

Expertise zur

« Eignung von Indikatoren zur Messung und Darstellung von Ergebnisqualität in der stationären Pflege im Bereich der sozialen Pflegeversicherung »



Prof. Dr. Renate Stemmer

Kath. Hochschule Mainz
Fachbereich Gesundheit und Pflege
Saarstraße 3 | 55122 Mainz

Dipl.-Psych. Jens Arnold

Centrum für angewandte
Wirkungsforschung (CAW)
IKJ ProQualitas GmbH
Saarstraße 1 | 55122 Mainz

Unter Mitwirkung von
Arne Nils Evers,
Martin Müller
und **Jana Wirz**

Mainz, im Juni 2014



Spitzenverband

erstellt im Auftrag des
GKV-Spitzenverbandes

ISBN 978-3-00-051281-0

Inhaltsverzeichnis

Management Summary	4
1 Einleitung	8
1.1 Hintergrund.....	8
1.2 Aufgabenstellung.....	9
1.3 Konzeptionelle Überlegungen.....	11
1.3.1 Indikatorenentwicklung.....	11
1.3.2. Interne Qualitätsentwicklung, Qualitätsberichterstattung externe Qualitätsprüfung.....	13
2 Methodisches Vorgehen	16
2.1 Vorgehen bei der Recherche.....	16
2.1.1 Recherche internationaler Einrichtungen des Gesundheitswesens.....	16
2.1.2 Recherche in Fachdatenbanken und Datenbankportalen	17
3 Zusammenfassender Überblick über pflegebezogene Outcomemessung im internationalen Raum	24
4 Beantwortung der Fragen	39
4.1 Reichweite der Indikatoren	39
4.1.1 Bildet das Indikatorenset von Wingenfeld die für eine vergleichende Qualitätsberichterstattung relevanten Pflegebereiche in der erforderlichen Themenbreite und -tiefe hinreichend ab? Auf welche von Wingenfeld et al. vorgesprochenen Themenbereiche/Indikatoren kann verzichtet werden, welche sollten ergänzend berücksichtigt werden?.....	39
4.1.2 Wenn relevante Themenbereiche nicht abgedeckt sind, wie könnten diese einbezogen werden, wenn eine Indikatorenbildung methodisch nicht valide möglich ist?	53
4.1.3 Bestimmte Personen werden aus der Bewertung ausgeschlossen. Sind die jeweils definierten Ausschlusskriterien (z. B. Schlaganfall oder akute Verschlechterung einer Gesundheitsstörung), die zum Ausschluss aus der Bewertung führen, akzeptabel oder führen sie zu einer systematischen Verzerrung der Qualitätsberichterstattung?	56
4.1.4 Fazit	60
4.2 Datengrundlage/Externe Qualitätsprüfung.....	62
4.2.1 Sind für eine Bewertung der Indikatoren neben der Pflegedokumentation weitere Datenquellen notwendig?.....	62
4.2.2 Wie kann erreicht werden, dass die Daten des internen Qualitätsmanagements flächendeckend eine zuverlässige und valide Datengrundlage liefern?	64

4.2.3	Können mit den von Wingefeld et al. vorgeschlagenen Plausibilitätsprüfungen im Rahmen externer Qualitätsprüfungen die Daten über die Qualität der Pflege verifiziert werden	68
4.2.4	Fazit	69
4.3	Bewertungssystematik.....	71
4.3.1	Stellen die von Wingefeld et al. gegebenen Empfehlungen zur Bewertungssystematik eine Form der Bewertung von Indikatoren/Indikatorenbereichen dar, die eine Differenzierung von Einrichtungen ermöglicht und sind die vorgenommenen Setzungen ausreichend fachlich begründet?	71
4.3.2	Gibt es Alternativen, z. B. die ergänzende Berücksichtigung von zu erfüllenden Mindestniveaus, und wie werden diese bewertet?	82
4.3.3	Fazit	86
4.4	Risikoadjustierung	87
4.4.1	Sind die der Risikoadjustierung zugrunde gelegten Kriterien (Items aus dem NBA) angemessen, um eine valide Risikoadjustierung zu ermöglichen? Gibt es Hinweise oder Vorschläge zu Alternativen oder zur Weiterentwicklung?	87
4.4.2	Fazit	103
4.5	Verbraucherperspektive.....	104
4.5.1	Welche Indikatoren sind aus der Verbraucherperspektive relevant?.....	105
4.5.2	Sind Indikatoren bzw. die Empfehlungen von Wingefeld et al. für eine verständliche, übersichtliche und vergleichbare Information der Verbraucher geeignet? Ist die verfügbare Information in ihrer Auswahl hinreichend, in ihrer Bewertung und Darstellung verständlich und erlaubt sie einen Vergleich der Einrichtungen/Dienste insbesondere bezogen auf Ergebnisqualität? Wenn nein, welche Voraussetzungen, Änderungen, Ergänzungen wären hierfür notwendig? 109	
4.5.3	Sind die Empfehlungen von Wingefeld et al. zur Bewertungssystematik auf Ebene der Indikatoren und der Indikatorenbereiche aus Verbraucherperspektive geeignet, eine hinreichende Differenzierung der Einrichtungen zu ermöglichen? Sind hierfür voraussichtlich fortlaufende Nachjustierungen am Bewertungssystem notwendig?	115
4.5.4	Fazit	116
5	Roadmap	117
	Literatur	120
	Anlage: Übersicht über die Fragestellungen der Expertise	132
	Abbildungsverzeichnis.....	134
	Tabellenverzeichnis.....	134

Management Summary

(Renate Stemmer)

Mit dem von Wingenfeld et al. (2011) vorgelegten Entwurf für Indikatoren zur Outcomemessung in der stationären Altenhilfe ist ein wichtiger Startpunkt gesetzt, der nun weiter verfolgt werden sollte.

Wenn Outcomemessung in der stationären Altenhilfe durchgeführt wird, dann gehören die Bereiche Selbständigkeit und Sicherheit sowie weitere Themen wie Fixierung und Schmerzmanagement in der Regel dazu. Betrachtet man jedoch die vorgeschlagene Reichweite und -tiefe der Wingenfeld-Indikatoren, dann zeigt sich Entwicklungsbedarf.

Die Indikatoren zu Mobilität und Selbständigkeitserhalt oder –verbesserung bei der Durchführung der Alltagsverrichtungen werden prinzipiell unterstützt. Die Erhebung des Selbständigkeitserhalts oder der -verbesserung bei der Gestaltung des Alltagskontakte allein als Fremdbeobachtung erscheint aber schwierig und sollte um den Abgleich mit der Erhebung des Erlebens der BewohnerInnen bezogen auf dieses Themenfeld ergänzt werden. Die Messung der Dekubitusinzidenz, der Stürze mit gravierenden Folgen und des unbeabsichtigten Gewichtsverlustes als Outcome entspricht prinzipiell dem state of the art. Einzig die Erhebung der Dekubitusinzidenz bei BewohnerInnen mit geringem Risiko scheint nicht zwingend erforderlich. Auch sollte der Gewichtsverlust nicht nur nach 6 Monaten, sondern schon nach 4 Wochen erkannt werden. Angesichts der hohen inhaltlichen Relevanz sollte die Harnkontinenz zusätzlich in den Kanon der Messthemen primär als Outputparameter aufgenommen werden. Im Bereich der spezifischen Bedarfslagen überzeugt der vorgeschlagene Indikator ‚Integrationsgespräch nach Heimeinzug‘ nicht, da er eine Leistung erfasst, die in ihrer Ausgestaltung nicht evidencebasiert ist und dessen Interpretation im Hinblick auf die BewohnerInnen auf Vermutungen angewiesen ist. Der Indikator zur Gurtfixierung wird in der vorgeschlagenen Form für sinnvoll erachtet. Anders sieht dies aus bei der Einschätzung von Verhaltensauffälligkeiten bei BewohnerInnen mit kognitiven Einbußen. Die grundsätzliche Relevanz des Indikators ist unbestritten, die Form der Operationalisierung erscheint aber nicht schlüssig. Zudem sollte dieser Outputparameter nicht allein das Vorhandensein einer instrumentengestützten Einschätzung erfragen, sondern um die Durchführung von Fallbesprechungen erweitert werden. Der Indikator zum Schmerzmanagement sollte ausdifferenziert (akut/chronisch), im Bereich des akuten Schmerzes zu einem Outcomeindikator ausgebaut und im Bereich des chronischen Schmerzes um einen individuellen Maßnahmeplan ergänzt werden.

Es wird begrüßt, dass die Qualitätsmessung in der stationären Altenhilfe in Richtung auf Outcomemessung weiterentwickelt wird, diese ist jedoch durch Prozess- und Strukturindikatoren zu ergänzen.

In der von Wingenfeld et al. (2011) vorgelegten Systematik sind Ausschlüsse letztendlich unumgänglich, da ein fairer Einrichtungsvergleich hier sonst nicht zu gewährleisten wäre. Der Ausschluss etwa von Kompatienten, deren Situation bei bestimmten Outcomes durch pflegerisches Handeln nicht beeinflusst werden kann, stellt in diesem Zusammenhang daher

auch eine Form der Risikoadjustierung dar. Da die Notwendigkeit von Ausschlüssen in der Literatur generell kritisch hinterfragt wird, sollten diese nur in begründeten Fällen vorgenommen und dann am besten über ein standardisiertes Testverfahren in transparenter und valider Weise dokumentiert werden. Auf Ausschlüsse komplett verzichtet werden könnte im Rahmen der von uns vorgeschlagenen alternativen Risikoadjustierungsmethode. Im Rahmen dieses multivariaten Ansatzes könnten die Ausschlusskriterien einfach als zusätzliche Risikofaktoren einfließen und kontrolliert werden.

