

# Quantitative Methoden – Grundlage für komplexes Handeln?<sup>1/2</sup>

Eva-Maria Panfil

■ In diesem Beitrag der Heidelberger Fachtagung des DV Pflegewissenschaft im September 2003 werden Stärken und Grenzen quantitativer Forschungsmethoden und Studiendesigns diskutiert. Zu vier Punkten werden Thesen entwickelt: zur Komplexitätsreduktion, zum Wesen von RCTs, zum Verhältnis von Wahrheit und Wahrscheinlichkeit sowie zu den Grenzen der Beforschbarkeit von Fragestellungen.

■ This presentation at the Heidelberg-Congress of DV Pflegewissenschaft in september 2003 discusses strengthes and limits of quantitative research methods and studydesigns. Four issues are presented: the reduction of complexity, the character of RCT, the relationship of truth and probability, and the limits of research itself.

Pflege ist gefordert, den Pflegeempfängern eine auf dem aktuellen Stand des Wissens basierende Pflege anzubieten. Die heute aktuelle Methodik zur wissenschaftlichen Basierung der Pflegepraxis wird unter dem Begriff »Evidence-based-Nursing (EBN)« diskutiert. Evidenzbasierte Pflegepraxis meint »den Einbezug von auf Forschung basierender Evidenz, klinischer Expertise und Patientenpräferenzen in Entscheidungen über die Gesundheitsversorgung von individuellen Patienten« (Mulhall 1998, S. 5).

Pflegeforschung als Basis für die Praxisdisziplin Pflege steht der Herausforderung gegenüber, handlungsleitendes Wissen für die Praxis zur Verfügung zu stellen (Moers 2000). Ziel der Forschung ist dabei die Identifikation, Generierung und Überprüfung von möglichst allgemeingültigen Regeln. Forschungsergebnisse erheben keinen Anspruch auf »Wahrheit«, sondern sind Ergebnis einer bestimmten Fragestellung, einer gewählten Methode und daraus resultierenden Ergebnissen. Sie sind primär nur unter diesen genannten Prämissen gültig. Je häufiger diese Ergebnisse jedoch belegt werden, desto »wahrer« werden sie, ohne jedoch den Anspruch auf »die« Wahrheit zu erheben. Forschung ist damit ein Instrument zum Verständnis unserer Welt, deren Muster und Regeln. Sie funktioniert nur so gut, wie seine Anwender sie nutzen und versucht allenfalls, sich der »Wahrheit« zu nähern. Die Herausforderung für die Praxis in der Anwendung von Forschungsergebnissen im Rahmen von komplexen Handlungen besteht deswegen in der Kenntnis der konkreten Bedingungen für die Gültigkeit und Verallgemeinerbarkeit dieser Ergebnisse.

Forschungsmethoden<sup>3</sup> lassen sich unterschiedlich klassifizieren. Eine der am häufigsten genutzten Differenzierungen unterscheidet zwischen »qualitativen« und »quantitativen« Studiendesigns wegen der ihnen unterschiedlichen zugrundeliegenden Methodologien. Fast wie ein roter Faden durchzieht die Forschung die scheinbare Unvereinbarkeit von qualitativen und quantitativen Designs. Vielfach wurde innerhalb der Diskussion aus einer Methodologie eine Ideologie. Der Tenor der heutigen Pflegewissenschaft geht davon aus, dass die Forschungsfrage die Methodologie bestimmt. Qualitative Designs beantworten Fragen zum »Verstehen«, sie identifizieren Konzepte und dienen der Entwicklung von Theorien. Quantitative Designs beantworten Fragen des »Erklärens«, sie beschreiben Häufigkeiten, Unterschiede und Kausalitäten. Während qualitative Designs die Perspektive der »Tiefe« einer Fragestellung einnehmen, fokussieren quantitative Designs die »Breite« einer Fragestellung. Folgerichtig gibt es nicht »die« beste Forschungsmethode zur Gewährleistung einer evidenzbasierten Pflegepraxis.

