

Roland Bröhe

**Überprüfung der Validität und Reliabilität
eines Instruments zur Einschätzung
des Dekubitusrisikos
am Beispiel einer Modifizierten Nortonskala**

Korrespondenzanschrift:

Roland Bröhe
Keupstraße 14
51063 Köln
E-Mail: bruehe@gmx.de

Mai 2002

Inhalt

Einleitung	2
Teil I: Überprüfungsmöglichkeiten einer Einschätzungsskala	3
1. Reliabilität.....	3
1.1 Die (Test)-Wiederholungsreliabilität	3
1.2 Parallel-Test	3
2. Validität.....	3
2.1 Inhaltsvalidität.....	3
2.2 Kriteriumsvalidität	4
2.3. Konstruktvalidität.....	5
Teil II: Beispielhafte Überprüfung der Vorhersagegültigkeit	6
1. Angaben über die Studie	6
2. Überprüfung der Vorhersagevalidität.....	7
2.1 Berechnung von Sensibilität und Spezifität sowie PVP- und PVN-Wert.....	7
2.2 Bedeutung der ermittelten Werte	8
3. Abschluss.....	10
Anhang	11
Literatur	12

Einleitung

Die Berufsgruppe der Pflegenden hat das Bedürfnis, ihr Handeln objektiver darzustellen. Gerade im Hinblick auf viele Risiken, die für einen Patienten im Krankenhaus bestehen können (z.B. Dekubitusgefahr, Pneumoniegefahr), sind vielfältige prophylaktische Maßnahmen notwendig, die kriteriengestützt ausgewählt werden müssen. Eine Einschätzungsmöglichkeit des Risikos nach einheitlichen und aussagekräftigen Kriterien kann das Wohl des Patienten schützen, ihn evtl. vor unnötigen Maßnahmen bewahren und für den Krankenhausträger Begründungen für Ausgaben liefern.

Die Einschätzung des Dekubitusrisikos ist schon seit den 60er Jahren des letzten Jahrhunderts ein wichtiges Thema, seitdem sich Doreen NORTON mit ihrer „Nortonskala“ die Frage stellte, wann ein Patient tatsächlich gefährdet ist. Sie ist bislang kaum hinreichend auf ihre Validität und Reliabilität getestet worden (JAKOBS, S. 17), was mehrere Autoren dazu veranlasste, Ergänzungen und Veränderungen vorzunehmen.

Diese kleine Arbeit stellt in aller Kürze Überprüfungsmöglichkeiten für Reliabilität und Validität dar und geht auf eine schwedische Studie ein, deren Daten eine beispielhafte statistische Kontrollrechnung ermöglichen.

Teil I: Überprüfungsmöglichkeiten einer Einschätzungsskala

1. Reliabilität

Mit der Reliabilität wird eine Aussage über die Zuverlässigkeit eines Instruments getroffen. Dabei bezieht sie sich auf die Stabilität und Genauigkeit der Messungen. Das Maß stellt der *Reliabilitätskoeffizient* dar. Er ist definiert „als Anteil der wahren Varianz an der Gesamtvarianz einer Messreihe“ (JAKOBS, S.48). Dabei liegt sein Wert zwischen 1 und 0. Ein reliables Instrument liefert dabei einen Wert näher der 1, als akzeptabler Wert „gilt ein Niveau von 0,70 oder höher“ (ebd.).

Es sind zwei Formen der Reliabilität zu unterscheiden.

1.1 Die (Test)-Wiederholungsreliabilität

Diese Überprüfung erfolgt mit der Messung durch dasselbe Instrument an zwei oder mehreren aufeinanderfolgenden Zeitpunkten. Um eine Veränderung des Probanden auszuschließen, sollte der Zeitraum zwischen den Messungen kurz sein. Als Vergleichswert der verschiedenen Messungen wird der *Korrelationskoeffizient* ermittelt. Je höher, desto beständiger ist das Messinstrument. Wird diese Überprüfung durch zwei oder mehr verschiedene Beobachter durchgeführt, spricht man von der „inter-rater reliability“.

