

Aus dem Institut für Medizin-/Pflegepädagogik und Pflegewissenschaft  
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

## **DISSERTATION**

Interrater-Reliabilität von Instrumenten zur Einschätzung des  
Dekubitusrisikos und des Dekubitus

zur Erlangung des akademischen Grades  
Doctor rerum curae (Dr. rer. cur.)

vorgelegt der Medizinischen Fakultät  
Charité – Universitätsmedizin Berlin

von

Jan Kottner

aus Berlin

Gutachter:     1. Prof. Dr. Theo Dassen  
                  2. Prof. Dr. Christa Lohrmann  
                  3. Prof. Dr. Peter Martus

Datum der Promotion: 20. April 2009

# Inhalt

	Seite
Zusammenfassung	3
Anteilerklärung	16
Publikationen	17

Kottner, J., & Dassen, T. (2008). An interrater reliability study of the Braden scale in two nursing homes. *International Journal of Nursing Studies*, 45, 1501-1511.

Kottner, J., Dassen, T., & Tannen, A. (2009). Inter- and intrarater reliability of the Waterlow pressure sore risk scale: A systematic review. *International Journal of Nursing Studie*, 46, 369-379.

Kottner, J., Tannen, A., Halfens, R., & Dassen, T. (2009). Does the number of raters influence the pressure ulcer prevalence rate? *Applied Nursing Research*, 22, 68-72.

Kottner, J., Tannen, A., & Dassen, T. (2009). Hospital pressure ulcer prevalence rates and number of raters. *Journal of Clinical Nursing*. doi: 10.1111/j.1365-2702.2008.02609.x

Kottner, J., Raeder, K., Halfens, R., & Dassen, T. (2009). A systematic review of interrater reliability of pressure ulcer classification systems. *Journal of Clinical Nursing*, 18, 315-336.

Lebenslauf	22
Vollständige Publikationsliste	23
Selbständigkeitserklärung	25

## Zusammenfassung

### Abstract

Dekubitus sind schwerwiegende Gesundheitsprobleme. Die Braden- und Waterlow-Skala sind verbreitete Instrumente zur standardisierten Einschätzung des Dekubitusrisikos. Beide Instrumente sind unter anderem nur dann für die klinische Praxis geeignet, wenn sie hohe Interrater-Reliabilitäten aufweisen. Die Einschätzung zum Vorhandensein von Dekubitus und die Klassifikation entsprechender Gewebeschädigungen sollten ebenfalls reliabel sein. Mit der vorliegenden Dissertation wurde untersucht, wie hoch die Interrater-Reliabilitäten der Braden- und Waterlow-Skala sind. Es wurde weiterhin untersucht, ob die Anzahl der Pflegekräfte (eins oder zwei), welche Hautinspektionen vornehmen, die Dekubitusprävalenz und die Wahrscheinlichkeit der Dekubitusdiagnose beeinflussen und welche Dekubitusklassifikation die höchste Interrater-Reliabilität aufweist. Die Interrater-Reliabilität der Braden-Skala wurde in zwei Pflegeheimen untersucht. Mit Hilfe eines Systematic Reviews wurden empirische Befunde zur Interrater-Reliabilität der Waterlow-Skala zusammengetragen. Der Einfluss der Anzahl der Pflegekräfte auf die Dekubitusprävalenz und auf die Wahrscheinlichkeit der Dekubitusdiagnose wurde durch sekundäre Datenanalysen von Pflegeheimbewohnern und Krankenhauspatienten untersucht. Um die Dekubitusklassifikation mit der höchsten Interrater-Reliabilität zu identifizieren, wurde ein umfangreiches Systematic Review durchgeführt. Die Interrater-Reliabilität der Braden-Skala im Setting Pflegeheim war hoch. Bis auf Ausnahmen waren die Differenzen zwischen den Pflegekräften gering. Es konnten jedoch deutliche Hinweise gefunden werden, dass bestimmte Items der Braden-Skala einen relativ hohen Fehleranteil aufweisen. Die Waterlow-Skala wurde in der klinischen Praxis bislang in nur vier Studien untersucht. Die Anteile der absoluten Übereinstimmungen und Differenzen sind mit denen der Braden-Skala vergleichbar. Interrater-Reliabilitätskoeffizienten wurden bislang nicht bestimmt. Da über die Interrater-Reliabilität der Waterlow-Skala nichts bekannt ist, sollte die Anwendung unter Vorbehalt erfolgen. Für beide Risiko-Skalen gilt, dass die in der Praxis häufig verwendeten Cut-Off-Punkte eine Genauigkeit vortäuschen, die kaum erreicht wird. Sowohl in Pflegeheimen als auch in Krankenhäusern unterschieden sich die ermittelten Dekubitusprävalenzen zwischen Teams von zwei Pflegekräften gegenüber einer einzelnen Pflegekraft nicht. Die Anzahl der Pflegekräfte, welche eine Hautinspektion vornahm, hatte ebenfalls keinen Einfluss auf die Wahrscheinlichkeit bei Patienten und Bewohnern einen Dekubitus zu identifizieren. Demzufolge könnte in zukünftigen Studien auf eine zweite Person verzichtet werden. Das würde personelle und zeitliche Ressourcen sparen. Untersuchungen zur Interrater-Reliabilität von Dekubitusklassifikationen sind heterogen und es liegen kaum aussagekräftige Ergebnisse für die klinische Praxis vor. Aufgrund der verfügbaren Evidenz ist es derzeit unmöglich zu entscheiden, welche Dekubitusklassifikation in der Praxis die höchsten Interrater-Reliabilitäten aufweist. Es fehlen Interrater-Reliabilitätsstudien, in denen vergleichbare Rater verschiedene Klassifikationssysteme unter vergleichbaren Bedingungen anwenden. Darüber hinaus sollte untersucht werden, wie die diagnostische Präzision bei der Klassifikation von Dekubitus in der Praxis verbessert werden kann.

## **Einleitung**

Dekubitus sind Bereiche lokalisierter Schädigungen der Haut und der darunter liegenden Gewebeschichten, welche durch Druck, Scherkräfte, Reibung oder eine Kombination dieser Faktoren verursacht wurden (European Pressure Ulcer Advisory Panel (EPUAP) 1998). Es handelt sich dabei um schwerwiegende Gesundheitsprobleme. Betroffene sind in ihrer Lebensqualität beeinträchtigt, die Behandlungen sind meist langwierig und kostenintensiv. Darüber hinaus wird das Auftreten von Dekubitus auf Mängel in der Gesundheitsversorgung zurückgeführt (von Renteln-Kruse 2004, Eberhardt et al. 2005, Püllen 2007).

Vor diesem Hintergrund hat die Vermeidung von Dekubitus höchste Priorität. Am Anfang einer effektiven Dekubitusprophylaxe steht die Einschätzung des individuellen Dekubitusrisikos. Zur Unterstützung dieser Risikoeinschätzung wird die Anwendung standardisierter Assessmentinstrumente empfohlen (Deutsches Netzwerk für Qualitätsentwicklung in der Pflege 2004; Eberhardt et al. 2005). Weltweit existieren circa 30 so genannter Risikoskalen (Schlömer 2003). Im deutschsprachigen Raum ist die Braden-Skala (Bergstrom et al. 1987) in Krankenhäusern und Pflegeheimen am weitesten verbreitet (Dassen et al. 2008). Sie besteht aus sechs Items. Jedes Item wird einzeln beurteilt und danach ein Summenwert berechnet. So können minimal 6 Punkte (maximales Dekubitusrisiko) bis maximal 23 Punkte (kein Dekubitusrisiko) erreicht werden. Die Waterlow-Skala (Waterlow 1985) wird vorwiegend in Krankenhäusern, so auch in der Charité – Universitätsmedizin Berlin, verwendet (Compton et al. 2008). Sie besteht aus 10 Items. Je höher der Waterlow-Summenwert, desto höher ist das Dekubitusrisiko. Die Validität dieser beiden Instrumente war bis heute Gegenstand zahlreicher Untersuchungen. Eine Meta-Analyse kam zu dem Schluss, dass die Braden-Skala das beste Verhältnis zwischen Sensitivität und Spezifität aufweist (Pancorbo-Hidalgo et al. 2006). Die Waterlow-Skala wird wegen ihrer zu niedrigen Spezifität kritisiert, dennoch gilt der Waterlow-Summenwert als ein guter Prädiktor für das Auftreten eines Dekubitus (Pancorbo-Hidalgo et al. 2006).