Welche Fragen stellt nun die Pflegepraxis? Basierend auf der Arbeitsmethode der Pflege zur Lösung der originären Pflegeaufgaben – dem sogenannten »Pflegeprozeß« – sind folgende allgemeinformulierten Forschungsfragen zu beantworten (Evers 2000):

- Werden die Menschen korrekt erkannt, die Pflege brauchen? (Assessment, Evaluation)
- Sind die für sie formulierten Pflegeziele realistisch? (Planning)
- Erhalten diese Menschen die für sie notwendige Pflege? (Intervention)

Entsprechend der aufgestellten Prämissen, die Methode sei abhängig von der Forschungsfrage, werden zur Beantwortung der vielfältigen Fragen aus der Pflegepraxis unterschiedliche Methoden benötigt:

- Anamnese:
  - nicht standardisiertes Assessment: phänomenologische Studien
  - standardisiertes Assessment: Diagnosestudien
- Planning (Ziele): Prognosestudien
- Intervention: Interventionsstudien
- Evaluation: Diagnosestudien.

Nach den Regeln des EBN gibt es für jeweils jede Fragestellung »optimale« Methoden, die durch die Hierarchie der best verfügbaren externen Evidenz beschrieben werden. Diese Hierarchie variiert je nach Fragestellung. Die Kritiker des EBN werfen dieser die ausschließliche Orientierung an RCTs (Randomisierte kontrollierte Studien) vor. Tatsächlich beschreiben diese jedoch nur den höchsten Evidenzgrad für Studien mit kausalen Fragestellungen, z.B. Fragen nach der Wirksamkeit von bestimmten Verfahren. Zur Beurteilung von Studien zur Beantwortung von Fragen nach der Prognose von Pflegebedürftigkeit, der Qualität von Messinstrumenten oder dem Erleben von Krankheit und Gesundheit werden andere Studiendesigns verwendet und entsprechend andere Hierarchiestufen formuliert.

Ich vertrete ebenfalls die Meinung, dass die in der Forschung genutzten Methoden von der Forschungsfrage abhängen, qualitative *und* quantitative Methoden ihre Notwendigkeit haben. Deswegen soll im folgenden auch keine methodologische Di-

skussion geführt werden. Basierend auf den einleitenden Ausführungen soll die Frage »Quantitative Methoden – Grundlage für komplexes Handeln?« anhand von vier Thesen problematisiert werden. Zur Erläuterung der Thesen werden Beispiele quantitativer pflegerischer Diagnose- und Interventionsstudien genutzt.

### These I:

#### **Die quantitative klinische Forschung muß Komplexität reduzieren, um Fragen erforschen zu können**

Diese These soll beispielhaft für die Forschung zu Assessmentinstrumenten diskutiert werden. Die in der Pflegepraxis derzeit am meisten diskutierten Instrumente sind die zur Erfassung des Dekubitusrisikos. Die Entwicklung eines Instrumentes verläuft im Spannungsfeld zwischen Konzeptualisierung, Operationalisierung, Itemgewichtung und Praktikabilität.

Für die Entwicklung eines Instrumentes muß geklärt werden, was mit dem Instrument genau gemessen werden soll. Dazu wird der Erhebungsgegenstand, hier Dekubitusrisiko, konzeptualisiert. In einem der ersten Dekubitusrisikoskalen, der Nortonskala, wurde Dekubitusrisiko konzeptualisiert als Ergebnis der Konzepte Beweglichkeit, Aktivität, Inkontinenz, körperlicher und geistiger Zustand (Norton 1975). In den darauf folgenden Jahren wurde ein neues Konzept in die Debatte um das Druckgeschwürrisiko eingeführt, das der Gewebetoleranz. Braden nutzte dieses in ihrer Skala (Bergstrom et al. 1987). Inzwischen sind nach Gosnell (in Halfens 2000) 126 verschiedene Risikofaktoren für die Dekubitusentstehung bekannt. In den gängigen Risikoskalen werden 5 bis 10 dieser Faktoren abgefragt.