1.2 Parallel-Test

Hierbei werden bei dem selben Probanden zwei verschiedene Messskalen verwendet, die das gleiche Phänomen messen. Eines der Instrumente ist dabei erwiesen reliabel, so dass eine Übereinstimmung gemessen werden kann, wiederum in Form des Korrelationskoeffizienten.

2. Validität

Da mit der Reliabilität alleine nur Aussagen dazu gemacht werden können, ob das gesamte Instrument stabil ist, kann über die Richtigkeit des Gemessenen damit noch nichts gesagt werden. Dies ist Aufgabe der Validität, die die Gültigkeit einer Skala angibt, d.h. ob das Instrument tatsächlich das misst, was es vorgibt zu messen.

Auch hier können verschiedene Formen unterschieden werden.

2.1 Inhaltsvalidität

In den Items einer Skala sollten sich die ihr zugrunde liegenden Hypothesen widerspiegeln. Ist dies der Fall und „der Inhalt der Test-Items (erfasst) das zu messende Konstrukt in seinen wichtigsten Aspekten erschöpfend“ (JAKOBS, S. 50) kann von Inhaltsvalidität gesprochen werden. Betrachtet ein Experte, ob die Items

bei oberflächlicher Betrachtung intuitiv als richtig zu bezeichnen sind und die Testmethode geeignet scheint, spricht man von „*face validity*“, die als eine gröbere und subjektive Einschätzung zu sehen ist.

2.2 Kriteriumsvalidität

2.2.1 Vorhersagegültigkeit

Die wohl wichtigste Form der Skalenüberprüfung bildet die Darstellung der Vorhersagegültigkeit. Sie macht Aussagen über die Beziehung zwischen den Aussagen des Messinstruments nach einer Messung und dem tatsächlich beobachtbaren Verhalten des Probanden.

Zur Berechnung dieser Form der Kriteriumsvalidität kann eine einfache Matrix hinzugezogen werden (Vgl. JAKOBS, S. 52):

		Krankheit	
		positiv	negativ
Test	positiv	Messung richtig positiv (a)	Messung falsch positiv (b)
	negativ	Messung falsch negativ (c)	Messung richtig negativ (d)

Betrachtet man die Kriterien aus der Sichtweise des tatsächlich eingetretenen Zustandes, also ob ein gemessenes Phänomen eingetreten ist oder nicht, ist die Berechnung der *Sensibilität* und *Spezifität* notwendig. Es gilt also zu schauen, ob ein vorhergesagtes Ereignis tatsächlich eingetreten ist. Dabei macht die Sensibilität Aussagen über den Anteil der Subjekte, bei denen ein Phänomen eingetreten ist und bei denen dieses auch vorhergesagt wurde. Die Spezifität ist der dagegen der Anteil zwischen denen, bei denen das Phänomen nicht eingetreten ist und für die dieses auch nicht vorhergesagt wurde.

Nach JAKOBS ist es möglich, aus den Messungen von Sensibilität und Spezifität den „cut-off-point“ zu bestimmen, „wenn Spezifität auf einer x-Achse und Sensibilität auf einer y-Achse dargestellt werden“ (ebd., S. 51). Der „cut-off-point“ liegt dann am Wendepunkt der Kurve.

Zur Überprüfung der Messungen von Sensibilität und Spezifität kann eine Berechnung der Werte aus dem Vergleich der Testergebnisse mit den in der Realität eingetretenen Zuständen erfolgen. Diese Überprüfung ermittelt den *PVP-Wert* (predictive value of a positive test), also die Vorhersagegültigkeit für ein positives Ergebnis, und den *PVN-Wert* (predictive value of negative results), der Vorhersagegültigkeit für negative Ergebnisse.

Die Unterschiede zwischen den beiden Messmethoden wird deutlich, wenn in Teil II die Ergebnisse einer Studie konkret angewendet werden.