Die anvisierte dreifache Nutzung der Daten zur internen Qualitätsentwicklung, zur externen Qualitätskontrolle und zur Qualitätsberichterstattung hat Konsequenzen auf drei Ebenen. Die Pflegedokumentation wird nun zum unmittelbaren Datenlieferanten für die Outcomemessung. Die aktuelle Vielfalt der Dokumentationssysteme und -verfahren lässt kaum eine Datenerhebung zu, die den Anforderungen an die Datenqualität für die Outcomemessung gerecht wird. Die Systematik der Pflegedokumentation ist äußerst heterogen und die Anschlussfähigkeit an ein Übermittlungssystem in keiner Weise gesichert. Wenn die aktuell breit gefächerten Dokumentationssysteme für die Datengrundlage genutzt werden sollen, hat dies umfangreiche Doppeldokumentationen verbunden mit hoher Fehleranfälligkeit zur Folge. In der Konsequenz ist eine edv-gestützte Pflegedokumentation verpflichtend einzuführen, die Items für die Datenerhebung in das Dokumentationssystem zu integrieren und die Schnittstelle für die Datenübermittlung zu definieren.

Wenn die dreifache Nutzung der Daten auch aus Sicht der Einrichtungen gewinnbringend gelingen soll, darf diese nicht nur die Aufgabe haben, für eine ausreichend Qualität der Rohdaten zu sorgen, sondern muss die Ergebnisse der Outcomemessung für die eigene Qualitätsentwicklung nutzen. Dazu sind die Auswertungsdaten den Einrichtungen zur Verfügung zu stellen. Darüber hinaus ist zu empfehlen, dass die Einrichtungen bei der Integration weitere Unterstützung erfahren. Dazu könnten die Benennung prioritärer Bereiche gehören, sowie das Angebot von Informationen, Strategien und Hilfen, um die interne Qualitätsentwicklung im Hinblick auf diese empfohlenen Prioritäten zu gestalten. Auf der Ebene der externen Qualitätskontrolle kommt nun der Auswertungsstelle in enger Zusammenarbeit mit den Einrichtungen eine wichtige Funktion bei der Sicherung einer ausreichenden Qualität der Daten und damit der Verlässlichkeit der Ergebnisse der Outcomemessung zu. Im Rahmen von Vor-Ort-Begehungen des MDKs können Outcomes nicht unmittelbar überprüft werden, aber sehr wohl im Sinne von Plausibilitätsprüfung zum Ausgangspunkt des Gespräches werden.

Der von Wingefeld et al. (2011) vorgelegte Vorschlag zur Risikoadjustierung ist in mehrfacher Hinsicht noch optimierungsfähig. Dabei ist nur ein Kritikpunkt, dass die Validität, Messgenauigkeit und Trennschärfe der aus dem NBA übernommenen Skalen zur Identifikation der Risikogruppen, z. B. BewohnerInnen mit hohen und niedrigen kognitiven Einbußen, nicht hinreichend belegt ist und hier daher ggf. besser erprobte Verfahren zum Einsatz kommen sollten. Die von Wingefeld et al. (2011) präferierte Adjustierungsmethode auf der Grundlage von Stratifizierung, weist an sich schon die wesentliche Limitierung auf, dass nur eine sehr begrenzte Anzahl an potenziellen Risiken einbezogen werden kann. Die internationale Literatur zeigt aber, dass bei allen zentralen Indikatorbereichen der Wingefeld-Systematik in der Regel eine Vielzahl von weiteren Risikofaktoren eine Rolle spielen. Mitunter kann selbst, die von Wingefeld et al. (2011) unterstellte, zentrale Rolle der kognitiven Beeinträchtigung nicht durchgehend belegt werden. Auf dieser Grundlage ist eine adäquate Risikoadjustierung nicht garantiert. Da diese für einen fairen Anbietervergleich unerlässlich ist und bei Wingefeld et

al. (2011) sogar in die Definition der Indikatoren einfließt, indem bspw. separate Indikatoren für BewohnerInnen mit hohem und niedrigen Risiken gebildet werden, empfehlen wir eine alternative Risikoadjustierungsmethode, auf der Basis von multiplen Regressionsanalysen, bei der eine Vielzahl von Risikofaktoren gleichzeitig betrachtet und adjustiert werden können. Da diese Technik unabhängig von der Indikatordefinition ist, kann auf diese Weise, bei vergleichbarer Reichweite, die Gesamtzahl der Qualitätsindikatoren deutlich gesenkt werden. Entsprechende Verfahren sind vor allem in der amerikanischen Pflegeforschung Standard und prinzipiell auch auf den deutschsprachigen Raum übertragbar. Dies bedeutet allerdings auch, dass die Datenerhebung um die entsprechenden Risiken erweitert werden müsste. Auf diesem Hintergrund wäre ebenfalls ein systematisches Literaturreview, der für die relevanten Indikatorbereiche im Einzelnen maßgeblichen Risikofaktoren, ratsam.

In der von Wingenfeld et al. (2011) vorgelegten Systematik sind Ausschlüsse letztendlich unumgänglich, da ein fairer Einrichtungvergleich hier sonst nicht zu gewährleisten wäre. Der Ausschluss etwa von Komapatienten, deren Situation bei bestimmten Outcomes durch pflegerisches Handeln nicht beeinflusst werden kann, stellt in diesem Zusammenhang daher auch eine Form der Risikoadjustierung dar. Da die Notwendigkeit von Ausschlüssen in der Literatur generell kritisch hinterfragt wird, sollten diese nur in begründeten Fällen vorgenommen werden und dann am besten über ein standardisiertes Testverfahren in transparenter und valider Weise dokumentiert werden. Auf Ausschlüsse komplett verzichtet werden könnte im Rahmen der von uns vorgeschlagenen alternativen Risikoadjustierungsmethode. Im Rahmen dieses multivariaten Ansatzes könnten die Ausschlusskriterien einfach als zusätzliche Risikofaktoren einfließen und kontrolliert werden.

Für eine differenzierte Einteilung von Einrichtungen erscheint die von Wingenfeld et al. (2011) vorgelegte dreistufige Bewertungssystematik als zu grob. Die AutorInnen der vorliegenden Expertise empfehlen analog zum internationalen Standard ein fünfstufiges Bewertungsraster, das in weit überdurchschnittliche, überdurchschnittliche, durchschnittliche, unterdurchschnittliche und weit unterdurchschnittliche Einrichtungen unterteilt. Die Differenzierungsfähigkeit wird allerdings nicht nur durch die Feingliedrigkeit der Bewertungsstufen, sondern auch durch die tatsächlich zu beobachtende Variabilität der Messwerte beeinflusst. In diesem Kontext erweisen sich insbesondere einige der für BewohnerInnen mit niedrigen Risikoausprägungen vorgesehenen Varianten der Wingenfeld-Indikatoren als ungeeignet. Die von uns vorgeschlagene alternative Risikoadjustierungsmethode, die ohne derartige Unterscheidungen auskommt, könnte hier allerdings Abhilfe schaffen. Auch andere von Wingenfeld et al. (2011) vorgenommene Setzungen können nicht durchgehend überzeugen. Von der Verwendung indikatorübergreifender Gesamtwerte zum Vergleich von Einrichtungen, raten wir auf der Grundlage des vorliegenden Forschungsstandes ab. Damit am Ende nicht sprichwörtlich „Schlendrian mit Schlendrian“ verglichen wird, sollte die Qualitätsbewertungen nicht ausschließlich am Anbieterdurchschnitt bemessen werden. Wir empfehlen daher dringend zukünftig verstärkt expertenbasierte Kriteriumsstandards zu formulieren. Damit sich bessere Qualität auch lohnen kann, sollte auf dem Hintergrund der Initiierung von zielführenden Qualitätsentwicklungsprozessen, aber nicht ausschließlich auf Mindestniveaus zurückgegriffen werden. Die Kriteriumsstandards sollten mindestens zwei Orientierungsmaßstäbe setzen, einen Grundstandard, der noch akzeptable Qualität umreißt und einen Zielstandard, ab dem höherwertige „gute“ pflegerische Qualität beginnt. Eine fortlaufende Nachjustierung wäre allerdings auch hier unerlässlich.

VerbraucherInnen haben andere Informationsbedürfnisse als Fachkräfte. Outcomeindikatoren nehmen hier grundsätzlich nur einen untergeordneten Stellenwert ein. Die VerbraucherInnen legen, im Rahmen der Entscheidung für oder gegen ein Pflegeheim, insgesamt mehr Wert auf Prozessqualität wie etwa die Freundlichkeit und Zuwendung des Personals oder strukturelle Merkmale der Einrichtung wie z. B. die geographische Nähe zum ursprünglichen Wohnort. Da der Verbraucherperspektive im Rahmen der Qualitätsberichtserstattung insgesamt eine zentrale Rolle zukommt, können die Informationsbedürfnisse der VerbraucherInnen nicht ignoriert werden. Dies heißt allerdings im Umkehrschluss nicht, dass Ergebnisindikatoren hier komplett außer Acht zu lassen wären. Neben Kriterien wie der Verständlichkeit der Sprache, ist insbesondere aber auf eine überschaubare Informationsmenge zu achten. Die Veröffentlichung von insgesamt mehr als drei bis sechs Indikatoren wäre aus Sicht der GutachterInnen eher nicht zielführend. Die Auswahl der zur Verfügung gestellten Indikatoren sollte dabei auf die jeweiligen Interessensgruppen abgestimmt werden. Mit Hilfe einer geeigneten Online-Plattform könnte die Auswahl und Gewichtung allerdings teilweise auch durch die NutzerInnen selbst geschehen. Sofern man unseren Empfehlungen zur Bewertungssystematik und zur Risikoadjustierung, bei denen die Nutzersicht bereits mitgedacht wurde, sowie den allgemeinen Gestaltungsleitlinien für „kundenfreundliche“ Qualitätsberichtserstattung folgt, wäre auch auf Grundlage vieler von Wingefeld et al. (2011) beschriebener Indikatorbereiche, ein angemessener Einrichtungsvergleich möglich, der sowohl im Hinblick auf Übersichtlichkeit als auch Verständlichkeit der Verbraucherperspektive gerecht werden kann. Dies gilt insbesondere in Bezug auf die von Wingefeld et al. (2011) entwickelten Indikatoren der Qualitätsbereiche 1 und 2, deutlich weniger aber für die Indikatoren des Qualitätsbereichs 3.