Validitätskriterien wie Sensitivität und Spezifität sind wichtig. Sie reichen jedoch nicht aus, den Einsatz entsprechend getesteter Instrumente zu rechtfertigen. Empirische Belege über die Reliabilität müssen ebenfalls erbracht werden. Da in der klinischen Praxis diagnostische Instrumente von vielen verschiedenen Anwendern benutzt werden, ist der Aspekt der Interrater-Reliabilität besonders entscheidend (Bartholomeyczik 2007). Unter Interrater-Reliabilität wird im Allgemeinen die Übereinstimmung zwischen verschiedenen Beobachtern oder Beurteilern, welche gleiche Patienten oder Sachverhalte einschätzen, verstanden (Szklo & Nieto 2007, Polit & Beck 2008). Bestehen zwischen den Urteilen große Unterschiede, dann deutet das auf einen hohen Fehleranteil innerhalb der Ratings hin. Die Interrater-Reliabilitäten der Braden- und Waterlow-Skala wurden bis heute weder systematisch analysiert noch gibt es verlässliche empirische Untersuchungen darüber (Kottner & Dassen 2008a, Kottner et al. 2009a).

Einschätzungsfehler sind bei Klassifikationen vorhandener Dekubitus ebenfalls von Bedeutung. Präzise Beurteilungen der Schweregrade von Dekubitus sind Voraussetzungen für eine eindeutige Kommunikation, Dokumentation, Behandlungsentscheidungen und aussagekräftige

Qualitätsbewertungen (internes und externes Benchmarking). Es existiert eine unüberschaubare Fülle von Dekubitusklassifikationen. Empirische Untersuchungen über deren Interrater-Reliabilität fehlen weitgehend oder sind kaum vergleichbar. Mit dem Ziel, die Präzision von Dekubituseinschätzungen zu erhöhen, empfiehlt beispielsweise das EPUAP (2002, 2005) das Hautassessment durch ein Team von zwei Pflegekräften gemeinsam durchzuführen. Diese Empfehlung wurde unter anderem in das Studienprotokoll der jährlich stattfindenden deutschlandweiten Dekubitusprävalenzerhebungen aufgenommen (Dassen et al. 2008). Es ist jedoch völlig offen, ob die Anzahl der Pflegekräfte tatsächlich einen Einfluss auf die Ergebnisse der Dekubitusdiagnostik hat. Der kürzlich veröffentlichte „Expertenstandard Pflege von Menschen mit chronischen Wunden“ (Deutsches Netzwerk für Qualitätsentwicklung in der Pflege 2008) empfiehlt ausdrücklich die Verwendung des EPUAP-Klassifikationssystems. Systematische Analysen empirischer Evidenz, die diese Empfehlungen stützen, gibt es nicht. Bis heute fehlen systematische Übersichtsarbeiten zur Beurteilung des Fehleranteils in der Dekubitusdiagnostik.

### **Zielstellung**

In der vorliegenden Promotion wurden folgende Fragestellungen untersucht:

- 1) Wie hoch ist die Interrater-Reliabilität der Braden-Skala? (**Studie 1**)
- 2) Wie hoch ist die Interrater-Reliabilität der Waterlow-Skala? (**Studie 2**)
- 3) Hat die Anzahl der Pflegekräfte, welche eine Hautinspektion durchführen, einen Einfluss auf die beobachtete Dekubitushäufigkeit? (**Studie 3**)
- 4) Welche Dekubitusklassifikation weist die höchste Interrater-Reliabilität auf? (**Studie 4**)

Für eine bessere Lesbarkeit und Nachvollziehbarkeit der Untersuchungen werden im Folgenden die Methodiken und Ergebnisse der einzelnen Studien (1 bis 4) jeweils getrennt dargestellt. Die Diskussion bezieht sich auf alle vier Teiluntersuchungen.

### **Methodik und Ergebnisse Studie 1**

#### Methodik

Eine Interrater-Reliabilitätsstudie wurde im Jahr 2007 in zwei Pflegeheimen durchgeführt. Pflegeheimbewohner ( $n = 152$ ) von insgesamt acht Wohnbereichen wurden anhand der Braden-Skala von Pflegekräften zweimal unabhängig voneinander in kurzem zeitlichem Abstand beurteilt. Die Pflegekräfte, welche das zweite Assessment durchführten, waren am ersten Assessment nicht beteiligt. Darüber hinaus hatten sie keinen Zugang zu den Ergebnissen des ersten Assessments. Eine Kommunikation zwischen den beteiligten Pflegekräften war nicht erlaubt.

Um ein differenziertes Bild über die Interrater-Reliabilität zu erhalten wurden prozentuale Übereinstimmungen und Interrater-Reliabilitätskoeffizienten für jedes Item der Braden-Skala und den Braden-Summenwert berechnet. Die prozentuale Übereinstimmung ( $p_o$ ) gibt Auskunft über den absoluten Anteil der Übereinstimmung. Daneben wurden  $\kappa$ -Koeffizienten (Cohen 1960) berechnet. Diese werden im Allgemeinen als zufallskorrigierte Maße interpretiert, das heißt sie zeigen an, wie

hoch die Übereinstimmung abzüglich der zufallsbedingten Übereinstimmung ist. Obwohl insbesondere die  $\kappa$ -Statistik zur Abbildung der Interrater-Reliabilität der Braden-Items häufig angewendet wurde, ist sie letztendlich dafür nicht geeignet (Kottner & Dassen 2008a). Die Items der Braden-Skala und die Braden-Summenwerte können als intervallskaliert betrachtet werden und somit wurden in dieser Studie parametrische Methoden angewendet. Es wurden Intraklassen-Korrelationskoeffizienten (ICC) (Shrout & Fleiss 1979) für die Items und die Summenwerte berechnet. Zur detaillierten Abbildung der Übereinstimmung für die Braden-Summenwerte wurde ein Bland-Altman-Plot (Bland & Altman 1999) erstellt und zugehörige Übereinstimmungsgrenzen berechnet. Um Aufschlüsse über einen möglichen Zusammenhang zwischen der Höhe der Interrater-Reliabilität und den Charakteristika der Rater zu erhalten, wurden die höchsten und niedrigsten ICC-Korrelationskoeffizienten pro Item und Wohnbereich den Berufsjahren der beteiligten Rater gegenübergestellt.

### Ergebnisse

Die Interrater-Reliabilität ausgedrückt in ICC-Koeffizienten für den gesamten Braden-Wert schwankte je nach Wohnbereich zwischen 0,73 (95% KI 0,26 bis 0,91) und 0,95 (95% KI 0,87 bis 0,98). Die absoluten Differenzen zwischen den beteiligten Ratern schwankten zwischen 0 und maximal 9 Punkten entsprechend der Braden-Skala. 95% der Unterschiede zwischen den Beurteilungen waren nicht größer als 3 Braden-Punkte.

