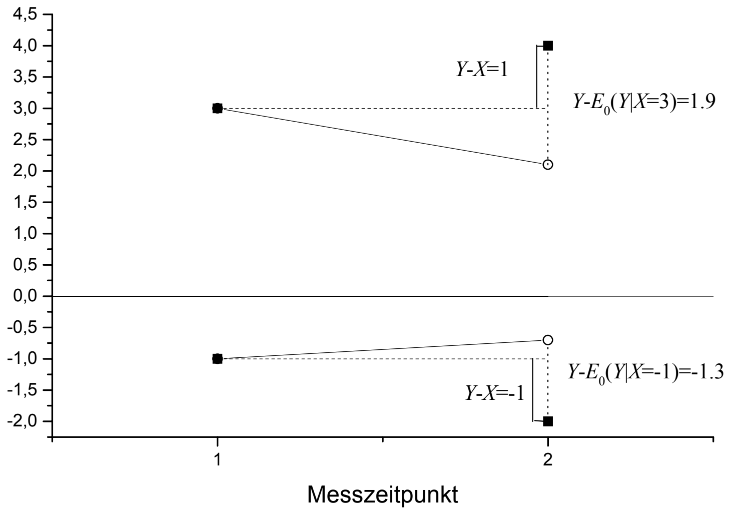


Abbildung 1: Veranschaulichung des deskriptiven Veränderungskennwertes $V_{\text{deskript}}$ (Erläuterung siehe Text)

[3] Das Anwendungsfeld von $V_{\text{infer}}$ liegt jedoch vornehmlich in der statistischen Testung von Veränderung, der Frage, welche statistischen Eigenschaften $V_{\text{infer}}$ besitzt, wird im zweiten Teil dieses Beitrages nachgegangen (Suhl und Nachtigall, 2002).

Die Abbildungen 1 und 2 veranschaulichen den zugrunde liegenden Mechanismus am Beispiel einer Prä-Post-Messung. Dargestellt sind zum einen verschiedene manifeste Differenzen, zum anderen die Werte, welche sich für den Veränderungskennwert $V_{\text{deskript}}$ anhand der vorgegebenen Zahlen errechnen lassen. Die ausgefüllten Kästchen geben die beobachteten Messwerte $X$ (Messzeitpunkt 1) und $Y$ (Messzeitpunkt 2) wieder.

Der unter der Nullhypothese „keine wahre Veränderung“ ausgehend vom Messwert $X$ zu erwartende Messwert ist durch einen kleinen Kreis (Messzeitpunkt 2) dargestellt. Wie man in beiden Graphiken erkennen kann, liegt der erwartete Posttestwert immer näher am Erwartungswert[4] als der zugehörige Prätestwert (Messwert zum Messzeitpunkt 1). Um dies zu verdeutlichen, ist in die Graphiken zum einen der Erwartungswert $E(X)=0$ jeweils durch eine durchgezogene horizontale Linie eingetragen. Zusätzlich sind aber auch die zueinander gehörenden Prätestwerte und die erwarteten Posttestwerte durch durchgezogene Linien miteinander verbunden. In den Abbildungen kann man erkennen, dass die Korrektur umso stärker ausfällt, je weiter der Prätestwert vom Mittelwert 0 abweicht. Dies sieht man an den unterschiedlichen Steigungen der Linien, die Prätestwert und erwarteten Posttestwert miteinander verbinden: In den oberen Teilen der Graphiken sind die Zahlenbeispiele so gewählt, dass die Prätestwerte stärker von Null abweichen als für die Zahlenbeispiele im unteren Teil der Graphiken. Die durchgezogenen Linien, die den Unterschied zum erwarteten Posttestwert veranschaulichen, weisen eine größere Steigung auf, als die entsprechenden Linien in der unteren Hälfte der Abbildungen. Für die Berechnung der erwarteten Posttestwerte wurde eine Reliabilität von $\text{Rel}(X)=0.7$ angenommen.

