

# Evaluation intraindividuelle Veränderung

## Ein Vergleich verschiedener Veränderungskennwerte

Christof Nachtigall<sup>1</sup> und Ute Suhl<sup>2</sup>

<sup>1</sup>Friedrich-Schiller-Universität, Jena

<sup>2</sup>Max-Planck-Institut für Kognitions- und Neurowissenschaften, Leipzig

**Zusammenfassung.** Zur Frage der Erfassung von intraindividuellen Therapieeffekten auf der Basis von Prä- und Post-Messungen liegt eine Reihe von Lösungsvorschlägen vor. Grawe und Braun (1994) schlugen als deskriptiven Kennwert die standardisierte Post-Prä-Differenz vor. Mit der Kritischen Differenz nach Lienert (1961) bzw. deren Umformulierung als Reliable Change Index (Jacobson & Truax, 1991), existieren weit verbreitete inferenzielle Kennwerte. Diese etablierten Kennwerte werden von Steyer, Hannover, Telser und Kriebel (1997) mit dem Argument kritisiert, dass die Unreliabilität der Messinstrumente sowie Regressionseffekte nicht adäquat berücksichtigt seien. Stattdessen werden dort alternative Kennwerte zur Erfassung intraindividuelle Veränderung vorgeschlagen. Der vorliegende Beitrag untersucht diese Alternativen und vergleicht sie mit den etablierten Kennwerten. Es zeigt sich, dass die alternativen Kennwerte mit großen Problemen sowohl hinsichtlich der Einhaltung des  $\alpha$ -Fehlers als auch bei der Berücksichtigung des Regressionseffektes behaftet sind, so dass von einer Verwendung abzuraten ist.

Schlüsselwörter: Psychotherapieforschung, Veränderungsmessung, Evaluation, Qualitätssicherung, intraindividuelle Veränderung, kritische Differenz, Reliable Change Index (RCI), Regression zur Mitte

Evaluation of intra-individual change: a comparison of different statistical measures

**Abstract.** A number of proposals have been put forward about how to calculate intra-individual therapy effects based on pre-post measurement. As a descriptive statistic, Grawe and Braun (1994) proposed the standardized pre-post difference. The Critical Difference according to Lienert (1961) as well as its reformulation as Reliable Change Index (Jacobson & Truax, 1991) are widely used inferential statistics. Steyer, Hannover, Telser, and Kriebel (1997) criticized these established statistics, arguing that they do not account appropriately for the unreliability of measurement instruments as well as for the problem of regression towards the mean, and provided alternative statistics to measure intra-individual change. The current article examines these alternatives and compares them with the established statistics. The results show that these alternative statistics have severe shortcomings with respect to type I errors, as well as with respect to the regression effect. Therefore, their application cannot be recommended.

Key words: Psychotherapy research, measurement of change, evaluation, quality assurance, intra-individual change, critical difference, reliable change index (RCI), regression towards the mean

Um den Effekt einer Behandlung bei einzelnen Patienten zu bewerten, ist es naheliegend, Prä- und Post-Messungen auf relevanten klinischen Skalen miteinander zu vergleichen. Zwar ist bekannt, dass Veränderungen zwischen einer Prä-Messung  $x$  und einer Post-Messung  $y$  nicht notwendigerweise ein Effekt der Behandlung sein müssen, dennoch stellt die Post-Prä-Differenz<sup>1</sup>  $y-x$  ein beliebtes deskriptives Maß für intraindividuelle Veränderung dar, das häufig zur Evaluation von Behandlungseffekten herangezogen wird. Ein Überblick über häufig verwendete Methoden psychotherapeutischer Veränderungsmes-

sung findet sich z. B. in Lambert und Hill (1994). Allerdings ist ein bloßer Differenzenwert  $y-x$  schlecht zu interpretieren, wenn die verwendeten Skalen dem Anwender nicht wohlbekannt sind. Ein Unterschied von z. B. 5 Punkten auf einer Skala kann eine große oder kleine Veränderung bedeuten, je nachdem, wie groß die Streuung der Messwerte auf dieser Skala ist. Daher schlugen Grawe und Braun (1994) vor, analog zu Effektstärkemaßen (siehe z. B. Cohen, 1988), die Post-Prä-Differenz zu standardisieren, in-

<sup>1</sup> Sprachlich gebräuchlicher ist der Begriff der *Prä-Post-Differenz*  $x-y$ . Jedoch ist in den Fällen, bei denen sich Behandlungserfolg in *höheren* Testscores niederschlägt, die Verwendung der Post-Prä-Differenz empfehlenswert, damit positive Veränderungen der Patienten sich auch in positiven Kennwerten ausdrücken. Die hier diskutierten Kennwerte können natürlich genauso auf Prä-Post-Differenzen ‚umgepolt‘ werden.