Berechnete Übereinstimmungs- und Interrater-Reliabilitätskoeffizienten für die einzelnen Items variierten stark: Die niedrigsten Übereinstimmungsmaße ergaben sich für die Items „Sensorische Wahrnehmung“ ( $p_o = 20\%$  bis  $67\%$ ,  $ICC(2,1) = 0,16$  (95% KI -0,10 bis 0,49) bis 0,68 (95% KI 0,14 bis 0,88) und „Ernährung“ ( $p_o = 29\%$  bis  $89\%$ ,  $ICC(2,1) = 0,06$  (95% KI -0,31 bis 0,48) bis 0,89 (95% KI 0,75 bis 0,95). Die höchsten Übereinstimmungsmaße ergaben sich für die Items „Aktivität“ ( $p_o = 52\%$  bis  $95\%$ ,  $ICC(2,1) = 0,74$  (95% KI 0,50 bis 0,88) bis 0,97 (95% KI 0,93 bis 0,99) und „Mobilität“ ( $p_o = 47\%$  bis  $83\%$ ,  $ICC(2,1) = 0,68$  (95% KI -0,38 bis 0,85) bis 0,85 (95% KI 0,66 bis 0,94). Ein Zusammenhang zwischen Jahren im Beruf und der Höhe der Interrater-Reliabilität war nicht erkennbar.

## **Methodik und Ergebnisse Studie 2**

### Methodik

Es wurde ein Systematic Review durchgeführt. In den Datenbanken MEDLINE (1985 – Juni 2008), EMBASE (1985 – Juni 2008), CINAHL (1985 – Juni 2008) und im World Wide Web wurde mit Hilfe verschiedener Begriffe für „Dekubitus“ in Kombination mit verschiedenen Begriffen für „Risiko“ und „Reliabilität“ gesucht. Literaturverzeichnisse wurden nach weiteren möglichen relevanten Quellen durchsucht. Einschlusskriterien waren: empirische Untersuchungen zur Interrater-Reliabilität der Waterlow Skala und Englische oder Deutsche Sprache. Daten von Studien, welche die Einschlusskriterien erfüllten, wurden extrahiert. Die Ergebnisse wurden übertragen oder neu berechnet, wenn entsprechende Rohdaten verfügbar waren.

Bis heute gibt es weder Leitlinien oder Standards zur Veröffentlichung von Interrater-Reliabilitätsstudien noch gibt es allgemein anerkannte Richtlinien zur Bewertung der Studienqualität. Basierend auf Forschungsarbeiten, in denen vergleichbare Probleme auftraten (Audigé et al. 2004, Brorson & Hróbjartsson 2008), und aufgrund relevanter Literatur (Dunn 2004, Shoukri 2004) untersuchten wir jede Studie hinsichtlich klarer Methodenbeschreibung, Unabhängigkeit der Beurteilungen, klare Beschreibung der Rater, Personen oder Objekte, Angemessenheit der statistischen Auswertungsmethoden und Klarheit der Ergebnisdarstellung. Weiterhin wurden die Studien danach klassifiziert, ob die untersuchten Rater zur Gruppe klinischer Anwender der Waterlow-Skala (z.B. Pflegende, Assistenten, Helfer) gehören und ob echte Patienten unter echten klinischen Bedingungen untersucht wurden.

Die Datenbank- und Literaturverzeichnissuche, die Datenextraktion und die Qualitätsbeurteilung wurden durch zwei Personen unabhängig voneinander durchgeführt. Die Ergebnisse wurden verglichen und Unstimmigkeiten diskutiert.

### Ergebnisse

Unter Praxisbedingungen wurde die Waterlow-Skala in nur vier Studien untersucht. Prozentuale Übereinstimmungen ( $p_o$ ) für den Waterlow-Summenwert variierten zwischen 0% und 57% (Watkinson 1996, Cook et al. 1999). Unter Einbeziehung von bis zu zwei Punkten Differenz reichten die  $p_o$ -Werte bis maximal 86% (Cook et al. 1999). Interrater-Reliabilitätskoeffizienten wurden bislang nicht berechnet. Die Mediane der Unterschiede waren für die Items „Mobilität“ (2,9 und 3,1), „Schlechte Ernährung“ (2,0 und 4,7) und „Hauttyp“ (3,7) im Vergleich zu den restlichen Items am höchsten was auf beträchtliche Diskrepanzen bei den Beurteilungen hindeutet (Cook et al. 1999, Hale et al. 1999).

Insgesamt lieferte die Datenbanksuche 705 Einträge. Die Anzahl der doppelten Einträge betrug 98 und 581 Studien wurden aufgrund von Titel oder Zusammenfassung ausgeschlossen. Von den verbleibenden 26 Studien erfüllten 6 die Einschlusskriterien. Eine Studie wurde in Literaturverzeichnissen entdeckt und eine Studie im World Wide Web. Alle acht Studien wurden zwischen 1989 und 2005 in Englischer Sprache veröffentlicht. Die Studienqualität war heterogen. Nur zwei Arbeiten (Cook et al. 1999, Hale et al. 1999) erfüllten nahezu alle Qualitätskriterien.

## **Methodik und Ergebnisse Studie 3**

### Methodik

Um die Qualität der Daten der jährlich stattfindenden Dekubitusprävalenzerhebungen zu evaluieren, wurden die teilnehmenden Institutionen stets gefragt, wie viele Pflegekräfte tatsächlich das Hautassessment durchführten. Bei einer ersten Durchsicht dieser Protokolle wurde deutlich, dass sich in jedem Jahr ein Anteil der teilnehmenden Institutionen nicht an das Studienprotokoll hielt, das heißt die Hautinspektion wurde von nur einer Pflegekraft durchgeführt. Um den Einfluss der Anzahl der Rater auf die jährlich ermittelten Dekubitusprävalenzen zu untersuchen, wurden zwei sekundäre Datenanalysen durchgeführt. Die erste Analyse bezog sich auf Pflegeheimdaten (Kottner et al. 2009b),

die zweite auf Krankenhausdaten (Kottner et al. 2009c). Die Trennung zwischen Pflegeheimen und Krankenhäusern schien angemessen, da sich die Charakteristika beider Stichproben erheblich unterscheiden. Das analytische Vorgehen war in beiden Fällen vergleichbar.

Die Variable „Anzahl der Rater (eins oder zwei)“ wurde in die bestehenden Datensätze eingefügt und entsprechend zwei Gruppen gebildet. Bewohner oder Patienten, welche nicht eindeutig einer Gruppe zugeordnet werden konnten, wurden von der weiteren Datenanalyse ausgeschlossen. Beide Gruppen (Gruppe 1: ein Rater und Gruppe 2: zwei Rater) wurden anhand ihrer demographischen Merkmale (Alter, Geschlecht, Body-Mass-Index (BMI)), Dekubitusrisiko (anhand der Braden-Skala) und Dekubitusprävalenz verglichen. Im Falle der Pflegeheimdaten wurden die Gruppen zusätzlich anhand der Pflegeabhängigkeit charakterisiert und verglichen, im Falle der Krankenhausdaten zusätzlich anhand der Anteile von Patienten pro Fachbereich. Die Vergleiche zwischen den beiden Gruppen bezüglich Alter, BMI, Braden- und Pflegeabhängigkeits-Wert erfolgten mit *t*- und *U*-Tests für zwei unabhängige Stichproben. Die Anteile zwischen Geschlecht, Fachgebiet und Dekubitusprävalenz zwischen beiden Gruppen wurden mit  $\chi^2$ - und Fishers exakten Tests verglichen. Ein  $\alpha$ -Level von 0,05 wurde für alle Tests festgelegt.