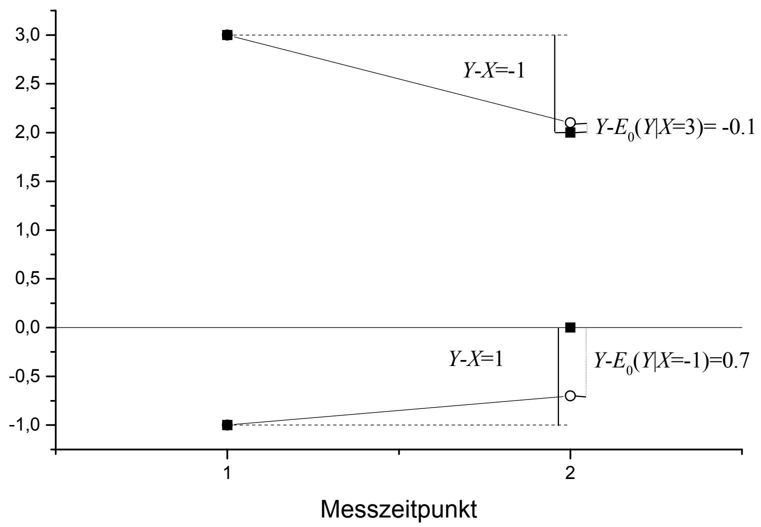


Abbildung 2: Veranschaulichung des deskriptiven Veränderungskennwertes $V_{\text{deskript}}$ nach Steyer et al. (Erläuterung siehe Text)

Die beiden Abbildungen unterscheiden sich im Bezug auf die ausgewählten Werte von $Y$: In Abbildung 1 weichen sie stärker von Null ab als die zugehörigen $X$-Werte; in diesen Fällen vergrößert der deskriptive Veränderungskennwert die manifeste Differenz. In Abbildung 2 liegen sie dagegen dichter am Erwartungswert als die zugehörigen $X$-Werte; in diesen Fällen verkleinert der deskriptive Veränderungsindex die manifeste Differenz. Die Zahlenbeispiele,

[4] Da wir zur Vereinfachung der Darstellung von standardisierten Variablen ausgehen, gilt $E(X)=0$.

die den beiden Graphiken zugrunde liegen, sind noch einmal in der nachfolgenden Tabelle 1 zusammengefasst.

Tabelle 1: Zahlenbeispiele zur Veranschaulichung der Eigenschaften des Veränderungskennwertes $V_{\text{deskript}}$. Für die Berechnung der erwarteten Posttestwerte $E_0(Y|X)$ wurde eine Reliabilität von Rel($X$)=0.7 zugrunde gelegt.

| | $Y$ | $X$ | $Y–X$ | $Y–E_0(Y\|X)$ | *Kommentar* |
|---|---|---|---|---|---|
| *Fall 1* | 4<br>-2 | 3<br>-1 | 1<br>-1 | 1.9<br>-1.3 | *Die manifeste Differenz wird vergrößert.* |
| *Fall 2* | 2<br>0 | 3<br>-1 | -1<br>1 | -0.1<br>0.7 | *Die manifeste Differenz wird verkleinert.* |
| *Fall 3* | 2.5 | 3 | -0.5 | 0.4 | *Das Vorzeichen der manifesten Differenz wird umgekehrt.* |
| *Fall 4* | 1 | 1 | 0 | 0.3 | *Im Falle identischer Wert von X und Y meldet* $V_{\text{deskript}}$ *eine Veränderung* |

Die Zahlenwerte aus der Zeile „Fall 1“ entsprechen den in Abbildung 1 dargestellten Gegebenheiten, die aus der Zeile „Fall 2“ den in Abbildung 2 veranschaulichten Verhältnissen. In der Zeile „Fall 3“ ist zusätzlich ein Zahlenbeispiel aufgeführt, bei dem die Verwendung von $V_{\text{deskript}}$ sogar zu einer Vorzeichenumkehrung führt: Obwohl $Y$ näher am Erwartungswert $E(X)$ liegt als $X$, signalisiert der Veränderungskennwert nach Steyer et al. (1997) dennoch eine Veränderung in die andere Richtung, weg vom Erwartungswert. „Fall 4“ zeigt, dass im Fall identischer Werte von $X$ und $Y$ der Kennwert trotzdem eine Veränderung anzeigt.