2.2.2 Konkurrierende Validität

Mit dem Korrelationskoeffizient wird der Vergleich zweier ebenbürtiger Instrumente angegeben. Beide Messinstrumente werden dabei parallel verwendet.

2.3. Konstruktvalidität

Zur Überprüfung der *Aussagegültigkeit* wird betrachtet, ob die im Instrument beschriebenen Beobachtungen genau identifiziert werden und eine funktionale Einheit bilden. Zum Beispiel kann überprüft werden, ob alle Beobachtungen bezüglich der Bewegung im Instrument eine funktionale Einheit bilden. Die Hypothesen müssen dabei als gültig identifiziert werden. So kann eine Aussage getroffen werden, ob alle Items bezeichnend für das zu messende Phänomen sind.

Die *Faktorielle Validität* überprüft in ähnlicher Weise, ob wechselseitig korrelierende Variablen, die zum zu messenden Phänomen beobachtbar sind, in wenigen Dimensionen (Faktoren) zusammengefasst werden.

Teil II: Beispielhafte Überprüfung der Vorhersagegültigkeit

In einer dieser Arbeit vorangegangenen Literaturrecherche wurden Studien gesucht, die eine „Modifizierte Nortonskala“ untersuchen (Datenbank Medline). Damit wurde abgezielt auf die in den achtziger Jahren des letzten Jahrhunderts vom damaligen Bildungszentrum des Deutschen Berufsverbands für Pflegeberufe (DBfK) veränderte Nortonskala zur Dekubitusrisikoeinschätzung. Die in der gefundenen Literatur¹ benannten Skalen waren jedoch in anderer Weise modifiziert und bezogen sich nicht auf das deutsche Instrument. Aus pragmatischen Gründen wurde eine der ermittelten Studien ausgewählt, um hier als exemplarisches Beispiel zur Verdeutlichung der Vorhersagegültigkeit zu dienen. Hilfreich war dabei, dass in der Studie umfangreiches Zahlenmaterial präsentiert wurde, wobei hier nur das Notwendigste wiedergegeben wird.

1. Angaben über die Studie

Die Autoren (GUNNINGBERG et al.) untersuchten die Prävalenz von Druckgeschwüren bei Patienten mit Oberschenkelhalsfrakturen, ob eine Modifizierte Norton-Skala (MNS, siehe Anhang) Dekubitus-Risikopatienten ermittelt und verglichen die Prävalenz von Dekubitalgeschwüren in einer Experimentalgruppe mit besonderer, täglicher Einschätzung.

Zum Zwecke der Untersuchung wurden Patienten mit Oberschenkelhalsfrakturen auf zwei Stationen verteilt nach einem bekannten Schema (Experimental- und Kontrollgruppe). Dem Pflegepersonal der Experimentalgruppe wurden als Instrumente die MNS und eine Dekubituseinschätzungsskala mit ausführlicher Beschreibung der verschiedenen Stadien in schriftlicher und abbildender Form überreicht. Eine Hauteinschätzung der Patienten geschah dort täglich. Bei einem cut-off-point von <21 wurden dekubitusprophylaktische Maßnahmen eingesetzt, die jedoch nicht vorgegeben waren. Eine Einschätzung mittels der MNS geschah bei Aufnahme in der Aufnahmeabteilung, 14 Tage postoperativ und bei Entlassung.

Die Studie ergab:

1. Bei Aufnahme in das Krankenhaus hatten ca. 20% aller Patienten in beiden Gruppen Dekubitalgeschwüre. Bei der Entlassung steigerte sich diese Rate auf 40% (Experimentalgruppe) bzw. 36% (Kontrollgruppe).
2. Die Anwendung der MNS konnte den Hauptanteil der gefährdeten Patienten bestimmen.