Zur Erfassung des Dekubitusrisikos müssen die einzelnen Faktoren operationalisiert werden. Innerhalb der Skalen wird dies unterschiedlich gelöst. Feuchtigkeit wird beispielsweise als Hautfeuchtigkeit, Inkontinenz, Stuhlinkontinenz und/oder Urininkontinenz abgefragt. Die Ernährungssituation wird über den Ernährungsstatus, Appetit, Nahrungsaufnahme oder Gewicht erfaßt (Panfil 2002). Je nach Art der Erfassung eines bestimmten Parameters verändert sich damit das Ergebnis der Skala.

Das Dekubitusrisiko setzt sich aus unterschiedlichen Faktoren zusammen. Jeder dieser Faktoren hat einen unterschiedlich hohen Einfluß auf das Risiko der Dekubitusentstehung. Ebenfalls verändert sich deren Einfluß, wenn sie zusammen mit anderen Faktoren wirken, also Wechselwirkungen auftreten. Wie diese Faktoren jedoch innerhalb der Skalen gewichtet werden sollen, ist weitgehend unbekannt (Halfens 2000).

Skalen müssen nicht nur inhaltlich valide sein (beschrieben über geeignete Konzeptualisierung, Operationalisierung und Itemgewichtung), sondern für einen Einsatz in der täglichen Pflegepraxis auch praktikabel sein. Schwierigkeitsgrad und Umfang einer Skala haben Einfluß auf dessen Praktikabilität. Die vorangegangenen Ausführungen verdeutlichen, wie schwierig es ist, die Komplexität des Dekubitusrisikos anhand einer praktikablen Skala zu erfassen. Deutlich wurde auch, dass der Anspruch einer bei »allen« Patienten anwendbaren Skala zu Lasten ihrer Validität geht.

Es scheint also praktisch unmöglich zu sein, ein perfektes Instrument zu entwickeln. Zur standardisierten Messung des Dekubitusrisikos ist es unumgänglich, ein komplexes Konzept auf seine wesentlichen Aspekte zu reduzieren. Im Rahmen der Operationalisierung und Itemgewichtung werden weitere einschränkende Entscheidungen getroffen. Diese Entscheidungen haben Auswirkungen auf die Ergebnisse der Risikoskalen. Risikoskalen sind deswegen allenfalls als Hilfsmittel zur Fundierung einer klinischen Entscheidung in Ergänzung zur klinischen Erfahrung einer Pflegefachkraft aufzufassen (DNQP 2002).

### These II: RCTs gleichen »Labor-experimenten«

Interventionsstudien beantworten Fragen nach der Wirksamkeit von bestimmten Maßnahmen (Intervention) zur Erreichung eines bestimmten Zieles (Zielkriterium). Die höchste Aussagekraft zur Beantwortung dieser Fragestellungen wird randomisierten kontrollierten Studien (RCT) zugesprochen (Sackett et al. 1999). Die Durchführung einer qualitativ hochwertigen RCT stellt für Forscher eine große Herausforderung dar. Soll eine RCT klinisch

übertragbare Ergebnisse liefern, muß sie erstens über ausreichend innere Validität verfügen, zweitens geeignete Zielkriterien messen und drittens diese auch in einem für die Praxis sinnvollen Zeitraum erheben.

Die innere Validität einer RCT wird vor allem durch die Reduktion bzw. Kontrolle von Faktoren, die ebenfalls eine Wirkung auf die Zielgröße haben können, wie z.B. dem Placeboeffekt oder die Zielgruppe ebenfalls beeinflussende Variablen, bestimmt. Maßnahmen zu einer effektiven Kontrolle sind die Definition von Ein- und Ausschlusskriterien zur Bestimmung geeigneter Studienteilnehmer, Vergleich der Intervention mit geeigneten Kontrollgruppen, Verblindung der Teilnehmer, Therapeuten, Datensammler und –analysten und Randomisierung. Die Cochrane Library betont in ihren Reviews jedoch die oft mangelhafte Qualität der RCTs. Grund dafür sind u.a. mangelhafte Randomisierung und Verblindung.