Wie die Zahlenbeispiele belegen, lässt sich die konkrete Auswirkung der in $V_{\text{deskript}}$ eingebauten Korrektur nicht in Form einer einfachen „Gesetzmäßigkeit“ zusammenfassen. Vielmehr hängt sie von der konkreten Messwertkonstellation ab. Und zwar ist die Korrektur umso stärker wirksam,

- je geringer die Reliabilität des verwendeten Messinstrumentes ist, und
- je stärker die Messung $X$ vom Erwartungswert abweicht.

Ob die Korrektur zu einer Vergrößerung, einer Verkleinerung oder sogar zur einer Umkehrung des Vorzeichens der manifesten Differenz führt, lässt sich dagegen nur schwer abschätzen, da es von den konkreten Zahlenwerten abhängt.

## 4. Vergleich der verschiedenen Veränderungskennwerte

Die Analyse zeigt, dass das Verhalten von $V_{\text{deskript}}$ recht kompliziert ist und zu intuitiv schwer nachvollziehbaren Resultaten führt. Uns ist es – trotz intensiven Nachdenkens – nicht gelungen, einen einigermaßen plausiblen Anwendungsfall zu konstruieren, in dem es *aus inhaltlichen Gründen* sinnvoll wäre, diesen Kennwert zu verwenden und damit die beobachtbare Differenz in Abhängigkeit von der Messwertkonstellation mal zu vergrößern, mal zu verkleinern oder auch im Vorzeichen umzukehren. Die Argumente, die Steyer et al. (1997) für ihren Index ins Feld führen, sind allerdings auch nicht *inhaltlicher*, sondern *formaler* Natur: Ihre Argumentation suggeriert – ohne es jedoch explizit so zu formulieren -, dass a) ihr deskriptiver Veränderungskennwert im Unterschied zu $D$ (bzw. $D_{st}$) das Messfehlerproblem behebe, indem b) die Regression zur Mitte berücksichtigt werde (ebda., S. 292 u. 293). Betrachten wir diese beiden Argumente genauer:

 

Es ist zutreffend, dass bei den Kennwerten $D$ (bzw. $D_{st}$) eine angezeigte Veränderung sowohl eine wahre Veränderung ($\tau_Y$- $\tau_X \neq 0$) als auch eine allein messfehlerbedingte Veränderung widerspiegeln kann. Allerdings trifft dies genauso auf $V_{\text{deskript}}$ zu. Auch hier kann eine angezeigte Veränderung auf einer tatsächlich vorliegenden wahren Veränderung oder allein auf Messfehlern beruhen. Beispielsweise könnten in Fall 3 aus Tabelle 1 die wahren Werte $\tau_Y = 2.4$, $\tau_X = 3.2$ lauten. In diesem Fall würden sowohl $D$ (bzw. $D_{st}$) als auch $V_{\text{deskript}}$ eine Veränderung anzeigen, bemerkenswerter Weise $V_{\text{deskript}}$ mit dem falschen Vorzeichen. Aber auch die wahren Werte $\tau_Y = 2.75$, $\tau_X = 2.75$ wären möglich, in diesem Falle zeigen *beide* Kennwerte fälschlicherweise Veränderung an. Festzuhalten ist also, dass keiner der Kennwerte das Messfehlerproblem wirklich lösen kann. Solange einzelne Messwerte messfehlerbehaftet sind und die Größe der einzelnen Messfehler unbekannt ist, besteht für keinen Kennwert die Möglichkeit, auf Einzelfallebene die wahren Differenzen zu ermitteln.