Der Einfluss der Variable „Anzahl der Rater (eins oder zwei)“ auf die binäre Zielvariable „Dekubitus (ja/nein)“ wurde mit Hilfe logistischer Regressionen durchgeführt. Um Verzerrungen zu vermeiden, wurden die demographischen Variablen Geschlecht, Alter, BMI und Dekubitusrisiko mit berücksichtigt.

Es kann davon ausgegangen werden, dass die Diagnose eines Dekubitus Grad 1 schwierig ist (Bethell 2003, Defloor et al. 2006) und entsprechend einen hohen Fehleranteil aufweist (Kottner und Dassen 2008b). Somit wurden alle Vergleiche zwischen Prävalenzen und die logistischen Regressionen getrennt für Dekubitus Grad 1, 2, 3 oder 4 und Dekubitus Grad 2, 3 oder 4 durchgeführt.

### Ergebnisse

Die Ergebnisse deuten darauf hin, dass die Anzahl der Pflegekräfte, welche die Hautinspektion vornahmen, keinen Einfluss auf die beobachtete Dekubitushäufigkeit hatten. In den Pflegeheimen waren die Gruppen mit einem Rater ( $n = 1269$ ) und zwei Ratern ( $n = 4224$ ) hinsichtlich Alter, Geschlecht, BMI, Pflegeabhängigkeit und Dekubitusrisiko sehr gut vergleichbar. Die beobachteten Dekubitusprävalenzen in der Gruppe mit einem Rater betrug 6,2% für Dekubitus Grad 1, 2, 3 oder 4 und 3,9% für Dekubitus Grad 2, 3 oder 4. In der Gruppe mit zwei Ratern betrug die Prävalenz 6,7% Dekubitus Grad 1, 2, 3 oder 4 und 4,2% für Dekubitus Grad 2, 3 oder 4. In beiden Fällen waren die Unterschiede statistisch nicht signifikant ( $p > 0,05$ ). Die adjustierten Odds Ratios für einen Rater im Vergleich zu zwei Ratern mindestens einen Bewohner mit Dekubitus zu erkennen waren 0,9 (95% KI 0,7 bis 1,2) für Grad 1, 2, 3 oder 4 und 0,9 (95% KI 0,7 bis 1,3) für Grad 2, 3 oder 4. Die einzige Variable, welche im multivariaten Modell die Diagnose „Dekubitus (ja/nein)“ erklären konnte war der Braden-Summenwert (Odds Ratio = 0,8 (95% KI 0,8 bis 0,9) für Grad 1, 2, 3 oder 4 und 0,9 (95% KI 0,8 bis 0,9) für Grad 2, 3 oder 4. Demnach sinkt die Wahrscheinlichkeit, dass bei einem Bewohner mindestens ein Dekubitus erkannt wird mit steigendem Braden-Wert.

Die Gruppen der Krankenhauspatienten mit einem ( $n = 2420$ ) und zwei ( $n = 15009$ ) Ratern waren hinsichtlich Geschlecht, Alter, BMI und Dekubitusrisiko vergleichbar. Die Prävalenzen für Dekubitus Grad 1, 2, 3 oder 4 betragen 8,5% (Gruppe mit einem Rater) und 9,3% (Gruppe mit zwei Ratern). Die Prävalenzen für Dekubitus Grad 2, 3 oder 4 betragen 5,3% (Gruppe mit einem Rater) und 5,0% (Gruppe mit zwei Ratern). In beiden Fällen waren die Unterschiede statistisch nicht signifikant ( $p > 0,05$ ). Das Odds Ratio, dass ein Rater im Vergleich zu zwei Ratern einen Patienten mit einem Dekubitus Grad 1, 2, 3 oder 4 identifiziert war 1,2 (95% KI 1,0 bis 1,4). Unter Ausschluss von Grad 1 betrug das Odds Ratio 1,4 (95% KI 1,2 bis 1,8). Eine Variable, welche in beiden Fällen einen signifikanten Einfluss auf die Zielvariable „Dekubitus (ja/nein)“ hatte war der Braden-Summenwert. Mit steigenden Braden-Summenwerten sank die Wahrscheinlichkeit, dass bei einem Patienten mindestens ein Dekubitus diagnostiziert wurde (0,8 pro Braden-Punkt (95% KI 0,8 bis 0,8)).

## **Methodik und Ergebnisse Studie 4**

### Methodik

Es wurde ein Systematic Review durchgeführt. Die Datenbanken MEDLINE (1965 - Juni 2007), EMBASE (1989 - Juni 2007) und CINAHL (1995 - Juni 2007) wurden mit der Software WebSPIRS (Ovid Technologies) systematisch durchsucht. Zur Suche wurden Kombinationen verschiedener Begriffe für Dekubitus und Klassifikationen verwendet. Literaturverzeichnisse wurden auf weitere relevante Quellen durchsucht. Mit Hilfe von Google wurde das World Wide Web durchsucht. Autoren von Kongressbeiträgen, welche im Internet zu finden waren, wurden für detaillierte Informationen kontaktiert. Einschlusskriterien für die Suche waren: primäre Studien, in denen Interrater-Reliabilitäten bestimmt wurden und Englische oder Deutsche Sprache. Ausschlusskriterien waren: retrospektive Studiendesigns, Reviews, Diskussionsbeiträge und Studien, in denen nicht auf primäre Daten zurückgegriffen wurde (z.B. Krankenakten). Daten relevanter Studien wurden in eine Datenextraktionstabelle übertragen. Interrater-Reliabilitätskoeffizienten und zugehörige 95% Konfidenzintervalle wurden übernommen oder neu berechnet, falls auf entsprechende Rohwerte zurückgegriffen werden konnte.

Eine intensive Suche nach Instrumenten zur Qualitätsbewertung von Interrater-Reliabilitätsstudien war erfolglos. Es gibt bis heute keine entsprechenden Leitlinien oder Kriterien. Basierend auf Forschungsarbeiten anderer Autoren (Audigé et al. 2004, Brorson & Hróbjartsson 2008), welche vor demselben Problem standen, und basierend auf bekannten Faktoren, welche die Interrater-Reliabilität beeinflussen (Dunn 2004), wurden neun Qualitätskriterien formuliert:

- (1) Wurden die Klassifikationssysteme und die verwendete Anzahl von Kategorien klar beschrieben?
- (2) Wurden die wiederholten Assessments unabhängig voneinander durchgeführt?
- (3) Wurde die Anzahl der beurteilten Personen/Objekte genannt?
- (4) Wurden die Assessmentbedingungen (z.B. echtes Hautassessment, Web-basierte Darbietung von Abbildungen) klar beschrieben?
- (5) Wurde die Anzahl der Rater genannt?

- (6) Wurden die Eigenschaften der Rater beschrieben?
- (7) Wurde die Anzahl der Ratings pro Rater beschrieben?
- (8) Wurde beschrieben, ob „normale“ Haut untersucht wurde?
- (9) Waren die Berechnungen der Interrater-Reliabilität nachvollziehbar und verständlich?

Jede relevante Studie wurde anhand der neun Kriterien beurteilt. Studien die sieben und mehr Kriterien erfüllten, wurden in die endgültige Datensynthese aufgenommen. Die Identifikation potentiell relevanter Studien in den Datenbanken, die Datenextraktion und die Qualitätsbeurteilung erfolgten durch zwei Reviewer unabhängig voneinander. Die Ergebnisse wurden verglichen und Diskrepanzen geklärt.