1 Einleitung

(Renate Stemmer)

Die Zielsetzungen und Verfahrensweisen interner und externer Qualitätsmessung in der Pflege erfahren seit Jahren national und international eine intensive Diskussion. Dabei sind die Möglichkeiten einer ergebnisbasierten Qualitätsmessung immer mehr in den Blick gerückt. Die Verpflichtung zur Qualitätsmessung im deutschen Gesundheitswesen ist seit Jahren sowohl in SGB V als auch in SGB XI verankert. Im Rahmen des SGB XI wird dies bislang überwiegend als Überprüfung von Prozessen durchgeführt. Zwar ist es im Bereich des SGB V gelungen, Outcomemessung einzuführen, jedoch konkretisiert sich hier der Blick auf Pflegequalität einzig über den Indikator ‚Dekubitusinzidenz‘, was eine sehr reduzierte Sichtweise darstellt.

Die Messung von Ergebnisqualität in der Pflege hat im deutschsprachigen Raum kaum begonnen. Das sieht international anders aus. Seit den 1970er Jahren werden Anstrengungen unternommen, um Ergebnisindikatoren zu entwickeln, die geeignet sind, den Beitrag der Pflegeprofession zum Gesundheitsstatus, zur Sicherheit und zum Wohlbefinden von Menschen mit Pflegebedarf abzubilden. Die Bemühungen zur Outcomeforschung und –messung im Gesundheitswesen wurden Ende der 1980er in Medizin und Pflege intensiviert. Es entstanden u. a. Kategoriensysteme, die das Ziel hatten, die Ergebnisse pflegerischer Interventionen zu systematisieren (Moorhead, Johnson, & Mass, 2004). Die Entwicklung von Outcomeindikatoren ist anspruchsvoll. Nicht zuletzt zeigte sich, dass Outcomeindikatoren hinsichtlich ihrer Beeinflussbarkeit und Aussagekraft zu differenzieren sind. So folgten nicht zuletzt in den USA umfangreiche Bemühungen zur Entwicklung, Implementierung und Modifizierung differenzierter Systeme für die Qualitätsmessung u. a. mit besonderem Blick auf die Outcomemessung in der ambulanten und stationären Pflege. Andere Länder, auch im europäischen Raum wie Großbritannien, die Niederlande und Schweden, entwickelten ihrerseits eigene Verfahren und Instrumente für die Qualitäts- bzw. Outcomemessung in der Pflege.

Die Definition und Messung von Pflegequalität wird immer auch durch Kontext und Kultur beeinflusst. Für pflegebezogene Outcomemessung in Deutschland bedeutet dies, dass es zwar grundsätzlich möglich und sinnvoll ist, sich von internationalen Erfahrungen anregen zu lassen, dass eine unmittelbare Übertragung aber nicht zulässig ist.

1.1 Hintergrund

Mit dem Ziel die Outcomemessung in der stationären Altenhilfe in Deutschland auf den Weg zu bringen, erteilten das Bundesgesundheitsministerium (BMBF) und das Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) 2008 den Auftrag zur „Entwicklung und Erprobung von Instrumenten zur Beurteilung der Ergebnisqualität in der stationären Altenhilfe“. Der Auftrag wurde an das Institut für Pflegewissenschaft an der Universität Bielefeld (IPW) und das Institut für Sozialforschung und Gesellschaftspolitik GmbH (ISG) vergeben.

Das zu entwickelnde Instrumentarium sollte mehrere Anforderungen erfüllen. Es sollte geeignet sein für die vergleichende Qualitätsberichterstattung, integrierbar sein in das interne Qualitätsmanagement sowie einer externen Qualitätsprüfung grundsätzlich zugänglich sein. Das Projekt wurde von den beiden genannten Instituten in der Zeit von Dezember 2008 bis Dezember 2010 durchgeführt. In diesem Rahmen wurden potentielle Indikatoren zur Qualitätsmessung – überwiegend für die Outcomemessung – entwickelt und erprobt. Im Ergebnis wurden 29 Qualitätsindikatoren, verschiedene Bewertungssystematiken und Erhebungsinstrumente identifiziert. Der Abschlussbericht wurde im März 2011 vorgelegt (Wingenfeld et al., 2011).

Die von Wingenfeld et al. (2011) vorgeschlagenen Indikatoren und Verfahren wurden verschiedentlich öffentlich diskutiert, so am 19. Mai 2011 anlässlich des „Pflegeforums des MDS“ und am 11. September 2012 auf der gemeinsamen Fachtagung „Indikatoren für Ergebnisqualität in der stationären Pflege“ der Leistungserbringer und Leistungsträger. Weitere Diskussionsbeiträge finden sich in einem Statement von Hasseler (2011) sowie in einer Kurzexpertise zur Bewertung von Qualitätsindikatoren in der ambulanten und stationären Langzeitpflege. Diese Kurzexpertise widmete sich der Thematik mit besonderer Berücksichtigung der vom Institut für Pflegewissenschaft der Universität Bielefeld entwickelten und erprobten Instrumente zur Beurteilung der Ergebnisqualität in der stationären Altenhilfe. Die Kurzexpertise wurde 2012 im Auftrag der GKV erstellt (Stemmer & Arnold, 2012).

Der Diskurs brachte einige offene Fragen ans Tageslicht, die allerdings unterschiedlich gewichtet wurden. Angesichts der weiter bestehenden Unterschiede hinsichtlich der Einschätzung des Reifegrades des von Wingenfeld et al. (2011) vorgeschlagenen Instrumentariums (Indikatorenset, Bewertungssystematik, Risikoadjustierung) zeigt sich Klärungsbedarf, der in der vorliegenden Expertise, die wiederum vom GKV-Spitzenverband beauftragt wurde, bearbeitet werden soll.

1.2 Aufgabenstellung

Der Blick auf die international eingesetzten Verfahrensweisen und Instrumente zeigt, dass Outcomemessung in der Pflege recht unterschiedlich interpretiert werden kann. Die vorliegende Expertise diskutiert jedoch nicht die Bandbreite der Messmöglichkeiten, sondern bezieht sich ihrem Auftrag entsprechend unmittelbar und konkret auf von Wingenfeld et al. (2011) entwickelte Vorschläge.

Die von Wingenfeld et al. (2011) zur Anwendung empfohlenen Indikatoren und Ergebnisse werden verschiedenen Bereichen zugeordnet:

- Bereich 1: Erhalt und Förderung der Selbständigkeit (5 Indikatoren)
- Bereich 2: Schutz vor gesundheitlichen Schädigungen und Belastungen (6 Indikatoren)
- Bereich 3: Unterstützung bei spezifischen Bedarfslagen (4 Indikatoren)
- Bereich 4: Wohnen und (hauswirtschaftliche) Versorgung (4 Indikatoren)
- Bereich 5: Tagesgestaltung und soziale Betreuung (10 Indikatoren)
- Sonderbereich: Ergebnisse der Zusammenarbeit mit Angehörigen

Der Auftrag für die vorliegende Expertise bezieht sich ausschließlich auf die Bereiche 1-3. Auch für die Indikatoren dieser Bereiche geht es nicht um eine grundsätzliche Auseinandersetzung, sondern um die Beantwortung der vom Auftraggeber dezidiert ausformulierten Fragen. Diese Fragen sind inhaltlicher und methodischer Art und in fünf Themenbereiche gegliedert:

- Themenbereich 1: Reichweite der Indikatoren
- Themenbereich 2: Datengrundlage/ Externe Qualitätsprüfung
- Themenbereich 3: Bewertungssystematik
- Themenbereich 4: Risikoadjustierung
- Themenbereich 5: Verbraucherperspektive

Jedem Themenbereich sind mehrere Fragen zugeordnet. Diese bilden in Kapitel 4 die Überschriften und sind zusätzlich in der Anlage als Übersicht aufgeführt.

Die Auftragserwartung besteht aus drei Teilen:

1. Anhand einschlägiger (pflege-)wissenschaftlicher und qualitätsmethodischer Literatur soll eine Recherche bzgl. der ausgewiesenen Fragen vorgenommen werden.
2. Die vorgefundenen Quellen sind hinsichtlich der Fragen einer wissenschaftlichen Bewertung zu unterziehen.
3. Die Rechercheergebnisse sollen erläutert werden. Sofern die empirische Grundlage keine eindeutige Beantwortung der Fragen zulässt, ist diese auf der Grundlage des Expertenurteils abwägend zu diskutieren.

Der Auftrag zur „Durchführung einer Expertise zur Eignung von Indikatoren zur Messung und Darstellung von Ergebnisqualität in der stationären Pflege im Bereich der sozialen Pflegeversicherung“ wurde von der GKV im Dezember 2013 erteilt. Er wurde von Renate Stemmer, Katholische Hochschule Mainz und Jens Arnold, Centrum für angewandte Wirkungsforschung (CAW)/ IKJ ProQualitas GmbH, Mainz mit Unterstützung von Arne Evers (MScN), Matthias Müller (BScN) und Jana Wirz (M.A.) zwischen Dezember 2013 und Juni 2014 durchgeführt.

Nach einer kurzen Phase des Aufbaus von Projektstrukturen erfolgte die Literaturrecherche in einschlägigen Datenbanken sowie weiteren geeigneten Quellen (s. Kapitel 2) als Grundlage für die Erstellung der Expertise.

In den Themenbereichen und Fragestellungen überschneiden sich vielfach inhaltliche und statistisch-methodische Aspekte, so dass sie gemeinsam bearbeitet wurden. In der vorliegenden Expertise werden die pflegewissenschaftlich-inhaltlichen Ausführungen von Renate Stemmer, die statistisch-methodischen Erörterungen von Jens Arnold verantwortet.

1.3 Konzeptionelle Überlegungen

Die Definitionen von Qualität sind heterogen. Als gemeinsames Merkmal enthalten viele Definitionen jedoch den Hinweis auf die relative Natur von Qualität. Qualität bezeichnet nach der Definition des US-Institute of Medicine "das Ausmaß, in dem Gesundheitsleistungen für Individuen und Bevölkerungsgruppen die Wahrscheinlichkeit erwünschter gesundheitlicher Behandlungsergebnisse erhöhen und mit dem gegenwärtigen professionellen Wissensstand übereinstimmen" (zit. n. Ollenschläger, 2007).