Aber liefert $V_{\text{deskript}}$ zumindest im Durchschnitt den richtigen Wert? Wir hatten in (3) gezeigt, dass die Differenz $D$ für jeden Patienten u erwartungstreu hinsichtlich der Differenz der wahren Werte ist. Für $V_{\text{deskript}}$ gilt hingegen:

$$\begin{aligned} E(V_{\text{deskript}} \mid P_U=u) &= E(Y\text{-Rel}(X)X \mid P_U=\text{u}) \\ &= E(Y \mid P_U=\text{u}) - \text{Rel}(X)E(X \mid P_U=\text{u}) \\ &= \tau_Y(\text{u}) - \text{Rel}(X)\,\tau_X(\text{u}) \end{aligned} \tag{11}$$

Demzufolge ist $V_{\text{deskript}}$ bei der Kennzeichnung einzelner Patienten *nicht* erwartungstreu für die Differenz der wahren Werte - ein erheblicher Nachteil für die Evaluation intraindividueller Veränderung, bei der gerade die Betrachtung von Einzelfällen von zentralem Interesse ist.

Es stellt sich natürlich die Frage, was die Aufgabe eines deskriptiven Kennwertes für intraindividuelle Veränderung sein soll. Für $V_{\text{deskript}}$ läßt sich sagen, dass er das Messfehlerproblem nicht löst (was auch kein anderer deskriptiver Kennwert kann) und dass er für einzelne Patienten im Mittel nicht richtig liegt. Welche Vorteile könnte $V_{\text{deskript}}$ haben? Betrachten wir das Argument von Steyer et al., wonach $V_{\text{deskript}}$ den Regressionseffekt berücksichtige. Vereinfacht gesprochen handelt es sich beim Regressionseffekt, auch Regression zur Mitte genannt, um das Phänomen, dass Messwerte bei einer weiteren Messung tendenziell näher am Mittelwert liegen. Dies klingt recht mystisch. Man kann und sollte sich den zugrundeliegende Mechanismus verständlich machen, da der Regressionseffekt ein häufig diskutiertes (und unserer Ansicht nach eher selten verstandenes) Problem darstellt. Eine ausführliche Erklärung findet sich z. B. in Nachtigall und Suhl (2002).

Regression zur Mitte zeigt sich folgendermaßen: Gibt es keine wahre Veränderung und liegt ein Messwertes $X$=x vor, dann ist für eine weitere Messung $Y$ der Wert

$$E_0(Y \mid X=\text{x}) = \text{Rel}(X)\text{x} \tag{12}$$

zu erwarten. Für die Herleitung von (12) berücksichtigen wir die Tatsache, dass wir z-standardisierte Variablen betrachten und die Regression zwischen $Y$ und $X$ als linear vorausgesetzt wird[5]. Dann gilt:

$$E_0(Y \mid X) = \text{Kor}(Y,X)X \tag{13}$$

(siehe z.B. Wirtz und Nachtigall, 2002, S. 107). Der Index $_0$ soll dabei deutlich machen, dass es sich um die Regression unter der Annahme gleicher wahrer Werte (also keine wahre Veränderung) handelt. Mit Gültigkeit der Annahmen i, ii und iii gilt wiederum

[5] Diese Annahme ist z.B. dann erfüllt, wenn $X$ und $Y$ gemeinsam normalverteilt sind.

$$\text{Kor}(Y,X) = \text{Rel}(X) = \text{Rel}(Y) \tag{14}$$

(siehe z.B. Steyer und Eid, 2002, Kap. 10). Durch Einsetzen von (14) in (13) folgt (12).