¹ Berglund (1995), Ek (1987), Gunningberg (1999), Stotts (1988)

3. Es konnte kein signifikanter Unterschied zwischen den beiden Gruppen bzgl. der Prävalenz der Dekubitalgeschwüre gefunden werden.

2. Überprüfung der Vorhersagevalidität

2.1 Berechnung von Sensibilität und Spezifität sowie PVP- und PVN-Wert

Zur Überprüfung, ob die Patienten, die einen Dekubitus bei Entlassung vorwiesen, auch durch die MNS als dekubitusgefährdet eingeschätzt wurden, ist zunächst die Berechnung der Sensibilität und Spezifität notwendig. Dazu verwenden wir die o.g. Matrix, in die wir die Werte einer Tabelle der Studie einfügen. Wir betrachten dabei exemplarisch die Werte zum Entlassungszeitpunkt.

Table 5 Number of patients with low (<21) or high (≥21) MNS scores in the A & E Department, related to pressure ulcers reported in the A & E Department, at discharge from the hospital and 14 ± 6 days post-surgery

	Experimental group (n = 58)		Control group (n = 66)	
	Low MNS (n = 33)	High MNS (n = 21)	Low MNS (n = 43)	High MNS (n = 21)
A & E Department ^{a,b}	7	4	10	2
Discharge	14	6 ^b	13	4 ^b
14 ± 6 days post-surgery	8	5 ^c	12	4

^a Six patients were not assessed in the A & E.

^b One patient with pressure ulcer at discharge was not assessed in the A & E.

^c Two patients with pressure ulcer 14 days post-surgery were not assessed in the A & E.

Tabelle: GUNNINGBERG et al. 1999

		Dekubitus	
		positiv	negativ
MNS- Skala	positiv (n=33)	14 (a)	19 (b)
	negativ (n=21)	6 (c)	15 (d)

Matrix zur Vorhersagegültigkeit mit Werten der Studie (Zeitpunkt Entlassung)

Als Test-positiv gelten hierbei alle Patienten, die einen MNS Score von weniger als 21 Punkten aufweisen. Dekubitus-positiv gelten alle, die ein Dekubitalgeschwür ab Grad 1 aufweisen.

Die Berechnung der Sensibilität und Spezifität erfolgt nach den folgenden Formeln:

Sensibilität	Spezifität
$\frac{a}{a + c} \times 100$	$\frac{d}{b + d} \times 100$
Aussage	
Die Formel setzt die Anzahl von Patienten zueinander in Beziehung, die <u>in der Realität</u> einen Dekubitus entwickelten.	Die Formel setzt die Anzahl von Patienten zueinander in Beziehung, die <u>in der Realität</u> keinen Dekubitus entwickelten.
Rechnung	
$\frac{14}{14 + 6} \times 100 = 70$	$\frac{15}{19 + 15} \times 100 = 44,1$

Die ermittelten Werte decken sich mit denen, die in der Studie genannt werden, wobei dies die einzige Berechnung zur Validität ist, die in der Studie erwähnt wird (GUNNINGBERG, S. 404).

Zur Überprüfung können nun PVP- und PVN-Wert ermittelt werden.

PVP-Wert	PVN-Wert
$\frac{a}{a + b} \times 100$	$\frac{d}{c + d} \times 100$
Aussage	
Die Formel setzt die Anzahl von Patienten zueinander in Beziehung, die <u>in der Messung</u> ein Dekubitusrisiko aufweisen.	Die Formel setzt die Anzahl von Patienten zueinander in Beziehung, die <u>in der Messung</u> kein Dekubitusrisiko aufweisen.
Rechnung	
$\frac{14}{14 + 19} \times 100 = 42,4$	$\frac{15}{6 + 15} \times 100 = 71,43$