### Ergebnisse

Aufgrund der derzeitigen Studienlage scheint es unmöglich die Dekubitusklassifikation mit der höchsten Interrater-Reliabilität zu bestimmen. Von 339 potentiell relevanten Studien erfüllten 47 die Einschlusskriterien. Davon wurden die meisten Studien in den USA ( $n = 16$ ) und in Großbritannien ( $n = 9$ ) durchgeführt. Nach Durchführung des Qualitätsassessments wurden 24 Studien (51%) in die endgültige Datensynthese aufgenommen.

Interrater-Reliabilitäten auf der Grundlage eines echten Hautassessments wurden in 10 Studien bestimmt.  $\kappa$ -Koeffizienten für das Klassifikationssystem des EPUAP variierten zwischen 0,97 (95% KI 0,92-1,00) (Bours et al. 1999) und 0,31 (Pedley 2004). Zugehörige prozentuale Übereinstimmungen ( $p_o$ ) variierten zwischen 100% (Bours et al. 1999) und 49% (Pedley 2004).  $p_o$ -Werte für das Klassifikationssystem des National Pressure Ulcer Advisory Panel (NPUAP) (1989) aus den USA reichten von 100% (Gawron 1994) bis zu 58% (Lyder et al. 1999).  $\kappa$ -Koeffizienten wurden nicht berechnet.

Interrater-Reliabilitäten auf der Grundlage von Abbildungen und Fotos wurden in 14 Studien bestimmt.  $p_o$ -Werte für die NPUAP-Klassifikation reichten von 39% (Buckley et al. 2005) über 70% (Arnold & Watterworth 1995) und 90% (Noonan et al. 2006) bis zu 100% (Buckley et al. 2005).  $\kappa$ -Werte schwanken zwischen 0,56 (Hart et al. 2006) und 0,75 (Groeneveld et al. 2004). Die Interrater-Reliabilität zwischen Dekubitusexperten welche 56 Fotografien beurteilten betrug  $\kappa = 0,80$  (Defloor & Schoonhoven 2004). Die Interrater-Reliabilität für das gleiche Bilderset zwischen über 500 Pflegenden war bedeutend geringer:  $\kappa = 0,40$  (Defloor et al. 2006).

Ungeachtet der durchgeführten Vergleiche und systematischen Darstellungen der Studienergebnisse war eine sinnvolle Datensynthese aufgrund der Heterogenität der Studien nahezu unmöglich. Die Studien wiesen hinsichtlich der Qualität, der Assessment-Methoden, der Qualifikation und Eigenschaften der Rater, der verwendeten Klassifikationssysteme und der Stichprobeneigenschaften große Unterschiede auf.

### **Diskussion**

Die Interrater-Reliabilität für den Braden-Summenwert kann als hoch angesehen werden. Die Ratings nach der Beurteilung der Pflegeheimbewohner waren identisch oder wichen mehrheitlich um maximal

3, auf der Skala von 6 bis 23 möglichen Punkten, ab. Bei einer vergleichbaren Stichprobe von Ratern und beurteilten Personen wäre somit eine Abweichung von bis zu drei Punkten zu erwarten. Vor dem Hintergrund der Spannweite möglicher Differenzen (17 Punkte) ist die klinische Relevanz dieser Abweichung gering einzuschätzen. Dennoch hat dieses Ergebnis für die klinische Praxis enorme Bedeutung, denn es stellt die Absolutheit verwendeter Cut-Off-Punkte und Risikogruppenbildungen infrage. Beispielsweise ist es international üblich, Personen als dekubitusgefährdet zu betrachten, wenn sie einen Braden-Wert  $\leq 18$  haben (Ayello & Braden 2002). Bei einer zu erwartenden Präzision von  $\pm 3$  ist es fraglich, ob beispielsweise Personen mit dem Braden-Wert 19 als nicht gefährdet zu betrachten sind. Cut-Off-Punkte täuschen eine Genauigkeit vor, die es in der Praxis nicht gibt. Besonders riskant sind Kategorisierungen von Patienten, wenn daran bestimmte Verfahrensanweisungen gekoppelt sind. Sieht beispielsweise eine einrichtungsinterne Versorgungsrichtlinie vor, Patienten mit Braden-Werten  $\leq 12$ , was einem „hohen Risiko“ (Ayello & Braden 2002) entspricht, auf speziellen Antidekubitusmatratzen zu lagern, so stellt sich die Frage was mit Patienten ist, bei denen ein Braden-Wert von 13 ermittelt wurde.

Unabhängig von den eher geringen Differenzen wurden in Einzelfällen Abweichungen bis zu maximal 9 Punkten gemessen. Diese Unterschiede sind beträchtlich und stellen den klinischen Wert dieser Summenwerte grundsätzlich infrage. Die Gründe für diese großen Abweichungen konnten mit der durchgeführten Untersuchung nicht eindeutig festgestellt werden. Alle Rater hatten einen vergleichbaren Ausbildungs- und Schulungsstand. Zwischen den Jahren im Beruf und der Interrater-Reliabilität war kein Zusammenhang erkennbar. Die Bewohner waren hinsichtlich ihrer demographischen Merkmale ebenfalls vergleichbar. Es konnte jedoch festgestellt werden, dass die Items „Sensorische Wahrnehmung“ und „Ernährung“ den höchsten Fehleranteil aufwiesen. Der Ausschluss dieser Items oder eine bessere Operationalisierung und klarere Beschreibung könnte die Interrater-Reliabilität des Braden-Summenwertes erhöhen.

Die Waterlow-Skala wurde in der klinischen Praxis bislang kaum untersucht. Bezogen auf den Waterlow-Summenwert sind die erzielten Übereinstimmungen zwischen verschiedenen Ratern mit denen der Braden-Skala vergleichbar. Die Studien deuten darauf hin, dass die Items „Schlechte Ernährung“, „Hauttyp“ und „Mobilität“ zu besonders unpräzisen Einschätzungen führen. Da bislang keine Interrater-Reliabilitätskoeffizienten bestimmt wurden, können keine Aussagen über den zu erwartenden Fehleranteil in der klinischen Praxis gemacht werden. Das steht im Widerspruch zu der weiten Verbreitung und Anwendung der Skala. Solange über die Verlässlichkeit der Waterlow-Summenwerte nichts bekannt ist, sollten klinische Entscheidungen aufgrund dieser Werte unter Vorbehalt getroffen werden. So gelten beispielsweise nach der „Verfahrensregelung Dekubitusphrophylaxe“ der Charité – Universitätsmedizin Berlin (2008) Patienten dann als „Risikopatienten“, wenn sie einen Waterlow-Summenwert  $> 9$  Punkte aufweisen. Vergleichbar mit der Braden-Skala, ist es sehr wahrscheinlich, dass diese Einschätzungsgenauigkeit in der Versorgungspraxis nicht erreicht werden kann. Daraus ergibt sich die Forderung, die Interrater-

Reliabilität der Waterlow-Skala dringend zu untersuchen. Erst dann können Empfehlungen für deren Nutzen, Anwendung oder Modifikation ausgesprochen werden.