Zielkriterien beschreiben den gewünschten Effekt einer Intervention. Die klinische Relevanz dieser Effekte hinsichtlich ihrer qualitativen Aussagekraft ist jedoch manchmal fraglich. So konnte beispielsweise bei der Literaturrecherche zum Expertenstandard Dekubitusprophylaxe festgestellt werden, dass bei einigen Studien anhand falsch gewählter Zielkriterien falsche Schlußfolgerungen gezogen wurden. Der Effekt von Schulungsmaßnahmen zur besseren Prävention von Druckgeschwüren wurde anhand von Prävalenzmessungen anstatt von Inzidenzmessungen erhoben. Die Wirksamkeit von Dekubitusmatratzen wird heute immer noch häufig anhand von reinen Druckmessungen (oft genug noch für Dummies oder gesunde Probanden) belegt. Notwendiger wären Aussagen über die Dekubitusinzidenz bei Anwendung dieser Hilfsmitteln. Gleiches gilt für die Wirksamkeit von Wundauflagen. Hier interessiert nicht die prozentual erreichte Verkleinerung der Wundfläche, sondern die Häufigkeit der erreichten Wundheilungen.

Zielkriterien sollten zusätzlich innerhalb eines für die Praxis relevanten Zeitraumes erhoben werden. Was nutzt es, wenn der Effekt von Patientenschulungen nach einem Monat oder einem halben Jahr gemessen wurde, die Praxis jedoch Effekte für weitaus längere Zeiträume benötigt. Eine Cochrane Review zum Effekt von Patien-

tenschulungen bei Rheuma (Riemsma et al. 2002) beschrieb, dass die Schulungen nach drei Wochen signifikante Verbesserungen von Fähigkeitsstörungen und depressiven Stimmungen bewirkt hatten und tendenziell weniger Schmerzen zu verzeichnen waren. Nach drei Monaten bzw. 14 Monaten war jedoch kein Effekt mehr meßbar. In wie vielen Studien werden jedoch follow-ups nach mehr als einem Jahr bzw. nach noch längeren Zeiträumen durchgeführt?

Studiensituationen unterscheiden sich zum Teil erheblich von klinischen Alltagssituationen. Earl-Slater (1998, zit. in Pientka 2000) bewertet RCTs als künstliche Welten, die den methodischen Anforderungen an hochwertige RCTs geschuldet wären:

- Studienpatienten repräsentieren durch strenge Ein- und Ausschlusskriterien nie das gesamt Spektrum einer Erkrankung sondern immer nur ausgesuchte Teile. Selten spiegeln diese mehr als 50% aller potentiellen Patienten wider (Raspe 2000).
- Die Überwachung von Studienpatienten unterscheidet sich vom Alltag.
- Professionelle handeln in RCTs anders als unter Alltagsbedingungen.

Dies hat entsprechende Auswirkungen auf die Evidenzsicherung. Evidente Interventionen müssen sich deswegen nicht automatisch auch im klinischen Alltag als wirksam erweisen. Multimorbide Patienten, wechselnde Qualität der durchgeführten Interventionen und der Ergebnismessungen und weniger aufmerksame Professionelle haben Auswirkungen auf das Ergebnis der Intervention. Deswegen fordern Methodiker der Evidenzbewegung auch die Ergänzung der RCTs durch Beobachtungsstudien, um die Wirksamkeit der als evident belegten Maßnahmen auch im Praxisalltag zu untersuchen.

### **These III: Ergebnisse von quantitativen klinischen Forschungen stellen keine »Wahrheiten« sondern »Wahrscheinlichkeiten« dar**

Allen quantitativen Designs ist die Grundorientierung am Prohabilitismus gemein. Diese bei den Anwendern von Forschung noch eher wenig verbreitete Erkenntnis bedeutet, dass diagnostische, prognostische und therapeutische Aussagen keine absolute Wahrheit, sondern allein

eine mehr oder weniger große Wahrscheinlichkeit für sich beanspruchen können (Raspe 2000). Die Gültigkeit von Forschungsergebnissen wird mit Hilfe von Konzepten und Methoden der klinischen Epidemiologie geprüft. Diese beschreiben in Wahrscheinlichkeiten ausgedrückte Chancen und Risiken des klinischen Handelns. Während bereits die Wettervorhersage in den letzten Jahren dazu übergegangen ist, nicht mehr den »Regen« an sich sondern das »Regenrisiko« anzugeben, gehen Forschungsanwender noch allzu häufig davon aus, dass eine durch Forschung belegte »Wirksamkeit« oder »Überlegenheit« einer Intervention bei den Studienpatienten für alle vergleichbaren Patienten im Praxisalltag gilt. Dies ist jedoch nicht so, die Wirksamkeit wird nicht dichotom sondern anhand von Wahrscheinlichkeiten angegeben. Beispiele sollen dies illustrieren.