2.2 Bedeutung der ermittelten Werte

Es ist den Zahlen zu entnehmen, dass eine hohe Anzahl (Sensibilität 70%) der Patienten, die einen Dekubitus entwickelten, auch als dekubitusgefährdet eingestuft wurden. Gleichzeitig kann aber auch ausgesagt werden, dass bei den positiv eingeschätzten Patienten die dekubitusprophylaktischen Maßnahmen, die ja mit Sicherheit eingeleitet wurden, versagt haben. Hier zeigt sich eine wichtige Problematik, nämlich die der nicht unbeeinflussten Messergebnisse. Ein ermittelter

Score hat Konsequenzen – und zwar Durchführung oder Unterlassung von spezifischen Pflegemaßnahmen. Eine optimale Testung würde bedeuten, trotz geringen MNS-Scores keine Prophylaxen durchzuführen und dann den Zustand erneut zu messen. Dies ist jedoch aus ethischen Gründen in keiner Weise vertretbar. Also muss diese Messungsgenauigkeit bei der Betrachtung mitgedacht werden.

Betrachtet man die PVP- und PVN-Werte, wird eine andere Perspektive eingenommen. Hier wird angeschaut, in welchem Verhältnis die als Test-positiv ausgewiesenen Patienten einen Dekubitus entwickelten oder nicht. Damit kann eine Aussage gemacht werden, inwieweit das Instrument zuverlässig das Dekubitusrisiko vorhersagt. Mit 42,4% weist die MNS einen geringen Wert auf. Viele der als positiv eingeschätzten Patienten entwickelten in der Tat keinen Dekubitus. Hier wirkt jedoch die o.g. Einschränkung ein. Die Bedeutung dieser Aussage hat zum einen für den Patienten Auswirkungen – er wird unnötigen Maßnahmen ausgesetzt, die seine Lebensqualität einschränken (z.B. regelmäßige Lagerungen), zum anderen hat das Ergebnis wirtschaftliche Folgen – unnötige Maßnahmen verursachen unnötige Kosten (Material, Personalzeit).

Der PVN-Wert ist mit 71,43% in einem akzeptablen Rahmen. Ein großer Teil der als nichtgefährdet eingestuften Patienten haben in der Tat keinen Dekubitus entwickelt. Inwieweit hier trotzdem prophylaktische Maßnahmen durchgeführt wurden, ist jedoch nicht in der Studie erwähnt.

Die gleichen Berechnungen sind natürlich mit der Kontrollgruppe ebenso möglich. Aus Platzgründen wird an dieser Stelle darauf verzichtet.

Erst die Aussagen von Sensibilität/Spezifität und PVP- bzw. PVN-Wert erlauben eine Beurteilung, ob ein Messinstrument wie die Modifizierte Nortonskala von seinen inhaltlichen Gesichtspunkten her ein gültiges Verfahren darstellt. Dabei betrachten beide Wertepaare die Ergebnisse einer Untersuchung aus verschiedenen Blickwinkeln: Von der Perspektive der Patienten (Dekubitus ja oder nein) bzw. von der Perspektive des Test (Gefährdung ja oder nein). Diese Perspektiven können auf die bekannte Matrix übertragen werden:

		Dekubitus	
		positiv	negativ
MNS-Skala	positiv (n=33)	14 (a)	19 (b)
	negativ (n=21)	6 (c)	15 (d)

Perspektiven der Vorhersagevalidität

Ein ausschließliche Reliabilitätsüberprüfung kann nicht zu einer derartigen Aussage kommen. Positiv ausgefallen würde mit ihr sogar eher ein irrtümlicher und für den Patienten evtl. fataler Nutzen bescheinigt, denn es könnte keine Aussage zur Güte des Ergebnisses getroffen werden. Dazu bedarf es der Überprüfung der Validität.

Zur genaueren Untersuchung der Validität des Instruments könnten außerdem auch einzelne Items kontrolliert werden in Bezug auf ihren Einfluss auf den Gesamtscore.