Basierend auf den Daten der jährlich stattfindenden deutschlandweiten Prävalenzerhebungen scheinen die Anzahl der Pflegekräfte, welche die Hautinspektion durchführen, die gemessene Dekubitusprävalenz nicht zu beeinflussen. Weiterhin konnten keine Hinweise gefunden werden, dass die Anzahl der Pflegekräfte die Wahrscheinlichkeit, Pflegeheimbewohner oder Krankenhauspatienten mit Dekubitus zu erkennen, beeinflusst. Daraus folgt zum einem, dass die jährlichen Dekubitusprävalenzschätzungen durch die wechselnde Anzahl der beteiligten Pflegekräfte nicht verzerrt sind. Zum anderen kann die Empfehlung der EPUAP (2002, 2005) mit den vorliegenden Daten nicht gestützt werden. Dekubitusprävalenz- oder –inzidenzmessungen sind personal- und kostenintensiv. Wenn zwischen den Assessments zwischen einer einzelnen Pflegekraft und einem Team bestehend aus zwei Pflegekräften kein Unterschied besteht, könnte man in zukünftigen Studien auf eine zweite Pflegekraft verzichten. Diese Schlussfolgerung müsste jedoch in zukünftigen Studien weiter empirisch untermauert werden, da aufgrund des Studiendesigns der Einfluss der Charakteristiken der Rater nicht untersucht werden konnte.

Aufgrund der derzeitigen Studienlage scheint es unmöglich, für die klinische Praxis eine bestimmte Dekubitusklassifikation zu empfehlen. Bisher durchgeführte Untersuchungen sind kaum vergleichbar. Besonders problematisch ist, dass mehr qualitativ hochwertige Studien unter künstlichen Bedingungen auf der Grundlage von Abbildungen oder Fallbeispielen als unter klinischen Bedingungen durchgeführt wurden. Über die Vergleichbarkeit beider Assessmentsituationen oder die Übertragbarkeit der Ergebnisse ist nichts bekannt. Ungeachtet der Heterogenität der Studienlage wurden Klassifikationen des NPUAP und des EPUAP am häufigsten untersucht. Die vorliegenden Daten deuten darauf hin, dass die Interrater-Reliabilitäten für beide Klassifikationen, wenn von Pflegekräften angewendet, eher gering ist. Das stellt die Güte und Verlässlichkeit der Dekubitusklassifikation und Dokumentation infrage. Darüber hinaus kann auch die Empfehlung des „Expertenstandards Pflege von Menschen mit chronischen Wunden“ (Deutsches Netzwerk für Qualitätsentwicklung in der Pflege 2008), das EPUAP-System zu verwenden, nicht gestützt werden. Es fehlen Interrater-Reliabilitätsstudien, in denen vergleichbare Rater verschiedene Klassifikationssysteme unter vergleichbaren Bedingungen anwenden. Darüber hinaus sollte untersucht werden, wie die diagnostische Präzision bei der Klassifikation von Dekubitus in der Praxis verbessert werden kann.

## Literatur

- Arnold, N., & Watterworth, B. (1995). Wound staging: can nurses apply classroom education to the clinical setting? *Ostomy Wound Management*, 41, 40-44.
- Audigé, L., Bhandari, M., Hanson, B., & Kellam, J. (2005). A concept for the validation of fracture classifications. *Journal of Orthopaedic Trauma*, 19, 401-406.
- Ayello, E.A., & Braden, B. (2002). How and why to do pressure ulcer risk assessment. *Advances in Skin & Wound Care*, 15, 125-133.
- Bartholomeyczik, S. (2007). Einige kritische Anmerkungen zu standardisierten Assessmentinstrumenten in der Pflege. *Pflege*, 20, 211-217.
- Bergstrom, N., Braden, B.J., Laguzza, A., & Holman, V. (1987). The Braden scale for predicting pressure sore risk. *Nursing Research*, 36, 205-210.
- Bethell, E. (2003). Controversies in classifying and assessing grade I pressure ulcers. *Journal of Wound Care*, 12, 33-36.
- Bland, J.M., & Altman, D.G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8, 135-160.
- Bours, G.J.J.W., Halfens, R.J.G., Lubbers, M., & Haalboom, J.R.E. (1999). The development of a National Registration Form to measure the prevalence of pressure ulcers in the Netherlands. *Ostomy Wound Management*, 45, 28-40.
- Brorson, S., & Hróbjartsson, A. (2008). Training improves agreement among doctors using the Neer system for proximal humeral fractures in a systematic review. *Journal of Clinical Epidemiology*, 61, 7-16.
- Buckley, K.M., Tran, B.Q., Adelson, L.K., Agazio, J.G., & Halstead, L. (2005). The use of digital impanges in evaluating homecare nurses' knowledge of wound assessment. *Journal of Wound Ostomy and Continence Nursing*, 32, 307-316.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Compton, F., Strauß, M., Hortig, T., Frey, J., Hoffmann, F., Zidek, W., & Schäfer, J.-H. (2008). Validität der Waterlow-Skala zur Dekubitusrisikoeinschätzung auf der Intensivstation: eine prospektive Untersuchung an 698 Patienten. *Pflege*, 21, 37-48.
- Cook, M., Hale, C., & Watson, B. (1999). Interrater reliability and the assessment of pressure-sore risk using an adapted Waterlow Scale. *Clinical Effectiveness in Nursing*, 3, 66-74.
- Dassen, T., Lahmann, N., Heinze, C., Kottner, J., Mertens, E., Schmitz, G., Tannen, A., Wilborn, D., Kuntz, S., Kutz, F., Lützkendorf, D., Pöhler, A., Raeder, K., & Schröer, F. (2008). *Prävalenz Erhebung 2008*. Berlin: Charité – Universitätsmedizin Berlin, Zentrum für Human- und Gesundheitswissenschaften, Institut für Medizin-/Pflegepädagogik und Pflegewissenschaft.
- Defloor, T., & Schoonhoven, L. (2004). Inter-rater reliability of the EPUAP pressure ulcer classification system using photographs. *Journal of Clinical Nursing*, 13, 952-959.

- Defloor, T., Schoonhoven, L., Vanderwee, K., Weststrate, J., & Myny, D. (2006). Reliability of the European Pressure Ulcer Advisory Panel classification system. *Journal of Advanced Nursing*, 54, 189-198.
- Deutsches Netzwerk für Qualitätsentwicklung in der Pflege (Hrsg.). (2004). *Expertenstandard Dekubitusprophylaxe in der Pflege* (2. Aufl.). Osnabrück.
- Deutsches Netzwerk für Qualitätsentwicklung in der Pflege (Hrsg.). (2008). *Expertenstandard Pflege von Menschen mit chronischen Wunden*. Osnabrück.
- Dunn, G. (2004). *Statistical Evaluation of Measurement Errors: Design and Analysis of Reliability Studies* (2nd ed.). London: Arnold.
- Eberhardt, S., Heinemann, W., Kulp, W., et al. (2005). *Dekubitusprophylaxe und –therapie*. Deutsche Agentur für Health Technology Assessment des Deutschen Instituts für Medizinische Dokumentation und Information. Zugriff von [http://gripsdb.dimdi.de/de/hta/hta\\_berichte/hta128\\_bericht\\_de.pdf](http://gripsdb.dimdi.de/de/hta/hta_berichte/hta128_bericht_de.pdf) am 30. Oktober 2008.
- European Pressure Ulcer Advisory Panel (2002). Draft EPUAP statement on prevalence and incidence monitoring. *EPUAP Review*, 4, 13-15.
- European Pressure Ulcer Advisory Panel (2005). EPUAP statement on prevalence and incidence monitoring of pressure ulcer occurrence 2005. *EPUAP Review*, 6, 74-80.
- European Pressure Ulcer Advisory Panel (1998). *Pressure ulcer treatment guidelines*. Zugriff von <http://www.epuap.org/gltreatment.html> am 10. Juni 2008
- Gawron, C.L. (1994). Risk factors for and prevalence of pressure ulcers among hospitalized patients. *Journal of Wound Ostomy and Continence Nursing*, 21, 232-240.
- Groeneveld, A., Anderson, M., Allen, S., Bressmer, S., Golberg, M., Magee, B. & al. (2004). The prevalence of pressure ulcers in a tertiary care pediatric and adult hospital. *Journal of Wound Ostomy and Continence Nursing*, 31, 108-120.
- Hale, C., Cook, M., & Watson, B. (1999). Assessing the interrater reliability of the Waterlow pressure sore risk assessment scale. Unpublished report. Zugriff von [http://www.personal.leeds.ac.uk/~hcscah/pdfs/waterlow\\_report.pdf](http://www.personal.leeds.ac.uk/~hcscah/pdfs/waterlow_report.pdf) am 30. Oktober 2008,
- Hart, S., Bergquist, S., Gajewski, B., & Dunton, N. (2006). Reliability testing of the national database of nursing quality indicators pressure ulcer indicator. *Journal of Nursing Care Quality*, 21, 256-265.
- Kottner, J., & Dassen, T. (2008a). Interpreting interrater reliability coefficients of the Braden Scale: A discussion paper. *International Journal of Nursing Studies*, 45, 1238-1246.
- Kottner, J., & Dassen, T. (2008b). Dekubitusprävalenzmessungen und Interrater-Reliabilität. *Pflegewissenschaft*, 10, 499-503.
- Kottner, J., Dassen, T., & Tannen, A. (2009a). Inter- and intrarater reliability of the Waterlow pressure sore risk scale: A systematic review. *International Journal of Nursing Studies*, 46, 369-379.
- Kottner, J., Tannen, A., Halfens, R., & Dassen, T. (2009b). Does the number of raters influence the pressure ulcer prevalence rate? *Applied Nursing Research*, 22, 68-72.