Die Forschungsergebnisse, auf die sich der vorliegende Artikel bezieht, entstanden in der Zeit, als die Koautorin als wissenschaftliche Mitarbeiterin am Institut für Psychologie der FSU Jena tätig war.

dem die Differenz  $y-x$  durch die Streuung des Prätestes  $s_x$  geteilt wird:

$$\text{Standardisierte Post-Prä-Differenz} = \frac{y-x}{s_x}. \quad (1)$$

Soll geschaut werden, ob sich die Post-Prä-Differenz mehr als nur zufällig von Null unterscheidet, wird häufig die *kritische Differenz* (Lienert, 1961, S. 455) bzw. der *Reliable Change Index (RCI)* (Jacobson & Truax, 1991) verwendet. Die *kritische Differenz* gibt die minimale Post-Prä-Differenz an, ab der eine intraindividuelle Veränderung signifikant von Null verschieden ist. Beim *RCI* wird demgegenüber die Differenz  $y-x$  durch deren Standardabweichung  $s_{y-x}$  geteilt:

$$RCI = \frac{y-x}{s_{y-x}}. \quad (2)$$

Zur Prüfung, ob eine Post-Prä-Differenz signifikant von Null verschieden ist, werden sowohl bei der *kritischen Differenz* als auch beim *RCI* normalverteilte Testwertvariablen  $x$  und  $y$  mit unkorrelierten Messfehlern und gleicher Messfehlervarianz vorausgesetzt. Beide Kennwerte sind lediglich Umformungen derselben Formel und werden daher in diesem Beitrag gemeinsam betrachtet.

## Die Kritik von Steyer et al. (1997)

In ihrem 1997 in der Zeitschrift für Klinische Psychologie erschienenen Artikel kritisieren Steyer, Hannover, Telser und Kriebel die oben genannten Veränderungskennwerte und schlagen Alternativen vor. Hinsichtlich der standardisierten Post-Prä-Differenz wird eingewendet, dass „...dabei das Meßfehlerproblem vernachlässigt wird. Die Konsequenz davon ist, daß eine Post-Prä-Differenz, die ausschließlich auf Meßfehlern beruht, genauso groß aussehen kann, wie eine Post-Prä-Differenz, die auf perfekt reliablen Messungen basiert“ (Steyer et al., 1997, S. 292). Ihr Verbesserungsvorschlag greift eine Idee von Dubois (1957) bzw. Manning und Dubois (1967) auf: Es wird nicht die Post-Prä-Differenz  $y-x$ , sondern die Differenz zwischen dem tatsächlichen Posttestwert  $y$  und dem zu *erwartenden* Posttestwert  $y'$  betrachtet<sup>2</sup>. Der erwartete Posttestwert  $y'$  ist der Wert, der bei einem Prätestwert  $x$  zu erwarten ist, wenn bei den wahren Werten der Person keine Veränderung vorlag, wenn also lediglich Messfehler zu Fluktuationen führen. Anders ausgedrückt: Anstelle des Prätestwertes wird ein korrigierter Wert verwendet, bei dem die *Regression zur Mitte* berücksichtigt wird (für eine zusammenfassende Darstellung der Problematik der Regression zur Mitte vgl. Nachtigall & Suhl, 2002a). Wird diese ‚korrigierte‘ Differenz  $y-y'$  zur Standardisierung ana-

log zum Vorgehen von Grawe und Braun (1994) durch die Streuung des Prätestes  $s_x$  geteilt, so erhält man den von Steyer et al. (1997) vorgeschlagenen deskriptiven Veränderungskennwert.