Dieses Argument führt Steyer et al. (1997) auch zu ihren Formeln für $V_{\text{deskript}}$ und $V_{\text{infer}}$. Der Messwert $Y$ wird bei diesen Kennwerten statt mit $X$ mit $\text{Rel}(X)X$ verglichen. Allerdings wird bei diesem Vorgehen nicht berücksichtigt, dass auch bei Vorliegen des Messwertes $Y$=y in analoger Weise hinsichtlich der Messung $X$ der Wert $\text{Rel}(Y)$y zu erwarten ist. Anders ausgedrückt: Regression zur Mitte ist kein „gerichtetes“ Phänomen, welches z. B. nur von der Messung $X$ auf die Messung $Y$ zu berücksichtigen wäre. In Nachtigall und Suhl (2002) wird dieser Punkt anhand der klassischen Fragestellung der Körpergröße von Eltern und Kindern demonstriert. Bei sehr großen Eltern liegt die mittlere Körpergröße der Kinder näher am Gesamtmittelwert als die der Eltern. Doch auch umgekehrt liegt die mittlere Größe der Eltern von besonders großen Kindern näher am Gesamtmittel. In diesem Sinne können $V_{\text{deskript}}$ und $V_{\text{infer}}$ als ‚halbe Bereinigungen‘ der Messwertedifferenz $Y$-$X$ von der Regression zur Mitte aufgefasst werden. Lediglich bei einer der beiden Variablen wird der Regressionseffekt berücksichtigt. Gerade diese halbe Bereinigung führt zu der unliebsamen Eigenschaft, dass $V_{\text{deskript}}$ nicht erwartungstreu ist.

*Konsequente Bereinigung*

Es liegt nahe, einen Kennwert auf der Basis konsequenter Bereinigung *beider* Messungen bzgl. des Regressionseffektes zu entwickeln. Für den Kennwert verwenden wir, wie bei $V_{\text{deskript}}$ geschehen, statt $X$ die korrigierte Größe $X_k := \text{Rel}(X)\, X$. Konsequenter Weise ersetzen wir aber auch $Y$ durch die korrigierte Größe $Y_k := \text{Rel}(Y)Y = \text{Rel}(X)Y$. Zur Erinnerung: Es war gleiche Reliabilität von $X$ und $Y$ vorausgesetzt. Wir erhalten den ‚konsequent bereinigten Kennwert‘

$$\begin{aligned} V_{\text{deskript k}} &:= \frac{Y_k - X_k}{Std(X_k)} = \frac{\text{Rel}(X)(Y-X)}{\text{Rel}(X)Std(X)} \\ &= \frac{Y-X}{Std(X)} = D_{\text{st}} \end{aligned} \tag{15}$$

Demnach ist der konsequent bereinigte deskriptive Veränderungskennwert identisch mit dem von Grawe und Braun (1994) vorgeschlagenen Kennwert $D_{st}$. In $D_{st}$ ist der Regressionseffekt also konsequent berücksichtigt, während in $V_{\text{deskript}}$ nur eine halbe ‚Bereinigung‘ stattfindet.

Wenden wir diese Idee der konsequenten Bereinigung auf den inferenziellen Kennwert $V_{\text{infer}}$ an, so erhalten wir

$$\begin{aligned} V_{\text{infer b}} &= \frac{Y_b - X_b}{Std(Y_b - X_b)} = \frac{\text{Rel}(X)(Y-X)}{\text{Rel}(X)Std(X)\sqrt{2(1-\text{Rel}(X)}} \\ &= \frac{Y-X}{Std(X)\sqrt{2(1-\text{Rel}(X)}} = RCI \end{aligned} \tag{16}$$

Bei konsequenter Bereinigung erhalten wir als inferenziellen Kennwert wieder den wohlbekannten *RCI* bzw. bei Umstellung der Formel die kritische Differenz (vgl. (6)). Die Regression zur Mitte ist also auch in diesem gängigen Kennwert bereits berücksichtigt, auch wenn es an den Formeln nicht direkt zu sehen ist. $V_{\text{deskript}}$ und $V_{\text{infer}}$ stellen hingegen Kennwerte dar, bei denen der Regressionseffekt nur einseitig berücksichtigt wird, was zu asymmetrischen und schwer interpretierbaren Eigenschaften dieser Kennwerte führt.