Meyer et al. (2003) untersuchten anhand einer RCT die Wirksamkeit der Art des Angebotes von Hüftprotektoren auf sturzbedingte Hüftfrakturen. An der Studie nahmen 924 Bewohner aus 49 Hamburger Altenheimen teil. Die Altenheime bzw. deren Abteilungen wurden randomisiert auf Interventions- bzw. Kontrollgruppe verteilt. In der Interventionsgruppe wurden strukturierte Schulungen für Pflegende, Informationsmaterialien für Angehörige und Ärzte und kostenlose Hüftprotektoren angeboten. In der Kontrollgruppe erhielt die verantwortliche Pflegefachkraft eine kurze Information, eine 10-minütige Demonstration und zwei Exemplare von Hüftprotektoren. Nach 18 Monaten waren in der Interventionsgruppe 4,6% Hüftfrakturen, in der Kontrollgruppe 8,1% Hüftfrakturen aufgetreten. Die Studie belegt damit die Reduktion von sturzbedingten Hüftfrakturen nach strukturierten Schulungen zum Einsatz von Hüftprotektoren und deren kostenloser Abgabe.

Die durch eine Studie belegte Wirksamkeit einer Intervention bedeutet jedoch nicht, dass jeder potentielle Patient von dieser Art der Intervention profitiert. Die relative Risikoreduktion<sup>4</sup> (RRR) durch die Intervention beträgt in dem genannten Beispiel 43%. Eine beeindruckende Zahl, mit der vor allem Firmenvertreter gerne die Wirksamkeit ihre Produkte anpreisen (Greenhalgh 2000). Vertreter der Evidence-Bewegung empfehlen jedoch, Angaben zur

»absoluten Risikoreduktion« und »Number needed to treat« zu nutzen, um die Wirksamkeit von Maßnahmen seriös einschätzen zu können. Die absolute Risikoreduktion<sup>5</sup> (ARR) erfaßt den wahren Interventionsnutzen und beträgt in diesem Fall 3,5%. Eine rein numerisch vergleichsweise weniger beeindruckende Zahl als die Höhe der RRR. Die Kennzahl »Number needed to treat« (NNT) beschreibt die Anzahl der Patienten, die über einen bestimmten Zeitraum behandelt werden müssen, um ein zusätzliches unerwünschtes Ereignis zu verhindern<sup>6</sup> (Kunz et al. 2000). Im Beispiel beträgt die NNT 29. 29 Patienten müssen also über 18 Monate mit der untersuchten Intervention behandelt werden, damit eine unerwünschte Hüftfraktur verhindert wird. Wer nun enttäuscht das Blatt sinken läßt, möge darauf hingewiesen werden, dass dies ein vergleichsweise guter Wert ist.

Für die Bewertung von Assessmentinstrumenten liefert der Kennwert »Likelihood Ratio« die erforderlichen Wahrscheinlichkeitsangaben. Gängig sind Angaben zur Spezifität, Sensitivität, positiven und negativen prädiktiven Validität, um die Qualität von Risikoinstrumenten einschätzen zu können. Eine vergleichende Bewertung verschiedener Instrumente auf Basis der vier Werte ist jedoch umständlich und liefert der Praxis nur wenige praktische Informationen zur Aussagekraft eine Skala. Likelihood Ratio (LR) beschreibt die Wahrscheinlichkeit, dass ein positives (+) (bzw. negatives (-)) Testergebnis bei einer Person mit der Erkrankung auftritt im Verhältnis zu der Wahrscheinlichkeit, dass dieses positive (bzw. negative) Ergebnis bei einer Person ohne diese Erkrankung auftritt (Kunz et al. 2000).