$$V_{\text{deskript}} = \frac{y-y'}{s_x}. \quad (3)$$

Als inferenziellen Veränderungskennwert  $V_{\text{infer}}$  schlagen Steyer et al. eine Alternative zur *kritischen Differenz* bzw. zum *RCI* vor. Sie betrachten wie beim deskriptiven Index  $V_{\text{deskript}}$  den Ausdruck  $y-y'$  und teilen durch dessen Streuung:

$$V_{\text{infer}} = \frac{y-y'}{s_{y-y'}}. \quad (4)$$

Diese Größe ist unter den üblichen Bedingungen (normalverteilte Testwertvariablen  $x$  und  $y$ , unkorrelierte Messfehler mit gleicher Messfehlervarianz) standardnormalverteilt. Eine Testung von intraindividuelle Veränderung bei einem Signifikanzniveau von  $\alpha = 0.05$  kann demnach durch den Vergleich von  $|V_{\text{infer}}|$  mit dem kritischen Wert 1.96 erfolgen.

Damit stehen zur Evaluation intraindividuelle Veränderung von Patienten verschiedene Alternativen zur Verfügung, die den praktischen Anwender vor die Frage stellen, in welcher Situation welcher Kennwert vorzuziehen ist. Der folgende Abschnitt gibt dazu klare Empfehlungen.

## Vergleich der verschiedenen Kennwerte

Die von Steyer et al. (1997) vorgeschlagenen Kennwerte finden in neueren klinischen Studien bereits Anwendung (z. B. Mestel, Vogler & Klingelhöfer, 2001). Dies hat praktische Konsequenzen für die Evaluationsergebnisse, da *RCI* und  $V_{\text{infer}}$  unterschiedliche Werte annehmen und für den einzelnen Patienten zu unterschiedlichen Aussagen über den Therapieerfolg führen können. Daher ist es wichtig, die Eigenschaften dieser Kennwerte genauer zu beleuchten.

Beginnen wir mit dem inferenziellen Kennwert  $V_{\text{infer}}$ : Dessen statistische Eigenschaften haben wir sowohl theoretisch als auch im Rahmen einer Simulationsstudie eingehend untersucht und mit denen des *RCI* verglichen. Die Details können in Suhl und Nachtigall (2002) nachgelesen werden. Zur Erinnerung: Beide Kennwerte sollen zu einem vorgegebenen Signifikanzniveau testen, ob sich der wahre Wert eines Patienten verändert hat. Es zeigt sich, dass beide Tests insgesamt das vorgegebene Signifikanzniveau  $\alpha$  einhalten. Dies gilt beim *RCI* unabhängig davon, wie groß der wahre Wert der untersuchten Person ist. Diese Eigenschaft ist gerade für einen Test der intraindividuellen Veränderung von besonderer Wichtigkeit. Bei  $V_{\text{infer}}$  hingegen kann man *nicht* von einem einheitlichen  $\alpha$ -Fehler für alle Personen ausgehen, vielmehr schwankt der

<sup>2</sup> Der erwartete Posttestwert  $y'$  berechnet sich aus dem Prätestwert  $x$  durch den Ausdruck  $y' = \text{Rel}(X)(x-E(X)) + E(X)$ . Dabei ist  $E(X)$  der Erwartungswert vom Test  $X$  und  $\text{Rel}(X)$  seine Reliabilität. Zur Herleitung vgl. Steyer et al. (1997) oder Suhl und Nachtigall (2002).

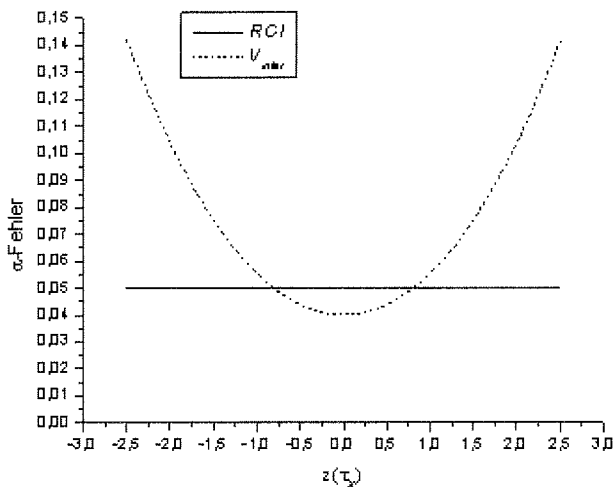


Abbildung 1.  $\alpha$ -Fehler-Wahrscheinlichkeit in Abhängigkeit vom wahren Wert  $\tau_x$  für die Kennwerte  $RCI$  (durchgezogene Linie) und  $V_{infer}$  (gepunktete Linie).