- Kottner, J., Tannen, A., & Dassen, T. (2009c). Hospital pressure ulcer prevalence rates and number of raters. *Journal of Clinical Nursing*, doi: 10.1111/j.1365-2702.2008.02609.x
- Lyder, C.H., Yu, C., Emerling, J., Mangat, R., Stevenson, D., Empleo-Frazier, O., & McKay, J. (1999). The Braden scale for pressure ulcer risk: evaluating the predictive validity in Black and Latino/Hispanic elders. *Applied Nursing Research*, 12, 60-68.
- National Pressure Ulcer Advisory Panel. (1989). Pressure ulcers prevalence cost and risk assessment: Consensus development conference statement. *Decubitus*, 2, 24-28.
- Noonan, C., Quigley, S., & Curley, M.A.Q. (2006). Skin integrity in hospitalized infants and children: a prevalence study. *Journal of Pediatric Nursing*, 21, 445-453.
- Pancorbo-Hidalgo, P.L., Garcia-Fernandez, F.P., Lopez-Medina, I.M., & Alvarez-Nieto, C. (2006). Risk assessment scales for pressure ulcer prevention: a systematic review. *Journal of Advanced Nursing*, 54, 94-110.
- Pedley, G.E. (2004). Comparison of pressure ulcer grading scales: a study of clinical utility and inter-rater reliability. *International Journal of Nursing Studies*, 41, 129-140.
- Polit, D.F. & Beck, C.T. (2008). *Nursing Research: Principles and Methods*. Wolters Kluwer and Lippincott Williams & Wilkins, Philadelphia, 8th ed.
- Püllen, R. (2007). Dekubitalulzera im höheren Lebensalter. *Klinikerarzt*, 36, 143-148
- Schlömer, G. (2003). Dekubitusrisikoskalen als Screeninginstrumente – Ein systematischer Überblick externer Evidenz. *Zeitschrift für ärztliche Fortbildung und Qualitätssicherung*, 97, 33-46.
- Shoukri, M.M. (2004). *Measures of Interobserver Agreement*. Chapman & Hall/CRC, Boca Raton.
- Shrout, P.E. & Fleiss, J.L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Szklo, M. & Nieto, F.J. (2007). *Epidemiology Beyond the Basics*. Jones and Bartlett Publishers, Sudbury, Massachusetts, 2nd ed.
- Verfahrensregelung Dekubitusprophylaxe*. (2008). Charité – Universitätsmedizin Berlin. Zugriff von [http://www.charite.de/qualitaetsmanagement/downloads/verfahren/VR\\_Dekubitusprophylaxe\\_Revision1.pdf](http://www.charite.de/qualitaetsmanagement/downloads/verfahren/VR_Dekubitusprophylaxe_Revision1.pdf) am 30. Oktober 2008
- Von Renteln-Kruse, W., Krause, T., Anders, J., Heinemann, A., & Püschel, K. (2004). Höhergradige Dekubitalulzera bei betagten pflegebedürftigen Hochrisiko-Personen. *Zeitschrift für Gerontologie und Geriatrie*, 37, 81-85.
- Waterlow, J. (1985). A risk assessment card. *Nursing Times*, 81, 49-55.
- Watkinson, C. (1996). Inter-rater reliability of risk-assessment scales. *Professional Nurse*, 11, 751-756.

## Erklärung über den Anteil des Promovenden an Publikationen

Publikation 1: **Kottner, J.,** & Dassen, T. (2008). An interrater reliability study of the Braden scale in two nursing homes.

90 Prozent

Beitrag im Einzelnen: Aufstellung der Forschungsfrage, Literaturrecherche, Festlegung von Designs und Methode, praktische Durchführung (Schulung, Datensammlung), Datenanalyse, Gestaltung und Schreiben des Manuskripts.

Publikation 2: **Kottner, J.,** Dassen, T., & Tannen, A. (2009). Inter- and intrarater reliability of the Waterlow pressure sore risk scale: A systematic review.

80 Prozent

Beitrag im Einzelnen: Aufstellung der Forschungsfrage, Festlegung des Designs, Durchführung der systematischen Recherche, Erstellung des Bewertungsinstruments, Bewertung des Studien, Datenextraktion, -analyse und Ergebnissynthese, Gestaltung und Schreiben des Manuskripts.

Publikation 3: **Kottner, J.,** Tannen, A., Halfens, R., & Dassen, T. (2009). Does the number of raters influence the pressure ulcer prevalence rate?

80 Prozent

Beitrag im Einzelnen: Aufstellung der Forschungsfrage, Literaturrecherche, Design, Vorbereitung und Zusammenfassung der Daten, Einpflegen der neuen Variablen, Datenanalyse, Gestaltung und Schreiben des Manuskripts.

Publikation 4: **Kottner, J.,** Tannen, A., & Dassen, T. (in press). Hospital pressure ulcer prevalence rates and number of raters.

80 Prozent

Beitrag im Einzelnen: Aufstellung der Forschungsfrage, Literaturrecherche, Design, Vorbereitung und Zusammenfassung der Daten, Einpflegen der neuen Variablen, Datenanalyse, Gestaltung und Schreiben des Manuskripts.

Publikation 5: **Kottner, J.,** Raeder, K., Halfens, R., & Dassen, T. (2009). A systematic review of interrater reliability of pressure ulcer classification systems. *Journal of Clinical Nursing*.

80 Prozent

Beitrag im Einzelnen: Aufstellung der Forschungsfrage, Festlegung des Designs, Durchführung der systematischen Recherche, Erstellung des Bewertungsinstruments, Datenextraktion, -analyse und Ergebnissynthese, Gestaltung und Schreiben des Manuskripts.

---

Jan Kottner (Antragsteller)

---

Prof. Dr. Theo Dassen (Betreuer)

## **Publikation 1**

Kottner, J., & Dassen, T. (2008). An interrater reliability study of the Braden scale in two nursing homes. *International Journal of Nursing Studies*, 45, 1501-1511.