Halek/Mayer (2002) testeten die prädiktive Validität der originalen Nortonskala in einem deutschen Altenheim. Sie fanden eine Sensitivität von 94,6%, eine Spezifität von 18,6%, eine positive prädiktive Validität von 29,7% und eine negative prädiktive Validität von 90,5%. Dies bedeutet, dass die originale Nortonskala 94,6% der »dekubitusgefährdeten« Personen korrekt als gefährdet erkennt und 18,6% der »nicht dekubitusgefährdeten« Bewohner richtig als nicht gefährdet erkennt. Aus der Perspektive der Nortonskala bedeutet dies, dass bei 29,7% der Personen mit Testergebnis »gefährdet« tatsächlich eine Dekubitusgefährdung vorliegt und bei 90,5% der Per-

sonen mit Testergebnis »nicht gefährdet« tatsächlich keine Dekubitusgefährdung vorliegt. Die originale Nortonskala neigt auf Basis dieser Studie zu einer Überschätzung des Dekubitusrisikos (Halek/Mayer 2002). Der Wert LR+ für die originale Nortonskala in dieser Studie beträgt 1,16. Damit ist es 1,16 mal wahrscheinlicher, dass ein positiver Test bei einem dekubitusgefährdeten Bewohner auftritt als bei einem nicht dekubitusgefährdeten Bewohner. Nach den gängigen Bewertungen des LR-Wertes (Bucher 2000) ist damit mit der Nutzung der Nortonskala zur Erkennung von dekubitusgefährdeten Patienten kaum ein Informationsgewinn verbunden. Der LR- beträgt 0,29, damit liefert die Skala zur Erkennung von nicht dekubitusgefährdeten Bewohnern manchmal relevante Informationen. Wünschenswert wären Ergebnisse von über 10 für die LR+ und kleiner 0,1 für die LR-, um die Dekubitusgefährdung sicher ein- bzw. ausschließen zu können.

#### **These IV: Nicht alle Fragestellungen sind erforschbar**

Nicht alle Fragestellungen, die aus Sicht der Praxis dringend erforscht werden sollten, können tatsächlich innerhalb von Studien bearbeitet werden. Einige der Gründe dafür sind pragmatischer Art. So kann man beispielsweise zur Erforschung der Wirksamkeit von bestimmten Ernährungsmaßnahmen auf die Verhinderung von Krebs-erkrankungen kaum eine RCT mit ausreichender Kontrolle, Verblindung und mehrjähriger Laufzeit planen. Hartgrink et al. (1998) untersuchten in einer anderen Studie die Wirksamkeit einer zusätzlichen Ernährung über Magensonde auf die Dekubitusinzidenz. Sie konnten die Ergebnisse der Studie letztlich wegen der mangelnden Akzeptanz der Magensonde durch die Studienteilnehmer nicht auswerten.

Fragestellungen können jedoch auch aus ethischen Gründen nicht zu realisieren sein. Häufiges Beispiel dazu ist die Erforschung der kausalen Zusammenhänge zwischen Rauchen und der Entstehung von Lungenkrebs. Bis heute gibt es keine RCT, die diese Frage untersucht hat. Da epidemiologische Studien diesen Zusammenhang fast zwingend nahelegen, ist es ethisch nicht vertretbar, die geplante Intervention

»Rauchen« zu untersuchen.