tatsächliche  $\alpha$ -Fehler in Abhängigkeit vom wahren Wert einer Person zum Teil erheblich. Der Grund dafür liegt darin, dass  $V_{infer}$  im Gegensatz zum  $RCI$  bei unterschiedlichen wahren Werten der Person *nicht* mehr standardnormalverteilt ist (der Beweis findet sich im Anhang). Dies führt dazu, dass bei Verwendung von  $V_{infer}$  das vorgegebene  $\alpha$ -Niveau teils unter-, teils deutlich überschritten wird.

Abbildung 1 illustriert diesen Zusammenhang zwischen tatsächlichem  $\alpha$ -Fehler und wahren Wert der Person. Zur besseren Interpretierbarkeit wurden z-transformierte Werte verwendet. Der tatsächliche  $\alpha$ -Fehler wurde auf der Basis des im Anhang beschriebenen Verfahrens

berechnet werden. Es zeigt sich folgendes: Gibt es keine wahre Veränderung und ist der wahre Wert einer Person nahe am Mittelwert (z-Wert nahe Null), dann ist der tatsächliche  $\alpha$ -Fehler bei  $V_{infer}$  (gepunktete Linie) kleiner als das vorgegebene Signifikanzniveau. Je weiter der wahre Wert einer Person vom Mittelwert abweicht, umso höher ist der tatsächliche  $\alpha$ -Fehler. Bei extremen Werten kann es zu mehr als einer Verdopplung kommen. Beim  $RCI$  hingegen ist der  $\alpha$ -Fehler unabhängig vom wahren Wert der Person, wie man an der durchgezogenen Linie erkennen kann. In Abbildung 1 wurden ein Signifikanzniveau von 5% und eine Reliabilität des Messinstruments von 0,80 zu Grunde gelegt, was einem typischen Anwendungsfall entspricht. Das Ergebnis ist jedoch nicht auf diese speziellen Werte beschränkt, sondern gilt in analoger Weise für andere Signifikanzniveaus und Reliabilitäten.

Kommen wir nun zu den Eigenschaften des deskriptiven Veränderungskennwertes  $V_{deskript}$ : Es stellt sich die Frage, was ein deskriptiver Veränderungsindex leisten kann und soll. Für die meisten Anwender dürfte es das zentrale Anliegen sein, das Ausmaß der wahren Veränderung (i. e. die Differenz der wahren Werte  $\tau_y - \tau_x$ ) auf Grund der Messwerte  $y$  und  $x$  möglichst genau abzuschätzen. In Nachtigall und Suhl (2002 b) wird gezeigt, dass für dieses Ziel die Post-Prä-Differenz einen erwartungstreuen Schätzer<sup>3</sup> darstellt, entsprechendes gilt für den von Grawe und Braun (1994) vorgeschlagenen Index. Bei  $V_{deskript}$  zeigt sich ähnlich wie beim inferenziellen Index  $V_{infer}$ , dass es in Abhängigkeit von der Größe der Werte einer Person zu Verzerrungen kommt. Die folgende Tabelle zeigt einige prototypische Fälle:

Tabelle 1 illustriert das Verhalten von  $V_{deskript}$  im Vergleich zur standardisierten Differenz  $(y-x)/s_x$  nach Grawe und Braun (1994) am Beispiel verschiedener möglicher wahrer Veränderungen von Patienten auf einer klinischen Skala im Rahmen einer Therapie. Zur besseren Verständ-

Tabelle 1. Veranschaulichung der Eigenschaften von  $V_{deskript}$  anhand von Beispielen wahrer Veränderung.  $\tau_x$  ist der wahre Prätestwert und  $\tau_y$  der wahre Posttestwert. In die konkrete Berechnung geht die Reliabilität des Messinstruments ein. Für diese Beispiele wurde eine Reliabilität von 0,70 zu Grunde gelegt

	$\tau_y$	$\tau_x$	Erwartungswert der standardisierten Differenz	Erwartungswert von $V_{deskript}$	Kommentar zum Verhalten von $V_{deskript}$
Fall 1	4 -2	3 -1	1 -1	1,9 -1,3	Die wahre Veränderung wird überschätzt.
Fall 2	2 0	3 -1	-1 1	-0,1 0,7	Die wahre Veränderung wird unterschätzt.
Fall 3	2,5	3	-0,5	0,4	$V_{deskript}$ zeigt Veränderung in falscher Richtung an.