3. Abschluss

Es wurde aufgezeigt, zwischen welchen Formen der Reliabilität und Validität unterschieden werden kann. Dabei wurde der Vorhersagevalidität als Form der Kriteriumsvalidität große Aufmerksamkeit geschenkt. Anhand einer Studie, die eine Modifizierte Nortonskala zur Ermittlung des Dekubitusrisikos an Patienten mit Oberschenkelhalsfrakturen untersuchte, wurde eine exemplarische Überprüfung der Sensibilität und Spezifität sowie des PVP- als auch des PVN-Wertes vorgenommen. Schlussfolgernd konnte ausgesagt werden, dass eine hundertprozentige statistische Aussage nicht möglich ist aufgrund ethischer Probleme bei der Untersuchung (durch Vorenthalten von notwendigen Interventionen). Jedoch geben die ermittelten Werte einen wichtigen Anhalt für die inhaltliche Aussagefähigkeit eines Instrumentes wie der Nortonskala.

Anhang

Modified Norton Scale

(Ek, A.-C. et al [1997]: Quality indicators for patients with or with risk of developing pressure ulcers. In: *Quality Indicators in Nursing*. Omvårdnad 1 (Idwall E., ed.), Spri publications, Stockholm)

MENTAL CONDITION

4. Fully orientated to time and space
3. Occasionally confused
2. Cannot answer adequately
1. No contact

PHYSICAL ACTIVITY

4. Ambulant without assistance
3. Walks with help by others (possibly wheelchair for independent transportation)
2. Chair-bound (all day)
1. Bed-ridden

MOBILITY

4. Full
3. Slightly limited (some assistance in changing position)
2. Very limited (full assistance in changing position, but can help some)
1. Immobile (full assistance in changing position, cannot help)

FOOD INTAKE

4. Normal portion (or full parenteral)
3. $\frac{3}{4}$ of normal portion (or corresponding parenteral)
2. $\frac{1}{2}$ of normal portion (or corresponding parenteral)
1. Less than $\frac{1}{2}$ portion (or corresponding parenteral)

FLUID INTAKE

4. More than 1000 ml/day
3. > 700 ml - < 1000 ml/day
2. > 500 ml - < 700 ml/day
1. Less than 500 ml/day

INCONTINENCE

4. None
3. Occasionally (usually continent but not just now)
2. Urinary or anal incontinence
1. Urinary and anal incontinence

GERNERAL PHYSICAL CONDITION

4. Good (afebrile, normal respiration, frequency, rhythm, normal pulse, blood pressure, no pain, normal skin colour, appearance corresponds to age)
3. Fair (afebrile-subfebrile, normal respiration, pulse, and blood pressure, possible mild tachycardia, latent hypo/hypertonia, no or mild pain, alert, possibly pale skin, mild oedema)
2. Poor (possibly fever, abnormal respiration, signs of curculatory insufficiency, tachycardia, oedema, hypo/hypertonia, pain, somnolent or alert but apathetic. Skin is pale, or cyanotic, warm humid or cold humid, or depressed turgor or oedema)
1. Very poor (possible fever, abnormal respiration, distinct signs of circulatory insufficiency or shock, much pain, somnolent, stupor, comatose. Skin is pale, or cyanotic, warm and humid or cold and humid, or depressed turgor or oedema)

Literatur

Berglund, Britta; Nordström, Gun: The Use of the Modified Norton Scale in Nursing-Home Patients. In: Scandinavian Journal of Caring Science 1995; 9:165-169

Ek, Anna-Christina; Bjurulf, Per: Interrater Variability in a Modified Norton Scale. In: Scandinavian Journal of Caring Science 1987; 1: 99-102

Gunningberg, Lena et al.: Implementation of risk assessment and classification of pressure ulcers as quality indicators for patients with hip fractures. In: Journal of Clinical Nursing. 1999; 8:396-406

Jakobs, Birgit: Skalen zur Einschätzung des Dekubitusrisikos. Reliabilität und Validität der Norton- und Bradenskala. Hannover: Schlütersche, 2000.

Stotts, N.A.: Predicting pressure ulcer development in surgical patients. In: Heart & Lung 1988; 17:641-647