An anderer Stelle wurde schon über Kennwerte zur Bestimmung der Qualität von Messinstrumenten gesprochen. Aus ethischen Gründen ist die Bestimmung der prognostischen Validität nicht so möglich, wie sie methodisch sinnvoll wäre. Die prognostische Validität beschreibt die Übereinstimmung eines bestehenden externen Kriteriums mit dem Ergebnis der Skala. Zur Testung der Dekubitusrisikoskalen wird als externes Kriterium in der Regel die Dekubitusinzidenz gewählt. Man geht davon aus, dass, wenn die Skala valide mißt, nur bei den Patienten ein Dekubitus entstehen kann, die auch von der zu testenden Skala als dekubitusgefährdet erkannt worden sind. Umgekehrt sollten die Patienten, die als nicht gefährdet getestet wurden, keinen Dekubitus entwickeln. Halfens (2000) kritisiert diese Art des Testens. Da es ethisch nicht vertretbar ist, gefährdeten Patienten präventive Maßnahmen vorzuenthalten, wird bei dieser Testung gemessen, wie viele gefährdete Patienten trotz pflegerischer Maßnahmen einen Dekubitus entwickeln. Damit wird eigentlich nicht die Validität des Instrumentes bestimmt, sondern die Qualität der Pflege. Unterschiedliche Ergebnisse zur prognostischen Validität sind deswegen auch immer den verschiedenen durchgeführten pflegerischen Maßnahmen geschuldet und weniger der Validität des Instrumentes in unterschiedlichen Settings. Halfens bewertet deswegen auch die von Braden/Bergstrom (1996) geforderte Bestimmung des cut off Punktes speziell für jede Gesundheitseinrichtung als wenig sinnvoll. Damit wäre das Dekubitusrisiko eine Funktion der Qualität der Pflege und weniger der Validität des Instrumentes.

#### **Schlussfolgerungen**

Die klinische Pflegepraxis besteht aus komplexen Handlungsabläufen. Die Komplexität wird u.a. durch die Interaktion der beteiligten Akteure, den internen und externen Bedingungen der Institution, deren Aufbau- und ablauforganisatorischen Prozesse und den interpersonellen Bedingungen aller beteiligten Akteure bestimmt. Quantitative Studien können komplexes Handeln nicht sicher fundieren, sie können allenfalls Hilfsstellungen für einzelne Entscheidungen geben. Vertreter der Evidenzbewegung sehen daher auch die Orientie-

rung an Forschungsergebnissen nur als einen Baustein für eine evidenzbasierte Praxis. Evidenzbasierte Praxis bedeutet, »integrating individual clinical expertise with the best available external clinical evidence from systematic research« (Muir Gray 2000, S. 32). Die Begründer der EBM betonen sowohl die Gleichwertigkeit klinischer Erfahrungen mit denen der extern vorhandenen Evidenz als auch die unzureichende Aussagekraft der einzelnen Komponenten:

»Ohne klinische Erfahrung und Sachverstand läuft die Praxis Gefahr, von externer Evidenz beherrscht zu werden, denn hervorragende externe Evidenz ist unter Umständen beim einzelnen Patienten nicht anwendbar oder sogar ungeeignet. Ohne die jeweils beste externe Evidenz droht die praktische Medizin zum Nachteil der Patienten schnell zu veralten.« (Sackett et al. 1999, S. 2)

Pflegende sollten quantitativen Studien wegen ihrer unzureichenden Fundierung komplexer Handlungen – was aber auch nicht deren Anspruch ist! – nicht ihre Legitimation absprechen. Sie sollten die Qualität der Studien vielmehr kritisch einschätzen, die Ergebnisse realistisch im Sinne von Wahrscheinlichkeiten bewerten und deren Übertragbarkeit auf den klinischen Einzelfall beurteilen können. Forschung ist ein Instrument, dessen Nutzen sowohl von der Kompetenz der Forscher als auch der der Forschungsanwender abhängt.

*Dr. Eva-Maria Panfil  
Professorin für klinische Pflegewissenschaft  
Fachhochschule Frankfurt am Main  
Nibelungenplatz 1  
60318 Frankfurt*

<sup>1</sup> Der ursprüngliche Titel dieses Vortrages sollte heißen: »Evidence based in der Pflegeforschung – Welche Kriterien sind einzuhalten?«. Da dieser Titel Beziehungen zwischen Begriffen herstellt, die per definitionem nicht bestehen, wurde ein neuer Titel für diesen Vortrag gewählt. »Evidence-based« bezieht sich auf die Nutzung der Forschung in der Praxis, Forschung kann deswegen nicht evidenzbasiert sein. Es können in der Forschung aber »evidente« Methoden gewählt werden, d.h. die Methoden, die derzeit die besten Verfahren zur

Beantwortung einer Forschungsfrage darstellen.

<sup>2</sup> Überarbeitete Fassung des Vortrages.

<sup>3</sup> Forschungsmethoden sind nicht zu wechseln mit Datenerhebungsmethoden.