Mit Blick auf ein Versorgungssystem und Versorgungsleistungen kann aus Qualitätssicht eine Dreiheit von Struktur, Prozess und Ergebnis differenziert werden. Klassischerweise werden der Struktur die System-, Mitarbeiter- und Patientencharakteristika zugeordnet, während der Prozess die technische Durchführung und die Art des interpersonellen Kontaktes reflektiert (Donabedian, 1980). Das Ergebnis kann als a) Versorgungsleistung und b) als Wirkung der Versorgungsleistung (häufig, aber nicht ausschließlich die Wirkung den Bewohner bzw. die Bewohnerin betreffend) verstanden werden. Der Output, die Versorgungsleistung, umfasst die Gesundheitsleistung und die Kontextleistung (Pfaff & Schrappe, 2011). Ein Beispiel für einen Output ist das Ausmaß der Erhebung vorhandener Schmerzen bei BewohnerInnen. Outputdaten reflektieren also die Leistungserbringung, machen aber keine Aussage über den bewohnerseitigen Nutzen dieser Leistungen. Damit sind im Hinblick auf die Erfassung von Versorgungsqualität die Outcomewerte stets aussagekräftiger als die Outputdaten.

Ein kundenbezogenes Outcome bildet sich in einer Verbesserung, Verschlechterung oder Stabilisierung eines Zustandes ab und kann vielfältig begründet sein (Stemmer, 2005). Es kann im Zusammenhang mit professionellen Interventionen, einem spontanen Verlauf oder weiteren Einflussgrößen stehen (Center for Medicare and Medicaid Services, 2011). Für die Outcomemessung sind jene Ergebnisse interessant, die - empirisch belegt - durch Gesundheitsdienstleistungen beeinflussbar, also versorgungssensitiv sind (Doran, Mildon, & Clarke, 2011). In Anlehnung an Moorhead et al. (2004) wird pflegesensitives Outcome verstanden als messbare Veränderungen des Gesundheitszustands, der Wahrnehmung und des Erlebens der BewohnerInnen, die durch die Unterstützung der Einrichtung bzw. durch das Handeln von professionellen Pflegekräften¹ bewirkt werden. Ergebnisqualität ist dementsprechend eine Eigenschaft von Versorgungsergebnissen, die mit einer bewertenden Aussage beschrieben wird.

1.3.1 Indikatorenentwicklung

Die Outcomemessung bedarf eines gut begründeten Messinstrumentariums. Zu unterscheiden sind Qualitätskriterien von Qualitätsindikatoren. Während Kriterien Eigenschaften definieren, deren Erfüllung typischerweise erwartet werden kann, wenn die Versorgungsleistung fach- und sachgerecht erfolgt (z. B. Wirksamkeit der Versorgungspraxis), stellen Indikatoren ein Maß dar, mit dem die Unterscheidung zwischen Ausprägungen z. B. des Nutzens erfasst werden soll (Expertenkreis Qualitätsindikatoren für Nationale VersorgungsLeitlinien, 2009).

¹ Angesprochen sind hier die professionell Pflegenden mit einer dreijährigen, zweijährigen oder einer einjährigen Pflegeausbildung sowie mit einem Pflegestudium mit klinischer Ausrichtung.

Qualitätsindikatoren können auf die Beschreibung der Pflege- und Versorgungsqualität für eine spezifische Gruppe oder die ganze Bevölkerung ausgerichtet sein. Es handelt sich also um Messgrößen, die ausgewählte Aspekte z. B. der Pflege und Versorgung retrospektiv abbilden. Nach ihren Häufigkeitsmaßen können Indikatoren auf der Grundlage aggregierter Daten von der Erfassung von Einzelereignissen (sog. sentinel events) unterschieden werden. Die Erfassung von Einzelereignissen kommt nur dann zum Tragen, wenn diese Ereignisse eine besondere Schwere darstellen (Expertenkreis Qualitätsindikatoren für Nationale VersorgungsLeitlinien, 2009). Pflegebezogene Outcomemessung nutzt in aller Regel aggregierte Daten.

Die Entwicklung von Qualitätsindikatoren, die eine zuverlässige Bewertung einer (Pflege-) Leistung ermöglichen, ist wissenschaftlich anspruchsvoll und bedarf einiger grundsätzlicher fachwissenschaftlicher Vorarbeiten. Dies zeigen die internationalen und nationalen Entwicklungen deutlich. Qualitätsindikatoren sind quantitative Instrumente, mittels derer bewertet werden soll, ob und in welchem Umfang Qualitätsziele erreicht wurden (AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH, 2010).

Eine Voraussetzung für zuverlässige Qualitätsindikatoren ist die Definition von Qualitätszielen und Qualitätskriterien, die ebenfalls quantitativ messbar sein müssen. International üblich ist es deshalb, Qualitätsindikatoren im Kontext von Standards und Leitlinien zu entwickeln. Sind zu Themen keine dieser Instrumente vorhanden, so müssen aus der vorliegenden Fachliteratur Kriterien zur Ableitung von Indikatoren herausgearbeitet werden.

Die Entwicklung von Qualitätsindikatoren folgt einer logischen Abfolge:

1. Klärung und Beschreibung der beabsichtigten Zwecke
2. Auswahl von Versorgungsaspekten (z. B. wegen Häufigkeit oder Schwere der unerwünschten Ereignisse, Verbesserungspotential)
3. Entscheidung über Qualitätskriterien (z. B. Zugänglichkeit, Wirksamkeit, Sicherheit)
4. Evidencebasierung und formale Konsentierung (auch Abwägung von Aufwand und Nutzen)
5. Ableitung von Qualitätsindikatoren aus Leitlinien, sofern vorhanden (Die methodische Qualität von Leitlinien und Standards erleichtert die Begründung und Ableitung von Indikatoren)
6. Strukturierte Einbeziehung von ExpertInnen, falls keine Leitlinien vorliegen und die Studienlage nur eine schwache Evidence ergibt
7. Evtl. Nutzung bestehender Indikatorensets (erfordert die Prüfung hinsichtlich der Passung)
8. Herstellung von Transparenz hinsichtlich des Entwicklungsprozesses und der Eigenschaften der Indikatoren
9. Umfassende Erprobung der Indikatoren (Expertenkreis Qualitätsindikatoren für Nationale VersorgungsLeitlinien, 2009; AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH, 2010; Kelley & Hurst, 2006)

10. Bei der Indikatorentwicklung sollten die Betroffenenverbände beteiligt werden (ActiZ, 2007; Kelley & Hurst, 2006).

Der beschriebene Prozess der Indikatorenentwicklung ist nicht nur aufwändig, sondern er greift auch auf vorhandene Strukturen, insbesondere auf Leitlinien und Standards, zurück. Für die Entwicklung von Outcomeindikatoren in der stationären Altenhilfe liegen diese jedoch nur punktuell vor. Dies stellt den Entwicklungsprozess für Outcomeindikatoren vor erhebliche Herausforderungen. Im beschriebenen Projekt gilt dies umso mehr, da Wingenfeld et al. (2011) für den gesamten Prozess nur zwei Jahre zur Verfügung standen. Angesichts dieser Rahmenbedingungen ist das Projektergebnis nachdrücklich zu würdigen. Gleichwohl erfordert gerade das noch weitgehend unbeackerte Feld die sorgfältige Klärung der übergeordneten Qualitätskriterien sowie der indikatorbezogenen Qualitätsziele. Beides fehlt jedoch.

Qualitätskriterien können auch politisch vorgegeben werden. Wenn dies nicht auf diesem Weg geschieht, sind sie zu Beginn des Entwicklungsprozess aber zu klären, da die Indikatoren entsprechend auszurichten sind. Dieser Klärungsprozess ist bei Wingenfeld et al. (2011) leider nicht ersichtlich. Es ist zu vermuten, dass Qualitätskriterien implizit einfließen. Da sie jedoch nicht explizit ausgewiesen sind, liegt in der Konsequenz mit dem Vorschlag von Wingenfeld et al. (2011) ein Entwurf für ein Indikatorenset vor, dem rahmensetzende Qualitätskriterien fehlen.

Dies gilt in gleicher Weise für die Qualitätsziele. Auch diese finden sich nicht. Jedoch fehlen den Indikatoren ohne den Ausweis von Qualitätskriterien und -zielen die Verankerung – die Auswahl und Ausrichtung erscheint vor diesem Hintergrund in gewisser Weise willkürlich.

1.3.2. Interne Qualitätsentwicklung, Qualitätsberichterstattung externe Qualitätsprüfung

Wingenfeld et al. (2011) wurden beauftragt, *ein* Instrumentarium für *drei* Ziele zu entwickeln: die vergleichende Qualitätsberichterstattung, die Eignung dieses Instrumentariums für die interne Qualitätsentwicklung sowie für die externe Qualitätsprüfung.

Wenn beachtet wird, dass die Bedürfnisse und Interessen der VerbraucherInnen und der Kontrollorgane bei der Qualitätsberichterstattung nicht identisch sind, handelt es sich hier sogar nicht nur um drei, sondern um vier Ziele.

Aus Gründen der Begrenzung administrativer Belastungen findet sich diese Kombination auch international, z. B. in Initiativen der Niederlande (ActiZ et al., 2007) und Australien (National Aged Care Alliance, 2014). Gleichwohl sind die damit verbundenen Herausforderungen nicht zu übersehen und bedürfen einer sorgfältigen Beachtung.

Verfahren der internen Qualitätsentwicklung zielen ganz wesentlich darauf ab, die Prozesse der Qualitätsentwicklung innerhalb der Einrichtung zu steuern. Dazu sind differenzierte Kenntnisse über das Zusammenspiel von Strukturen, Prozessen und Ergebnissen erforderlich. Anders als von Donabedian ursprünglich angenommen (Donabedian, 1980), ist keineswegs von einem linearen Bedingungsverhältnis zwischen den genannten Qualitätsdimensionen auszugehen. Stattdessen zeigt sich eher eine Beziehung, die auf Gegenseitigkeit beruht. Allerdings kann hierbei kein unmittelbarer Zusammenhang zwischen einer professionellen Intervention und dem bewohnerbezogenen Ergebnis erwartet werden (Mitchell, Ferketich, &

Jennings, 1998). Vielmehr werden die Wirkungen immer durch einrichtungs-, situations- und bewohnerbezogene Aspekte moduliert. Damit sind die im Rahmen der internen Qualitätsentwicklung errechneten zahlenmäßigen Ergebnisse nur bedingt aussagekräftig. Stattdessen stellen sie die Ausgangslage für einen Analyseprozess dar, der im Wesentlichen die Aufgabe hat, Bedingungsverhältnisse aufzudecken. Erst auf der Grundlage dieses Interpretationsprozesses sind Steuerungsmaßnahmen zu entwickeln. Um diesen Prozess anzustoßen, sind Datenerhebungen auch für die interne Qualitätsentwicklung durchaus sinnvoll (Europäisches Zentrum für Wohlfahrtspolitik und Sozialforschung et al., 2010).