<sup>3</sup> Man spricht von einem erwartungstreuen Schätzer, wenn die Schätzungen weder systematisch zu groß noch zu klein sind, genauer, wenn der Erwartungswert des Schätzers die gesuchte Größe ist. Dies ist bei der Post-Prä-Differenz  $y-x$  der Fall.

lichkeit sind in der Tabelle z-transformierte Werte wiedergegeben. In Fall 1 sind die wahren Posttestwerte weiter vom Mittelwert entfernt als die wahren Prätestwerte.  $V_{deskript}$  gibt als Erwartungswert eine größere Veränderung als der Erwartungswert der standardisierten Differenz an. Im 2. Fall liegen die wahren Posttestwerte näher am Mittelwert, worauf  $V_{deskript}$  vom Erwartungswert her kleinere Veränderungen signalisiert als der Erwartungswert der standardisierten Differenz. Der dritte Fall zeigt, dass die verschiedenen Kennwerte vom Erwartungswert her z.T. sogar auf unterschiedliche Richtungen der Veränderung schließen. Da die standardisierte Differenz erwartungstreu die Veränderung der wahren Werte schätzt, wird bei Verwendung von  $V_{deskript}$  in der Situation von Fall 1 die wahre Veränderung systematisch über-, in Fall 2 unterschätzt. In Fall 3 würde  $V_{deskript}$  hinsichtlich der Richtung der Veränderung zu systematisch falschen Schlüssen führen. Mithin ist  $V_{deskript}$  wenig geeignet, die Größe und Richtung der wahren Veränderung (ausgedrückt durch die Differenz der wahren Werte) zu beschreiben.

## Diskussion

Der Vergleich der statistischen Eigenschaften der (standardisierten) Post-Prä-Differenz bzw. des  $RCI$  und der kritischen Differenz mit den von Steyer et al. (1997) vorgeschlagenen Alternativen führt zu dem Schluss, dass sowohl für eine deskriptive Beschreibung der Größe der Veränderung als auch für inferenzielles Testen weiterhin die etablierten Kennwerte heranzuziehen sind und auf eine Verwendung von  $V_{deskript}$  und  $V_{infer}$  verzichtet werden sollte.

Neben diesen statistischen Aspekten gibt es auch theoretische Überlegungen, die für eine Verwendung der etablierten Kennwerte sprechen. Zwar sehen Steyer et al. (1997) den Vorteil ihres Kennwertes  $V_{infer}$  gegenüber der *kritischen Differenz* gerade darin, dass „... bei der Konstruktion dieser Statistik das Problem der Regression zur Mitte berücksichtigt wird“ (ebd., S. 293). In Nachtigall und Suhl (2002 b) wird jedoch gezeigt, dass der  $RCI$  bzw. die *kritische Differenz* die Regression zur Mitte durchaus berücksichtigen. Dieser Befund scheint zunächst paradox, geht doch in diese Kennwerte die ‚unbereinigte‘ Post-Prä-Differenz  $y-x$  ein. Regression zur Mitte ist jedoch im Allgemeinen kein zeitlich gerichtetes Phänomen. Nicht nur große Prätestwerte lassen kleinere Posttestwerte erwarten, auch große Posttestwerte lassen Prätestwerte erwarten, die näher am Mittelwert liegen. In Nachtigall und Suhl (2002 a) wird die Problematik der Regression zur Mitte systematisch dargestellt und am Beispiel der klassischen Fragestellung (Körpergröße von Eltern und Kindern) illustriert. Große Eltern haben im Schnitt weniger große Kinder, aber auch große Kinder haben im Schnitt weniger große Eltern. Von daher ist sowohl beim Prä- als auch beim Posttestwert ein Regressionseffekt zu berücksichtigen, mithin sind beide Werte zu ‚bereinigen‘. Dies ist beim  $RCI$  bereits auf Grund seiner Konstruktion berücksichtigt, nicht aber bei  $V_{infer}$ : Bei diesem wird, wie oben ausgeführt, nur der Prätestwert korrigiert. In diesem Sinn ist  $V_{infer}$  hinsichtlich