<sup>4</sup> Relative Risikoreduktion: Relative Senkung der Rate an unerwünschten Ereignissen in der experimentellen Gruppe im Vergleich zur Kontrollgruppe. Formel:  $RRR = \frac{CER - EER}{CER}$  (CER = Ereignisrate in der Kontrollgruppe, EER = Ereignisrate in der Experimentalgruppe) (Kunz et al. 2000)

<sup>5</sup> Absolute Risikoreduktion: Absolute Senkung der Ereignisrate in der Experimentalgruppe im Vergleich zur Kontrollgruppe. Formel:  $ARR = CER - EER$  (Kunz et al. 2000)

<sup>6</sup> Formel:  $NNT = 1/ARR$

#### Literatur

- Bergstrom, N.; Braden, B.; Laguzza, A.; Holman, V.: The Braden scale for predicting pressure sore risk. *Nursing Research*, 36, 4, 1987, 205-210
- Braden, B.; Bergstrom, N.: Risk-assessment and risk-based programs of prevention in various settings. *Ostomy Wound Management*, 42, 1996, Suppl. 10A, 6-12
- Bucher, H.: Kritische Bewertung von Screening und Diagnostik. In: Kunz et al., aaO, 2000, 108-119
- Deutsches Netzwerk für Qualitätsentwicklung in der Pflege (DNQP): Expertenstandard Dekubitusprophylaxe in der Pflege, Osnabrück 2002
- Evers, G.: Klinische Pflegeforschung. *Pflege*, 13, 3, 2000, 133-138
- Greenhalgh, T.: Einführung in die Evidence-based Medicine. Hans Huber Verlag, Bern u.a., 2000
- Halfens, R.: Risk assessment scales for pressure ulcers: A theoretical, methodological and clinical perspective. *Ostomy Wound Management*, 46, 8, 2000, 36-44
- Halek, M.; Mayer, H.: Die prädiktive Validität der originalen und erweiterten Norton-Skala in der Altenpflege. *Pflege*, 15, 6, 2002, 309-317
- Hartgrink, H.; Wille, J.; König, P.; Hermans, J.; Breslau, P.: Pressure sores and tube feeding in patients with a fracture of the hip: a randomised clinical trial. *Clinical nutrition*, 17, 6, 1998, 287-292
- Kunz, R.; Ollenschläger, G.; Raspe, H.; Jonitz, G.; Kolkmann, F.-W.: *Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis*. Deutscher Ärzte Verlag, Köln, 2000
- Meyer, G.; Warnke, A.; Bender, R.; Mühlhauser, I.: Effect on hip fractures of increased use of hip protectors in nursing homes: cluster randomised controlled trial. *BMJ*, 326, 2003, 76
- Moers, M.: Pflegewissenschaft: Nur Begleitwissenschaft oder auch Grundlage des Berufes? *Pflege & Gesellschaft*, 5, 1, 2000, 21-25
- Mulhall A.: Nursing, research, and the evidence. *Evidence-based Nursing*, 1, 1998, 4-6
- Muir Gray, J.: Die englische EbM-Story. In: Kunz et al., aaO, 2000, 30-37
- Norton, D.; McLaren, R.; Exton-Smith, A.: *An investigation of geriatric nursing problems in hospitals*. Churchill Livingstone, New York, 1975
- Panfil, E.: Literaturanalyse zur Dekubitusprophylaxe. In: DNQP, aaO, 2002, 46-62
- Pientka, L.: EbM in der Geriatrie – Beispiel »Alzheimer Demenz«. In: Kunz et al., aaO, 2000, 226-233
- Raspe, H.: Grundlagen und Theorie der evidenzbasierten Medizin. In: Kunz et al., aaO, 2000, 38-49
- Riemsma, R.; Kirwan, J.; Taal, E.; Rasker, J.: Patient education for adults with rheumatoid arthritis. *Cochrane Library*, Systematic review, 2002
- Sackett, D.; Richardson, W.; Rosenberg, W.; Haynes, R.: *Evidenzbasierte Medizin*. W. Zuckschwerdt Verlag, Bern, Wien, New York, 1999