Die Qualitätsberichterstattung dient demgegenüber hauptsächlich dem Vergleich zwischen Einrichtungen.

Wenn im Rahmen der internen Qualitätsentwicklung die erhobenen Zahlenwerte als Interpretationsgrundlage dienen, aber selbst nur bedingt relevant sind, sieht dies bei der externen Qualitätsberichterstattung ganz anders aus. Hier geht es in zentraler Weise um die zahlenmäßig erfassten Ergebnisse. Diese bilden die Grundlage für die externe Einschätzung der Pflegequalität und damit auch für das Benchmarking der Einrichtungen. Dies hat erhebliche Konsequenzen. Rein methodisch betrachtet ist eine Risikoadjustierung erforderlich, um die unterschiedlichen Bewohner- und Einrichtungsspezifika zu modulieren. Darüber hinaus bekommen nun die ausgewählten Indikatoren, Bewertungssystematiken, Verfahren der Risikoadjustierung etc. eine besondere Bedeutung, denn je nach deren Auswahl kann die Bewertung einer Einrichtung und deren Platz im Ranking beeinflusst werden.

Auch die Qualität der Daten erhält nun eine besondere Bedeutung. Die Rohdaten werden von den Einrichtungen geliefert. Diese müssen exakt, zuverlässig und fristgerecht erhoben werden - Anforderungen, die sowohl für die Einrichtungen als auch für die Durchführbarkeit der Outcomemessung nicht banal sind, wie dies auch Erfahrungen zur Qualitätsberichterstattung im Krankenhaus zeigen (AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH, 2013).

Ein weiterer Punkt betrifft die Anzahl der Indikatoren. Für die Zielsetzung der internen Qualitätsentwicklung ist eine größere Anzahl von Indikatoren zielführend und möglich als für die externe Qualitätsberichterstattung. Andersherum werden die Indikatoren, die für die externe Berichterstattung genutzt werden, selten für die interne Qualitätsentwicklung ausreichen. Vor diesem Hintergrund räumen die AutorInnen der australischen ‚Allianz für Versorgung und Pflege im Alter‘ ein, dass es schwierig sein kann, die Zielsetzung der mehrfachen Nutzung (interne und externe Qualitätsentwicklung und –sicherung) für jeden Indikator zu erreichen (National Aged Care Alliance, 2014).

Zum dritten Thema, der Eignung von Outcomeindikatoren für die externe Qualitätsprüfung: Derzeit findet externe Qualitätsprüfung in der stationären Altenpflege mittels Vor-Ort-Begehungen statt. Mit dem Ziel der Prüfung von Prozessqualität werden nicht nur Dokumente gesichtet, sondern auch das Gespräch mit BewohnerInnen geführt und stichprobenartig deren Pflegezustand verifiziert.

Bei den Daten für eine outcomeorientierte Qualitätsberichterstattung handelt es sich um aggregierte Daten. Diese können jedoch nicht unmittelbar überprüft werden. Eine Beobachtung z. B. der Dekubitusinzidenz oder einer Sturzrate mit schwerwiegenden Folgen ist im Rahmen einer Vor-Ort-Begehung nicht möglich, sehr wohl aber die Diskussion der Ergebnisse der Outcomemessung. Denkbar sind darüber hinaus eine Vor-Ort-Überprüfung der Verfahren der

Datenerhebung oder der Maßnahmen zur Sicherung der Datenqualität sowie die Erhebung zusätzlicher Struktur- und Prozessfaktoren.

2 Methodisches Vorgehen

(Arne Nils Evers & Renate Stemmer)

2.1 Vorgehen bei der Recherche

Zentrale Grundlage für die Erstellung der Expertise war eine umfangreiche Literaturrecherche. Es wurden zwei verschiedene Recherchen durchgeführt:

1. Internationale Recherche bei staatlichen Einrichtungen des Gesundheitswesens, Aufsichtsbehörden und sonstigen Institutionen oder Verbänden
2. Recherche nach Studien über Fachdatenbanken, Datenbankportale und unterschiedliche Benutzeroberflächen.

Für das Quellenmanagement wurde ein Literaturverwaltungsprogramm genutzt. Dieses unterstützte die Zusammenführung der Rechercheergebnisse, deren Codierung und Weiterbearbeitung im Projektteam und sicherte so die Übersichtlichkeit und Nachvollziehbarkeit der Rechercheergebnisse für alle Beteiligten.

Das Vorgehen bei den Recherchen und die weiteren Bearbeitungsschritte werden nachfolgend erläutert.

2.1.1 Recherche internationaler Einrichtungen des Gesundheitswesens

Qualitätsmessung in der Pflege ist auch international seit Jahre ein Thema. Teilweise werden von Ministerien, Versicherungen oder anderen Institutionen die Entwicklung von Qualitätsindikatoren in Auftrag gegeben oder deren Entwicklung begleitet. Nicht immer werden die so generierten Indikatoren und Instrumente in wissenschaftlichen Zeitschriften publiziert, so dass zusätzlich zur Fachdatenbankrecherche eine Recherche dieser Quellen erforderlich ist. Hierfür wurde gezielt auf den Internetseiten staatlicher Aufsichtsbehörden im Gesundheitswesen, Gesundheitsministerien und -behörden sowie großer Versicherungen für folgende Länder recherchiert:

- Australien, Belgien, Dänemark, Großbritannien, Irland, Japan, Kanada, Neuseeland, Niederlande, Norwegen, Österreich, Schweden, Schweiz, Spanien, USA

Rechercheergebnisse wurden nur dann eingeschlossen, wenn diese in englischer oder deutscher Sprache verfügbar waren und einen Bezug zur Outcomemessung bzw. zu Qualitätsindikatoren, vorzugsweise in der Langzeitpflege, aufgewiesen haben. In die Auswertung wurden Berichte, Projektergebnisse, Studien, Standards, Positionspapiere und Empfehlungen aufgenommen.

Ziel dieser Recherche war es, die Prinzipien, Vorgehensweisen und Erfahrungen in Gesundheitssystemen zu eruieren, die im weiteren Sinne mit dem deutschen System vergleichbar sind. Insbesondere die Verwendung selbstentwickelter Instrumente und Indikatoren oder der Rückgriff auf bestehende Instrumente und Indikatoren erschien lohnenswert, um einen Ver-

gleich zwischen den Ländern herzustellen und diese Erfahrungen für die Bearbeitung der Expertise verwertbar zu machen.

Im Ergebnis wurden 49 Dokumente aus dieser Recherche ausgewertet bzw. verwendet. Für Spanien und Japan endete die Suche in einer Sackgasse, da keine englischsprachigen Quellen auffindig gemacht werden konnten.

2.1.2 Recherche in Fachdatenbanken und Datenbankportalen

Die Datenbankrecherche diente der Suche nach Studien und Reviews im Kontext der Entwicklung und Überprüfung von versorgungs- sowie pflegesensitiven Outcomeindikatoren bzw. Indikatorensets für die interne Qualitätsentwicklung, die externe Qualitätsprüfung sowie für die Qualitätsberichterstattung im Kontext der stationären Pflege. Die Erstellung des Reviews orientierte sich in methodischer Hinsicht an gängigen Vorgaben zur Erstellung eines systematischen Reviews (Kunz, Kalid, Kleinen, & Antes, 2009). Da hier aber die Fragestellung breiter als in einem klassischen Review zur Darstellung der Evidenz einer spezifischen Intervention angelegt war, zog dies Konsequenzen für die Recherche nach sich, da diese mehrfach mit unterschiedlichen Begriffskombinationen angegangen werden musste.

Zur Strukturierung des Vorgehens wurde die Auswertung schrittweise vorgenommen (Abbildung 1).