des Regressionseffektes nur ‚halb‘ bereinigt. Auch das Argument der Berücksichtigung von Regressionseffekten spricht demnach für die Verwendung der kritischen Differenz anstatt  $V_{infer}$ . Eine Ausnahme stellt die Situation dar, in welcher klinische Gruppen erst auf Grund des Vortests selektiert werden. Wenn Personen auf Grund eines hohen (messfehlerbehafteten) Prätestwertes eine Behandlung beginnen, dann ist allein auf Grund der Regression zu Mitte eine Senkung der Symptomschwere zu erwarten, die Post-Prä-Differenz ist nicht mehr erwartungstreu und der Regressionseffekt nicht mehr symmetrisch von Prä- zu Post- und Post- zu Prätest. Bisher ist ungeklärt, welcher Veränderungskennwert in diesem Fall bessere Eigenschaften aufweist. Für den Praktiker empfiehlt es sich, Veränderungskennwerte *nicht* aus Messungen zu berechnen, die gleichzeitig zur Entscheidung über eine Behandlung führen. Dies kann dadurch sichergestellt werden, dass Patienten routinemäßig nach der Behandlungsentscheidung einen (weiteren) Prätest durchlaufen.

Auch bezüglich des alternativen deskriptiven Kennwertes  $V_{deskript}$  bleiben weitere kritische Fragen. Das erklärte Ziel von Steyer et al. lautet, ihr deskriptiver Index solle „...zwischen einer zufälligen, meßfehlerbedingten Fluktuation und einer systematischen, tatsächlichen Veränderung unterscheiden“ (Steyer et al., 1997, S. 295). Dies ist unserer Auffassung nach jedoch keine deskriptive, sondern eine inferenzielle Frage, in dem Sinne, dass von messfehlerbehafteten Werten auf wahre Werte geschlossen werden soll. Just zur Beantwortung dieser Frage wurden inferenzielle Verfahren wie der  $RCI$  bzw. die kritische Differenz entwickelt. Ein deskriptiver Kennwert wie  $V_{deskript}$  kann demgegenüber bei messfehlerbehafteten Daten niemals ‚unterscheiden‘, ob z.B. bei einem Posttestwert von 8 Punkten und ein Prätestwert von 7.5 Punkten eine wahre Veränderung oder nur eine messfehlerbedingte Veränderung vorliegt. Ist  $V_{deskript} = 0$ , so kann trotzdem eine wahre Veränderung vorliegen und ein von Null verschiedener Wert von  $V_{deskript}$  kann ebenso messfehlerbedingt zu Stande kommen. Diese Einschränkung trifft wohlge- merkt auf jeden deskriptiven Kennwert (also auch auf (standardisierte) Post-Prä-Differenzen) zu. Unserer Einschätzung nach sollte das Ziel eines deskriptiven Kennwertes für intraindividuelle Veränderung aber sein, die Größe der wahren Veränderung möglichst genau und erwartungstreu zu schätzen, und für dieses Ziel ist die Post-Prä-Differenz (bzw. deren standardisierte Variante von Grawe und Braun, 1994) in der Regel ein sehr gutes Verfahren. Von der Verwendung von  $V_{deskript}$  ist hingegen wegen der beschriebenen Verzerrung abzuraten.