## 5. Diskussion

In diesem Beitrag werden Kennwerte der intraindividuellen Veränderung einander gegenüber gestellt. Dabei handelt es sich einerseits um die einfache (standardisierte) Messwertedifferenz $Y$-$X$ bzw. den *RCI*, andererseits um die von Steyer et al. (1997) vorgeschlagenen komplexeren Kennwerte $V_{\text{deskript}}$ und $V_{\text{infer}}$. Zunächst wurden die Resultate einer Verwendung von $V_{\text{deskript}}$ an Beispielen illustriert. Dabei zeigt dieser Kennwert überraschende Eigenschaften, welche inhaltlich äußerst schwer zu interpretieren sein dürften. Als besonders problematisch erweist sich seine mangelnde Erwartungstreue beim Einsatz auf individueller Ebene. Dies beruht darauf, dass bei den neuen Kennwerten die Variable $Y$ nicht mit $X$ sondern mit einer vom Regressionseffekt ‚bereinigten' Variable $X_k$ verglichen wird. Allerdings wird nicht berücksichtigt, dass der Regressionseffekt ein symmetrisches Phänomen ist. Wendet man diese Idee der ‚Bereinigung vom Regressionseffekt' jedoch konsequent auf *X und Y* an, dann erhält man gerade die wohlbekannten, einfach zu berechnenden sowie statistisch vorteilhaften (da erwartungstreuen) ‚alten Kennwerte' (Standardisierte Differenz bzw. RCI).

Insgesamt fällt der hier vorgenommene theoretische Vergleich klar zu Gunsten der alten Kennwerte aus. Im zweiten Teil dieses Beitrages (Suhl und Nachtigall, 2002) werden darüber hinaus die teststatistischen Eigenschaften von $V_{\text{infer}}$ unter die Lupe genommen und mit denen des *RCI* verglichen. Auch bei diesem Vergleich zeigen sich deutliche Vorteile des *RCI.* Beide Teile des Beitrags führen zu dem Schluss, dass $V_{\text{deskript}}$ und $V_{\text{infer}}$ eindeutig die schlechtere Alternative im Vergleich zu $D$ (bzw. $D_{st}$) und *RCI* darstellen und von ihrer Verwendung in der klinischen Praxis abzuraten ist.

## 6. Literatur

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). New York: Academic Press.

Grawe, K. & Braun, U. (1994). Qualitätskontrolle in der Psychotherapiepraxis. *Zeitschrift für klinische Psychologie*, 23, 242-267.

Jacobson, N. S. & Truax, P. (1991). Clinical significance. A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 1, 12-19.

Lambert, M. J. & Hill, C. E. (1994). Assessing psychotherapy outcomes and process. In A. E. Bergin & S. L. Garfield (Ed.). Handbook of Psychotherapy and Behaviour change. pp. 73 - 113. New York: Wiley.

Lienert, G. (1961). *Testaufbau und Testanalyse*. (3. Auflage). Weinheim: Beltz.

Nachtigall, C. & Suhl, U. (2002). Der Regressionseffekt. Mythos und Wirklichkeit. *metheval report 4* (*2*).

Nachtigall, C. & Wirtz, M. (2002). *Inferenzstatistik und Grundzüge der Wahrscheinlichkeits-rechnung*. (2. Aufl.). Weinheim: Juventa.

Steyer, R., Hannöver, W., Telser, Ch. & Kriebel, R. (1997). Zur Evaluation intraindividueller Veränderung. *Zeitschrift für Klinische Psychologie*, 26, 291-299.

Suhl, U., & Nachtigall, C. (2002). Warum kompliziert, wenn es auch einfach geht. Teil 2. Ergebnisse einer Simulationsstudie zum Vergleich von Veränderungskennwerten. *metheval report 4* (*4*).

Wirtz, M. & Nachtigall, C. (2002). *Deskriptive Statistik*. (2. Aufl.). Weinheim: Juventa.