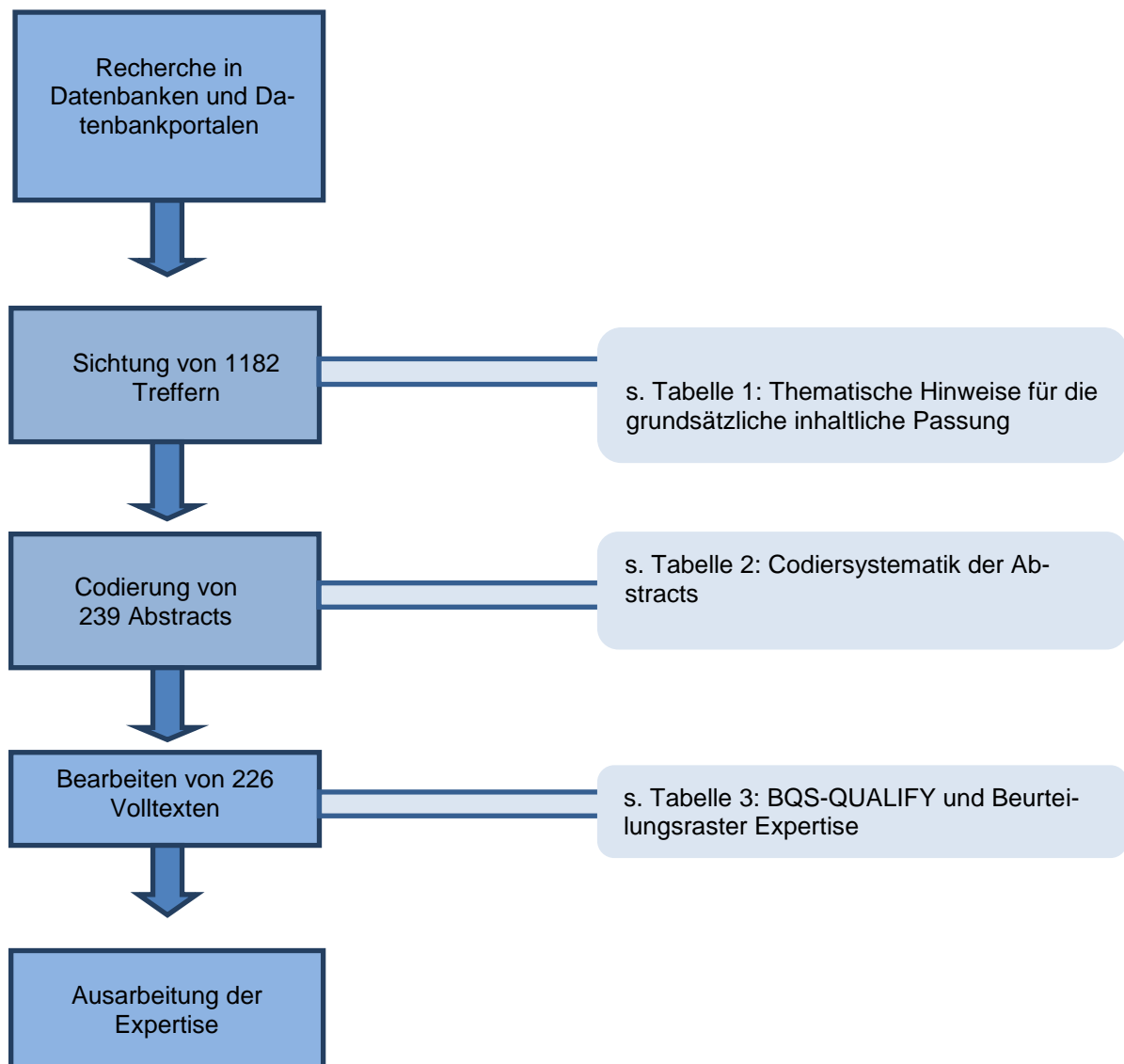
Initial erfolgte die Recherche in Datenbanken und Datenbankportalen unter Nutzung verschiedener Benutzeroberflächen. Im weiteren Verlauf wurden die ermittelten Quellen verschiedenen Zwischenauswertungen unterzogen. Diese werden im Folgenden erläutert.

Das Ziel dieser Recherche war das Erschließen von Abstracts/ Studien zur Bearbeitung der in dieser Expertise aufgeworfenen Fragestellungen. Für die Recherche wurden die nachstehenden Datenbanken, Datenbankportale und Benutzeroberflächen genutzt: Medline, CINAHL, PubMed, Cochrane Library, HECLINET, Psycinfo, Tripdatabase, Google Scholar, Psynindex, DIMDI, Medpilot, OVID.

Zu Beginn der Literaturrecherche wurde eine Sammlung relevanter Suchbegriffe erstellt und soweit möglich in MESH-Terms übersetzt. Als MESH-Terms wurden u. a. *outcome assessment (health care)*, *nursing homes*, *nursing outcomes*, *quality indicators (health care)* genutzt; Stichworte, die in Google scholar verwendet wurden, umfassten u. a. *outcome assessment*, *nursing homes*, *quality indicators*, *resident outcomes*, *quality reporting indicators*, *quality measure*, *care quality indicator*, *nursing sensitive quality indicator*.

Die Suchbegriffe wurden mit Bool'schen Operatoren verbunden. Eingeschlossen wurden englisch- und deutschsprachige Quellen der letzten zehn Jahre. Aus der Recherche heraus wurden weitere Suchbegriffe erschlossen und verwendet.

Abbildung 1: Vorgehen Datenbankrecherche



Die Treffer wurden auf grundsätzliche inhaltliche Eignung gesichtet.

Tabelle 1: Thematische Hinweise für die grundsätzliche inhaltliche Passung

Screeningkriterien für Abstracts
Qualitätsmessung im Setting der stationären Altenhilfe
Thematisierung von interner/ externer Qualitätsmessung
Entwicklung von Instrumenten zur Qualitätsmessung oder Überprüfung bekannter Instrumente mit Eignung für das Setting der stationären Altenhilfe
Diskussion von Qualitätsverbesserungs- oder –sicherungsprogrammen für das Setting der stationären Altenhilfe
Pflegebezogene Standards und Leitlinien
Verfahren der Erhebung

Es zeigte sich, dass sich viele Quellen auf die Messung von Ergebnissen beziehen, die die Wirksamkeit einer spezifischen Intervention untersuchen oder ein anderes Setting als die Langzeitpflege im Blick haben.

Zur Aufnahme für die weitere Bearbeitung war jedoch insbesondere die Passung auf das Setting der stationären Altenhilfe in Abgrenzung zur Versorgung in Krankenhäusern oder ambulanten Pflegeeinrichtungen bedeutsam. Weiterhin musste die Thematisierung von Qualitätsindikatoren erkennbar sein. Es wurden auch Abstracts aufgenommen, die ein bereits entwickeltes Verfahren untersuchten. Ebenfalls wurden solche Abstracts akzeptiert, die Qualitätsverbesserungs- oder Qualitätssicherungsprogramme beschreiben oder eigenständige Standards bzw. Leitlinien entwickeln und den Vorgang der Anwendung erläutern.

Nach einer Prüfung auf Duplikate wurden die vorhandenen Abstracts nun codiert, um sie verschiedenen Schwerpunkten (siehe Tabelle 2) zuordnen zu können. Das Ziel dieses Vorgehens war die Systematisierung der Abstracts – und daraus folgend auch die der Volltexte – um die Auswertung zu strukturieren und zu vereinfachen. Die Codierung wurde ins Literaturverwaltungsprogramm übertragen und so konnte fortan gezielt thematisch gesucht werden. In diesem Arbeitsschritt wurden noch einmal einige Quellen aussortiert, da diese sich bei der genauen Betrachtung als ungeeignet erwiesen haben.

Der erste Schwerpunkt „Ebene“ untergliedert sich in „Organisation“ und „nicht näher bezeichnet“. Dieser Schwerpunkt dient dazu, Abstracts zu identifizieren, die sich mit der Umsetzung oder Evaluierung von Outcomemessung innerhalb einer Organisation oder eines Systems auseinandersetzen. Insbesondere wurden hier Studien ausgeschlossen, die Einzelmaßnahmen überprüfen. Als geeignet wurden Studien eingestuft, welche z. B. die Implementierung eines Messinstruments in einer Altenpflegeeinrichtung beschreiben.

Der zweite Schwerpunkt „Indikatorthemen“ zielt auf die Zuweisung von potenziellen Outcomeindikatoren zu den Abstracts ab. Die benannten Codes in diesem Schwerpunkt entsprechen Indikatoren, die zur Ergebnisqualitätsmessung genutzt werden können. Der Code „Sonstiges Thema“ wurde für relevante Aspekte, die bei der Codierung entdeckt und bisher nicht aufgezählt waren, vorgehalten, um nachfolgend überprüfen zu können, ob diese sich zur Anwendung als Indikator eignen.

Der dritte Schwerpunkt „Grundlage“ differenziert nach der Art und Weise der Fundierung von Indikatoren. Der Code „Empirische Grundlage“ wurde immer dann zugewiesen, wenn die Nutzung statistischer Verfahren im Rahmen der Entwicklung oder Überprüfung von Indikatoren ersichtlich ist. Weiterhin gibt es den Code „Leitlinienbezug“, der immer dann benutzt wurde, wenn innerhalb eines Abstracts der Bezug auf eine Leitlinie oder einen Standard zur Entwicklung oder Anwendung eines Ergebnisqualitätsindikators oder eines -sets erkennbar war. Der dritte Code „Gütekriterien“ verweist auf die explizite Diskussion eben jener in den Studien. Häufig handelt es sich um die Frage der Reliabilitäts- und Validitätsprüfung. Sind Aussagen in den Abstracts erkennbar gewesen, die ein Berechnungsverfahren zur Bewertung von Ergebnismaßen beschreiben, ist dies mit dem Code „Bewertungssystematik“ gekennzeichnet worden.

Die „Risikoadjustierung“ wurde als ein weiteres Differenzierungsmerkmal verwendet, da die Angemessenheit verschiedener Risikoadjustierungen eine hohe Relevanz für die Bearbeitung der aufgeworfenen Fragestellungen darstellt.

Tabelle 2: Codiersystematik der Abstracts

1. Schwerpunkt: Ebene	2. Schwerpunkt: Indikatorthemen	3. Schwerpunkt: Grundlage
Organisation	Dekubitus	Empirische Grundlage
nicht näher bezeichnet	Sturz	Leitlinienbezug
	Mobilität	Gütekriterien
	Kontraktur	Bewertungssystematik
	Harnwegserkrankungen	Risikoadjustierung
	Lebensqualität	nicht näher bezeichnet
	Arzneimittel	
	Krankenhauseinweisung	
	Schmerz	
	Wunden	
	Ernährung	
	Grad der Selbständigkeit	
	Sonstiges Thema	

Nachdem diese Bearbeitungsschritte durchlaufen und weitere ungeeignete Abstracts ausgeschlossen worden waren blieb eine Gesamtzahl von 239 berücksichtigten Abstracts. Für diese wurden nun die Volltexte über die Elektronische Zeitschriftenbibliothek sowie, falls kein Zugang vorhanden war, über eine Bestellung bei Subito bezogen. Für 13 Studien waren die Volltexte auf diesem Weg nicht zugänglich, so dass im Ergebnis insgesamt 226 Volltexte erschlossen werden konnten. Diese mussten nun für eine detaillierte Auswertung erneut beurteilt werden.

Dafür wurde auf ein 21 Codes umfassendes Beurteilungsraster zurückgegriffen, welches eine erweiterte Form des Instruments „QUALIFY“ darstellt (Reiter et al., 2014). QUALIFY ist ein Instrument zur strukturieren Bewertung von Qualitätsindikatoren im Gesundheitswesen. Es wurde im Auftrag der Bundesgeschäftsstelle für Qualitätssicherung (BQS) entwickelt. Dieses Instrument wurde im Hinblick auf die Fragestellung der Expertise und die pflegerische Perspektive erweitert, um die pflegewissenschaftliche Fundierung von Qualitätsindikatoren abbilden und Verfahrensfragen beantworten zu können (siehe Tabelle 3).