## **Publikation 2**

Kottner, J., Dassen, T., & Tannen, A. (2009). Inter- and intrarater reliability of the Waterlow pressure sore risk scale: A systematic review. *International Journal of Nursing Studie*, *46*, 369-379.

### **Publikation 3**

Kottner, J., Tannen, A., Halfens, R., & Dassen, T. (2009). Does the number of raters influence the pressure ulcer prevalence rate? *Applied Nursing Research*, 22, 68-72.

#### **Publikation 4**

Kottner, J., Tannen, A., & Dassen, T. (2009). Hospital pressure ulcer prevalence rates and number of raters. *Journal of Clinical Nursing*. doi: 10.1111/j.1365-2702.2008.02609.x

## **Publikation 5**

Kottner, J., Raeder, K., Halfens, R., & Dassen, T. (2009). A systematic review of interrater reliability of pressure ulcer classification systems. *Journal of Clinical Nursing, 18*, 315-336.

## **Lebenslauf**

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

## Vollständige Publikationsliste

- Kottner, J.,** Dassen, T., Heinze, C. (in press). Diagnose und Klassifikation von Dekubitus und anderen Hautschäden: Interrater-Reliabilität und Übereinstimmung. *Pflegezeitschrift*.
- Kottner, J.,** Dassen, T., & Lahmann, N. (in press). Comparison of two skin examination methods for Grade 1 pressure ulcer identification. *Journal of Clinical Nursing*.
- Kottner, J.,** Wilborn, D., Dassen, T., Lahmann, N. (in press). The trend of pressure ulcer prevalence rates in German hospitals: Results of seven cross-sectional studies. *Journal of Tissue Viability*.
- Kottner, J.,** Tannen, A., & Dassen, T. (in press). Hospital pressure ulcer prevalence rates and number of raters. *Journal of Clinical Nursing*.
- Kottner, J.,** Dassen, T., & Tannen, A. (2009). Inter- and intrarater reliability of the Waterlow pressure sore risk scale: A systematic review. *International Journal of Nursing Studies, 46,* 369-379.
- Kottner, J.** (2009). Commentary on Sayar S et al. (2009) Incidence of pressure ulcers in intensive care unit patients at risk according to the Waterlow scale and factors influencing the development of pressure ulcers. *Journal of Clinical Nursing, 18,* 776-777.
- Kottner, J.,** Tannen, A., Halfens, R., & Dassen, T. (2009). Does the number of raters influence the pressure ulcer prevalence rate? *Applied Nursing Research, 22,* 68-72.
- Kottner, J.,** Raeder, K., Halfens, R., & Dassen, T. (2009). A systematic review of interrater reliability of pressure ulcer classification systems. *Journal of Clinical Nursing, 18,* 315-336.
- Kottner, J.,** Lahmann, N., Wilborn, D., & Dassen, T. (2009). Ein Vergleich der Dekubitushäufigkeiten zwischen 37 Pflegeheimen: Pflegequalität sichtbar machen. *Pflegezeitschrift, 62,* 34-35.
- Kottner, J.** (2009). Interrater reliability and the kappa statistic: A comment on Morris et al. *International Journal of Nursing Studies, 46,* 140-141.
- Dassen, T., Lahmann, N., Heinze, C., **Kottner, J.,** Mertens, E., Schmitz, G., Tannen, A., Wilborn, D., Kuntz, S., Kutz, F., Lützkendorf, D., Pöhler, A., Raeder, K., & Schröer, F. (2008). *Prävalenz Erhebung 2008*. Berlin: Charité – Universitätsmedizin Berlin, Zentrum für Human- und Gesundheitswissenschaften, Institut für Medizin-/Pflegepädagogik und Pflegewissenschaft.
- Kottner, J.,** & Dassen, T. (2008). An interrater reliability study of the Braden scale in two nursing homes. *International Journal of Nursing Studies, 45,* 1501-1511.
- Kottner, J.,** & Dassen, T. (2008). Dekubitusprävalenzmessungen und Interrater-Reliabilität. *Pflegewissenschaft, 10,* 499-503.
- Hoppe, C., Pöhler, A., **Kottner, J.,** & Dassen, T. (2008). Dekubitus: Neue Daten zu Risiko, Prävalenz und Entstehungsorten. *Pflegezeitschrift, 61,* 90-93.
- Kottner, J.,** & Dassen, T. (2008). Interpreting interrater reliability coefficients of the Braden Scale: A discussion paper. *International Journal of Nursing Studies, 45,* 1238-1246.
- Kottner, J.,** Tannen, A., & Dassen, T. (2008). Die Interrater-Reliabilität der Braden-Skala. *Pflege, 21,* 85-94.
- Kottner, J.,** & Wilborn, D. (2008). Dekubitusprophylaxe: Verfügbare wissenschaftliche Erkenntnisse nutzen. *Pflegezeitschrift, 61,* 433-437.
- Dassen, T., Heinze, C., **Kottner, J.,** Lahmann, N., Mertens, E., Schmitz, G., Tannen, A., Wilborn, D., Amend, C., Kuntz, S., Kutz, F., Lützkendorf, D., Pöhler, A., Raeder, K., & Siegmund, U. (2007). *Prävalenz Erhebung 2007*. Berlin: Charité – Universitätsmedizin Berlin, Zentrum für Human- und Gesundheitswissenschaften, Institut für Medizin-/Pflegepädagogik und Pflegewissenschaft.

- Dassen, T., & **Kottner, J.** (2007). Pflegewissenschaft im deutschsprachigen Raum – ein Positionspapier. *Pflegezeitschrift*, 60, 96-100.
- Kottner, J.**, Mertens, E., & Dassen T. (2007). Dekubitusprävalenz in Deutschland: Ergebnisse einer Querschnittsstudie 2006. *Pflegezeitschrift*, 60, 28-31.
- Kottner, J.**, Schmitz, G., & Dassen, T. (2007). Daten aus der Wissenschaft: Prävalenzerhebung lässt Pflegeprobleme erkennen. *Die Schwester Der Pfleger*, 46, 172-175.
- Kottner, J.**, & Wilborn, D. (2007). Darstellung der Studie kritisch zu bewerten [Leserbrief]. *Die Schwester Der Pfleger*, 46, 858.
- Dassen, T., Heinze, C., Lahmann, N., Lohrmann, C., Mertens, E., Tannen, A., Eichhorn, J., **Kottner, J.**, Raeder, K., & Siegmund, U. (2006). *Prävalenz Erhebung 2006*. Berlin: Charité – Universitätsmedizin Berlin, Zentrum für Human- und Gesundheitswissenschaften, Institut für Medizin-/Pflegepädagogik und Pflegewissenschaft.
- Kottner, J.**, Helberg, D., Mertens, E., Halfens, R., & Dassen, T. (2006). Dekubitusbehandlung: Vergleich zwischen Theorie und Praxis. *Pflegezeitschrift*, 59, 488-492.
- Kottner, J.**, Mertens, E., & Dassen, T. (2006). Dekubitusprävalenz in Deutschland: Neue Daten, neue Erkenntnisse. *Pflegezeitschrift*, 59, 30-33.

## **Erklärung**

„Ich, Jan Kottner, erkläre, dass ich die vorgelegte Dissertationsschrift mit dem Thema: Interrater-Reliabilität von Instrumenten zur Einschätzung des Dekubitusrisikos und des Dekubitus selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, ohne die (unzulässige) Hilfe Dritter verfasst und auch in Teilen keine Kopien anderer Arbeiten dargestellt habe.“

Berlin, 24. März 2009

Jan Kottner