## Literatur

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2<sup>nd</sup> ed.). New York: Academic Press.
- Dubois, P. H. (1957). *Multivariate correlational analysis*. New York.
- Grawe, K. und 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.
- Manning, W. H. & Dubois, P. H. (1967). Correlation methods in research on human learning. *Perceptual and Motor skills*, 15, 287–321.
- Mestel, R., Vogler, J. & Klingelhöfer, J. (2001). Rückmeldung der testpsychologisch ermittelten Ergebnisqualität in der stationären Psychosomatik. In M. Bassler (Hrsg.), *Störungsspezifische Therapieansätze – Konzepte und Ergebnisse*. Gießen: Psychosozialverlag.
- Nachtigall, C. & Suhl, U. (2002 a). Regression zur Mitte. Mythos und Wirklichkeit. *Metheval report 4* (2). Verfügbar unter: [www.uni-jena.de/svw/metheval/report/](http://www.uni-jena.de/svw/metheval/report/) [11.05.2005].
- Nachtigall, C. & Suhl, U. (2002 b). Warum kompliziert, wenn es einfach geht. Teil 1: Zur Analyse intraindividuelle Veränderung. *metheval report 4* (3) Verfügbar unter: [www.uni-jena.de/svw/metheval/report/](http://www.uni-jena.de/svw/metheval/report/) [11.05.2005].
- Steyer, R., Hannover, W., Telser, Ch. & Kriebel, R. (1997). Zur Evaluation intraindividuelle Veränderung. *Zeitschrift für Klinische Psychologie*, 26, 291–299.
- Suhl, U. & Nachtigall, C. (2002). Warum kompliziert, wenn es einfach geht. Teil 2. Ergebnisse einer Simulationsstudie zum Vergleich von Veränderungskennwerten. *metheval report*, 4 (4). Verfügbar unter: [www.uni-jena.de/svw/metheval/report/](http://www.uni-jena.de/svw/metheval/report/) [11.05.2005].

Manuskript eingereicht: 26. 06. 2003

Manuskript angenommen: 15. 10. 2004

Dr. Christof Nachtigall

Friedrich-Schiller-Universität Jena  
Institut für Psychologie  
Am Steiger 3  
Haus 1  
07743 Jena  
E-Mail: [christof.nachtigall@uni-jena.de](mailto:christof.nachtigall@uni-jena.de)

## Anhang

### Ableitung der Verteilungseigenschaften der Veränderungskennwerte

Es wird gezeigt, dass der  $RCI$  unabhängig vom wahren Wert  $\tau$  einer Person standardnormalverteilt ist, während sich bei  $V_{\text{infer}}$  Erwartungswert und Varianz ändern. Zur Vereinfachung der Ableitung werden zunächst nur die Zähler der beiden Prüfgrößen betrachtet. Vorausgesetzt werden die üblichen Eigenschaften (Normalverteilung der wahren Werte und der Messfehler, Unkorreliertheit der Messfehler sowie gleiche bedingte Varianz der Messfehler für alle wahren Werte  $\tau$ ). Ferner gehen wir ohne Einschränkung der Allgemeinheit zur Vereinfachung von z-standardisierten Messwerten  $X$  und  $Y$  aus. Die Ableitung erfolgt unter der Annahme der Gültigkeit der Nullhypothese, dass es keine wahre Veränderung gibt, dass also  $\tau_X = \tau_Y$  gilt. Hergeleitet werden Erwartungswert und Varianz der Prüfgrößen für  $\tau_X = \tau_a$ , dabei ist  $\tau_a$  ein beliebiger, aber fester wahrer Wert und  $\varepsilon$  der Messfehler.

#### Herleitung für den $RCI$ :

Erwartungswert:

$$\begin{aligned} E(Y - X \mid \tau_X = \tau_a) &= E(\tau_Y + \varepsilon_Y - \tau_X - \varepsilon_X \mid \tau_X = \tau_a) \\ &= E(\varepsilon_Y - \varepsilon_X \mid \tau_X = \tau_a) \\ &= E(\varepsilon_Y \mid \tau_X = \tau_a) - E(\varepsilon_X \mid \tau_X = \tau_a) \\ &= 0 \end{aligned}$$

Daraus folgt sofort

$$E\left(\frac{Y - X}{\sqrt{2(1 - \text{Rel}(X))}} \mid \tau_X = \tau_a\right) = \frac{1}{\sqrt{2(1 - \text{Rel}(X))}} E(Y - X \mid \tau_X = \tau_a) = 0$$

Varianz:

$$\begin{aligned} \text{Var}(Y - X \mid \tau_X = \tau_a) &= \text{Var}(\tau_Y + \varepsilon_Y - \tau_X - \varepsilon_X \mid \tau_X = \tau_a) \\ &= \text{Var}(\varepsilon_Y - \varepsilon_X \mid \tau_X = \tau_a) \\ &= \text{Var}(\varepsilon_Y \mid \tau_X = \tau_a) + \text{Var}(\varepsilon_X \mid \tau_X = \tau_a) \\ &= 2(1 - \text{Rel}(X)) \end{aligned}$$