Tabelle 3: BQS-QUALIFY und Beurteilungsraster Expertise

BQS-QUALIFY	Beurteilungsraster zur Auswertung der Volltexte und Beurteilung der Indikatoren dieser Expertise
	Pflegewissenschaftliche Fundierung
Risikoadjustierung	Risikoadjustierung
Indikatorevidenz	Indikatorevidenz
Klarheit der Definitionen (des Indikators und seiner Anwendung)	Klarheit der Definitionen (des Indikators und seiner Anwendung)
Reliabilität	Reliabilität
Statistische Unterscheidungsfähigkeit	Statistische Unterscheidungsfähigkeit
Sensitivität	Sensitivität
Spezifität	Spezifität
Validität	Validität
	Messmethoden / Verfahren zur Datenerhebung
	Umsetzung in der Routine, Modellprojekte
	Anforderungen an die Datenerhebung
Erhebungsaufwand	Erhebungsaufwand
Datenverfügbarkeit	Datenverfügbarkeit
	Übertragbarkeit auf das deutsche Pflegesystem
Verständlichkeit für Patienten und interessierte Öffentlichkeit	Verständlichkeit für Menschen mit Pflegebedarf und interessierte Öffentlichkeit
Verständlichkeit für Ärzte und Pflegende	Verständlichkeit für professionell Pflegende
	Übersichtlichkeit
	Eignung für vergleichende Qualitätsberichtserstattung
Beeinflussbarkeit der Indikatorausprägung	Beeinflussbarkeit der Indikatorausprägung
Überprüfbarkeit (Richtigkeit, Vollständigkeit, Vollzähligkeit)	Überprüfbarkeit (Richtigkeit, Vollständigkeit, Vollzähligkeit)

Fand die Codierung bisher auf der Ebene der Abstracts statt, so wurden nun alle Volltexte direkt im Dokument codiert. Jedes Kriterium des Beurteilungsrasters erhielt eine Nummer, die der Reihenfolge in Tabelle 3 entspricht. Die Volltexte wurden gelesen, passende Textstellen zu einem Kriterium markiert und der Zahlencode des Kriteriums hinterlegt. So erhielt beispielsweise eine Studie, die Aussagen zur „Risikoadjustierung“, „Indikatorevidenz“ und zu „Messmethoden/ Verfahren zur Datenerhebung“ umfasste, die Zahlencodes 02, 03 und 10. Die sich so ergebenden Zahlenketten für ein Dokument wurden in dem erneut dafür angepassten Literaturverwaltungsprogramm hinterlegt.

Zur Veranschaulichung sollen im Folgenden einige Auswertungscodes beispielhaft erläutert werden. So wurde der Code „Pflgewissenschaftliche Fundierung“ dann vergeben, wenn Aussagen zu einer Fundierung mittels pflgetheoretischer Annahmen oder anderer, eindeutiger Bezüge zu pflgewissenschaftlichen Begründungszusammenhängen getroffen wurden.

Wie bereits oben erwähnt, wurden erste Aussagen zu Gütekriterien bereits in den Abstracts codiert. Damit dieses Feld weiter differenziert werden konnte, wurden die Aspekte „Risikoadjustierung“, „Indikatorevidenz“, „Klarheit der Definitionen“ (des Indikators und seiner Anwendung), „Reliabilität“, „statistische Unterscheidungsfähigkeit“, „Sensitivität“, „Spezifität“ und „Validität“ in den Volltexten jeweils einzeln markiert und im jeweiligen Dokument kenntlich gemacht.

Der Code „Messmethoden/ Verfahren zur Datenerhebung“ greift Verfahren zur Erhebung der Daten auf. Berücksichtigt wurden Hinweise darauf, ob beispielsweise Kundenbefragungen durchgeführt wurden, ob eine Einschätzung mittels eines dafür entworfenen Fragebogens oder andere Messungen zur Datengewinnung stattgefunden haben. Auch Aussagen, welche die Praktikabilität einzelner Erhebungsmethoden beschreiben, wurden erfasst.

Unter „Umsetzung in der Routine/ Modellprojekte“ wurden Aussagen codiert, die strukturelle Voraussetzungen beschreiben oder Aussagen zu Erfahrungen mit Modellprojekten in Bezug auf eine Outcomemessung zulassen.

Wurde der Einsatz oder die Entwicklung einer speziellen Datenbank oder Aussagen über den Erhebungsaufwand von Daten diskutiert, wurde der Code „Anforderungen an die Datenerhebung“ hinterlegt. Das heißt, dieser Code wurde immer dann benutzt, wenn Anforderungen struktureller und/ oder inhaltlicher Art beschrieben werden, wie z. B. das Vorhandensein und die Implementierung eines Formblattes zur Erhebung von Indikatoren.

Des Weiteren wurden auch Passagen markiert, die sich explizit mit dem Erhebungsaufwand oder der Datenverfügbarkeit befassen. Aspekte, die hier zugeordnet wurden, betreffen u. a. die Möglichkeit des Rückgriffs auf pflegerische Routinedaten oder deren Anpassung bzw. Erweiterung für Zwecke der Qualitätsmessung.

Der Code „Übertragbarkeit auf das deutsche Pflegesystem“ beschreibt Erkenntnisse, die einen Rückschluss auf mögliche Herausforderungen zulassen, die mit dem potentiellen Einsatz international genutzter Instrumente und Verfahren zur Qualitätsmessung in Deutschland einhergehen können bzw. welche Anpassungen ggf. erforderlich sind.

Das BQS-QUALIFY kennt ein Kriterium, das zwischen der Verständlichkeit für PatientInnen, für die interessierte Öffentlichkeit, für ÄrztInnen und Pflegefachkräfte differenziert. Dieses Kriterium wurde dahingehend interpretiert, dass mit dem Begriff „Patient“ auch Menschen mit Pflegebedarf, die sich beispielsweise in der ambulanten Versorgung befinden, aber einen Heimeinzug planen, erfasst werden. Die professionelle Seite wurde mit „professionell Pflegenden“ konkretisiert, da die Ergebnisqualitätsindikatoren die Versorgung und Pflege in den Blick nehmen und einer ärztlichen Ergebnisqualität andere Maßstäbe zuzuordnen sind.

Der Code „Übersichtlichkeit“ ist in der Zielsetzung von Qualitäts- und Outcomemessungen begründet, Indikatoren mit einer hohen Aussagekraft zu identifizieren und zu erheben, die in der Lage sind, sowohl die Kontrollbehörden als auch die VerbraucherInnen zu informieren. Die Erfassung einer hohen Anzahl von Indikatoren birgt immer auch ein erhebliches Fehler-

potenzial, so dass möglicherweise Aufwand und Nutzen nicht in einem angemessenen Verhältnis zueinander stehen.

Weiterhin stellt sich die Frage, ob die Ergebnisqualitätsindikatoren für eine vergleichende Berichterstattung geeignet sind. Hier sind Texte hinterlegt, die die Aussagekraft der Indikatoren, die methodische Güte, aber auch die Praktikabilität der Datenerhebung für die vergleichende Berichterstattung diskutieren.

Waren eindeutige Aussagen zu einer pflegerischen Beeinflussung der Indikatorausprägung abzuleiten, wurde der Punkt „Beeinflussbarkeit der Indikatorausprägung“ hervorgehoben. Dazu wurden ergänzend Bezüge zu einer grundsätzlichen Beeinflussbarkeit von Indikatoren durch pflegerische Interventionen sowie auch strukturelle Voraussetzungen markiert.

Der letzte Code wurde gemäß des BQS-QUALIFY verwendet, das heißt, ob eine Richtigkeit und Vollständigkeit der Daten generell überprüft werden kann.

Anhand der Codierung der Abstracts nach Schwerpunkten und der Volltexte nach den genannten Kriterien konnte eine strukturierte Auswertung der Dokumente für die Beantwortung der Fragen dieser Expertise erfolgen.

3 Zusammenfassender Überblick über pflegebezogene Outcomemessung im internationalen Raum

(Jana Wirz & Renate Stemmer)

In vielen Ländern gibt es eine Reihe von Bemühungen, die Messung von Pflegequalität in unterschiedlichen Settings zu verbessern. Auch wenn diese Anstrengungen häufig rechtlich forciert werden, d.h. dass international betrachtet Fragen der Pflegequalität immer mehr in den Fokus der nationalen Gesetzgebung rücken, werden im länderspezifischen Kontext unterschiedliche Wege zur Ausgestaltung bestimmter Regelungen beschritten. So werden neben Qualitätsindikatoren Standards, Maßzahlen und Benchmarks definiert, denen wiederum die Zielsetzung, eine ausreichende Pflegequalität in Einrichtungen zu gewährleisten und nachzuweisen, gemein ist.

Diese verschiedenen Entwicklungen haben zu länderspezifischen Vorgehensweisen geführt, was das Instrumentarium bzw. die Verfahren zur Messung von Pflegeergebnissen betrifft. Gleichzeitig ist aber zu beobachten, dass insbesondere das Minimum Data Set (MDS) des Resident Assessment Instruments (RAI) weit verbreitet ist. So werden u. a. in Kanada, den USA und Irland Pflegeoutcomes in leistungserbringenden Einrichtungen häufig mit einem jeweils angepassten Inter-RAI-MDS-Instrument gemessen. Die hier erhobenen Daten werden für interne und externe Qualitätsbeurteilungen genutzt.

Neuere Entwicklungen zielen in vielen Ländern auf öffentliche Berichterstattungen ab, die insbesondere potentiellen Leistungsnutzerin und -nutzerinnen sowie ihren Angehörigen bei der Wahl der richtigen Pflegeheime Hilfestellungen geben sollen (Canadian Institute for Health Information, 2013; Health Information and Quality Authority, 2009; National Quality Forum, 2011).