Daraus lässt sich nun ableiten, dass die folgende Beziehung gilt:

$$\begin{aligned} \text{Var}\left(\frac{Y - X}{\sqrt{2(1 - \text{Rel}(X))}} \mid \tau_X = \tau_a\right) &= \frac{1}{2(1 - \text{Rel}(X))} \text{Var}(Y - X \mid \tau_X = \tau_a) \\ &= \frac{1}{2(1 - \text{Rel}(X))} \cdot 2(1 - \text{Rel}(X)) \\ &= 1 \end{aligned}$$

Damit ist gezeigt, dass unabhängig von der Ausprägung des wahren Wertes  $\tau_X = \tau_a$  für den  $RCI$  weiterhin  $E(RCI \mid \tau_X = \tau_a) = 0$  und  $\text{Var}(RCI \mid \tau_X = \tau_a) = 1$  gilt.

**Herleitung für  $V_{infer}$ :**

Erwartungswert:

$$\begin{aligned} E(Y - \text{Rel}(X) \cdot X \mid \tau_x = \tau_a) &= E(\tau_y + \varepsilon_y - \text{Rel}(X) \cdot (\tau_x + \varepsilon_x) \mid \tau_x = \tau_a) \\ &= E(\tau_y - \text{Rel}(X) \cdot \tau_x \mid \tau_x = \tau_a) + E(\varepsilon_y - \text{Rel}(X) \cdot \varepsilon_x \mid \tau_x = \tau_a) \\ &= \tau_a(1 - \text{Rel}(X)) \end{aligned}$$

Daraus folgt dann weiter

$$\begin{aligned} E\left(\frac{Y - \text{Rel}(X) \cdot X}{\sqrt{1 - \text{Rel}^2(X)}} \mid \tau_x = \tau_a\right) &= \frac{1}{\sqrt{1 - \text{Rel}^2(X)}} E(Y - \text{Rel}(X) \cdot X \mid \tau_x = \tau_a) \\ &= \frac{\tau_a(1 - \text{Rel}(X))}{\sqrt{1 - \text{Rel}^2(X)}} \end{aligned}$$

Varianz:

$$\begin{aligned} \text{Var}(Y - \text{Rel}(X) \cdot X \mid \tau_x = \tau_a) &= \text{Var}(\tau_y + \varepsilon_y - \text{Rel}(X) \cdot (\tau_x + \varepsilon_x) \mid \tau_x = \tau_a) \\ &= \text{Var}(\varepsilon_y \mid \tau_x = \tau_a) + \text{Rel}^2(X) \text{Var}(\varepsilon_x \mid \tau_x = \tau_a) \\ &= (1 - \text{Rel}(X)) + \text{Rel}^2(X)(1 - \text{Rel}(X)) \\ &= (1 - \text{Rel}(X)) \cdot (1 + \text{Rel}^2(X)) \end{aligned}$$

Daraus folgt nun für die Varianz der Prüfgröße  $V_{infer}$  in Abhängigkeit vom wahren Wert  $\tau$ :

$$\begin{aligned} \text{Var}\left(\frac{Y - \text{Rel}(X) \cdot X}{\sqrt{1 - \text{Rel}^2(X)}} \mid \tau_x = \tau_a\right) &= \frac{1}{1 - \text{Rel}^2(X)} \text{Var}(Y - \text{Rel}(X) \cdot X \mid \tau_x = \tau_a) \\ &= \frac{(1 - \text{Rel}(X)) \cdot (1 + \text{Rel}^2(X))}{1 - \text{Rel}^2(X)} = \frac{1 + \text{Rel}^2(X)}{1 + \text{Rel}(X)} \end{aligned}$$

Die Herleitung zeigt, dass  $V_{infer}$  bei gegebenem wahren Wert  $\tau_x = \tau_a$  nicht mehr standardnormalverteilt ist. Insbesondere der bedingte Erwartungswert  $E(V_{infer} \mid \tau_x = \tau_a)$  ist in der Regel von Null verschieden, da er von  $\tau_a$  abhängt. Nur für den Ausnahmefall  $\tau_a = 0$  gilt auch  $E(V_{infer} \mid \tau_x = \tau_a) = 0$ .