Mit den nationalen Traditionen gehen verschiedene Auslegungen des Qualitätsbegriffs einher – eine einheitliche, über Ländergrenzen hinweg gültige Definition fehlt. So verstehen die australischen Gesundheitsbehörden Qualität als das Ausmaß, zu dem Gesundheitsdienstleister oder ein Produkt das gewünschte Outcome herstellen, während irländische VertreterInnen Qualität zusätzlich in Beziehung zum aktuellen Stand des professionellen Wissens setzen. Weitgehend übereinstimmend finden sich ferner Dimensionen wie Effektivität, Gerechtigkeit, Sicherheit und Personenzentriertheit als Aspekte von Qualität (Australian Institute of Health and Welfare, 2009; Department of Health, 2012; Health Information and Quality Authority, 2013)

Die Entwicklung der entsprechenden Qualitätsmaße wird von unterschiedlichen Playern betrieben. Dazu gehören staatliche Behörden, Trägerverbände, Berufsverbände, Versicherungen und Vertreterorganisationen von Betroffenen sowie Konsortien, in denen die genannten Gruppen in wechselnder Zusammensetzung repräsentiert sind.

Für die nachfolgende Betrachtung wurden insbesondere Qualitätsindikatoren für Alten- bzw. Langzeitpflegesettings für folgende Länder recherchiert: Australien, Neuseeland, Dänemark, Großbritannien, Irland, Kanada, Niederlande, Norwegen und USA. Für Neuseeland und Dänemark wurden keine Indikatoren gefunden, die sich speziell auf Pflegeheime beziehen, sondern nur solche, die im gesamten Gesundheitssektor zum Einsatz kommen.

Wie bereits angedeutet, werden nicht immer Qualitätsindikatoren ausgewiesen, sondern auch Maßzahlen, Standards und Benchmarks, die in die zusammenfassende Tabelle 4 ebenfalls aufgenommen wurden, da sie relevante Bereiche von Pflegeoutcomes definieren und abbilden. Betrachtet man die zusammengetragenen Qualitätsindikatoren, -standards, -benchmarks und Maßzahlen (s. Tabelle 4), so fällt auf, dass folgende Bereiche häufig als relevant für die Outcomemessung definiert werden:

- Dekubitus
- Stürze
- Medikation
- Schmerzen
- Lebensqualität/ Wohlbefinden
- Inkontinenz
- Aktivitäten des täglichen Lebens
- Ernährung
- Fixierung

Daneben werden eine Reihe weiterer Indikatoren verwendet, die sich nicht nur auf die Ergebnisqualität beziehen, sondern auch die Strukturqualität, z. B. über Mitarbeiterprofile, abbilden.

Tabelle 4: Qualitätsmessung nach Ländern²

Land	Bereiche der Qualitätsmessung	Quelle
Australien	<ul style="list-style-type: none"> • Mundgesundheit • BewohnerInnen, die eine Überprüfung der Medikation erhalten • Mangelernährung • Dekubitusgeschwüre • Stürze, die zu Schäden bei BewohnerInnen führen • Erfahrungen der BewohnerInnen • Das Vorhandensein von geeigneten Verfahren zur fortlaufenden Dokumentation und Beobachtung von Zwischenfällen 	<i>(Australian Institute of Health and Welfare, 2009)</i>
Dänemark	<ul style="list-style-type: none"> • Informationen zur Organisation (u. a. mit Berichten zur geriatrischen Abteilung und dem Personal), • Funktionsbezogene Daten (Body-Mass-Index, Aktivität des täglichen Lebens) • Stammdaten zu den Klienten und KlientInnen 	<i>(Sektion for Klinisk Information & Data, 2014)</i>
Großbritannien	<p>Lebensqualität (bei langdauerndem Versorgungsbedarf)</p> <p>umfassender Indikator</p> <ul style="list-style-type: none"> • gesundheitsbezogene Lebensqualität <p>Verbesserungsbereiche</p> <ul style="list-style-type: none"> • Sicherstellen, dass Menschen sich unterstützt fühlen • Funktionale Fähigkeit • Reduzieren der Krankenhausaufenthalte • Sicherstellen der Lebensqualität der Angehörigen • Sicherstellen der Lebensqualität der Menschen mit mentalen Erkrankungen 	<i>(Department of Health, 2012)</i>

² In den länderspezifischen Konzepten differieren die Bezeichnungen für die Menschen mit Pflegebedarf. Die jeweilige Bezeichnung KlientIn, NutzerIn, Bewohnerin etc. wurde jeweils beibehalten.

Land	Bereiche der Qualitätsmessung	Quelle
	<ul style="list-style-type: none"> • Sicherstellen der Lebensqualität der Menschen mit Demenz <p>Sichere Umwelt (unabhängig vom Setting im Gesundheitswesen)</p> <p>umfassende Indikatoren</p> <ul style="list-style-type: none"> • Vorfälle, die die Patientensicherheit betreffen • Vorfälle, die einen Schaden oder Tod verursacht haben • Todesfälle, die Pflegeproblemen zugeschrieben werden <p>Verbesserungsbereiche</p> <ul style="list-style-type: none"> • Reduzieren der Vorfälle von vermeidbaren Schäden (Medikationsfehler, Infektionen, Druckgeschwüre etc.) 	
Großbritannien	<p>Fokus auf vier Bereiche:</p> <ul style="list-style-type: none"> • Nahrung, Flüssigkeit und Ernährung • Stürze • Prävention von Druckgeschwüren • Beobachtung, Monitoring <p>Darüber hinaus auch Datensammlung zu Mitarbeiterprofilen, um nicht nur Prozess und Outcomes, sondern auch Struktur der Pflegequalität abzubilden.</p>	<i>(The Scottish Government, 2008)</i>
Großbritannien	<p>Benchmarks für folgende Bereiche:</p> <ul style="list-style-type: none"> • Kontrolle über Blase und Darm • Pflegeumwelt • Kommunikation • Essen und Trinken • Schmerzprävention und –management • persönliche Hygiene • Dekubitusprävention und –management • Förderung der Gesundheit und des Wohlbefindens • Führen von Aufzeichnungen 	<i>(Department of Health, 2010)</i>

EXPERTISE ERGEBNISQUALITÄTSINDIKATOREN IN DER STATIONÄREN PFLEGE

Land	Bereiche der Qualitätsmessung	Quelle
	<ul style="list-style-type: none"> • Respekt und Würde • Sicherheit • Selbstpflege 	
Irland	<p>Standards für folgende Bereiche:</p> <ul style="list-style-type: none"> • Rechte • Schutz • Gesundheit und Pflege- und Unterstützungsbedarf • Lebensqualität • MitarbeiterInnen • Pflegeumwelt • Management, Leitung • Zusätzliche Kriterien für demenzspezifische Pflegeeinheiten für ältere Menschen 	<i>(Health Information and Quality Authority, 2009)</i>
Kanada	<p>Fokus auf drei Bereiche (physische Funktion, Lebensqualität, Sicherheit)</p> <p>Physische Funktion</p> <ul style="list-style-type: none"> • Gehen, Rollen • Inkontinenz <p>Lebensqualität</p> <ul style="list-style-type: none"> • Depressionen • Verhaltenssymptome • Schmerz <p>Sicherheit</p> <ul style="list-style-type: none"> • Stürze • Antipsychotische Medikation • Druckgeschwüre 	<i>(Canadian Institute for Health Information, 2013)</i>
Kanada	<p>Fokus auf funktionalen Status, Schmerz, Ermüdung, Dyspnoe, Übelkeit, Stürze und Druckgeschwüre</p> <p>Funktionaler Status</p>	<i>(Canadian Nurses Association, 2008)</i>

EXPERTISE ERGEBNISQUALITÄTSINDIKATOREN IN DER STATIONÄREN PFLEGE

Land	Bereiche der Qualitätsmessung	Quelle
	<ul style="list-style-type: none"> • Selbst-Performance – Bewertung der Performance der BewohnerInnen • Bettmobilität • Transfer • Gehen auf dem Zimmer • Gehen im Korridor • Fortbewegung • Ankleiden • Essen • Toilette • Persönliche Hygiene <p>Schmerz</p> <ul style="list-style-type: none"> • Häufigkeit, in der sich BewohnerInnen über Schmerzen beschweren oder Anzeichen von Schmerzen zeigen • Schmerzintensität <p>Ermüdungserscheinungen</p> <p>Dyspnoe</p> <p>Übelkeit</p> <p>Stürze</p> <p>Druckgeschwüre</p>	
Kanada	<p>Qualitätsindikatoren für unterschiedliche Qualitätsdimensionen (Erreichbarkeit, Effektivität, Sicherheit, Ressourcen, Fokus auf öffentliche Gesundheit)</p> <p>Erreichbarkeit</p> <ul style="list-style-type: none"> • Wartezeiten <p>Effektivität</p> <ul style="list-style-type: none"> • Inkontinenz • Aktivitäten des täglichen Lebens • Kognitive Funktion • Schmerz 	<i>(Health Quality Ontario, 2013)</i>

EXPERTISE ERGEBNISQUALITÄTSINDIKATOREN IN DER STATIONÄREN PFLEGE

Land	Bereiche der Qualitätsmessung	Quelle
	<ul style="list-style-type: none"> • Aufenthalte auf der Intensivstation <p>Sicherheit</p> <ul style="list-style-type: none"> • Stürze • Druckgeschwüre • Fixierung • Medikationssicherheit <p>Angemessene Ressourcen</p> <ul style="list-style-type: none"> • Ressourcen bezogen auf die menschliche Gesundheit <p>Fokussiert auf die öffentliche Gesundheit</p> <ul style="list-style-type: none"> • Infektionen 	
Neuseeland	<p>Für den gesamten Gesundheitssektor gelten Standards in sechs Bereichen (Sicherheit, Patientenerfahrung, Effektivität, Gleichheit, Zugang, Effizienz)</p> <p>Sicherheit</p> <ul style="list-style-type: none"> • Stürze, die zu Schäden führen • Infektionen, die mit der Pflege zusammenhängen • Chirurgische Schäden • Medikation • Druckgeschwüre <p>Patientenerfahrung</p> <ul style="list-style-type: none"> • Patientenerfahrung • Absage von chirurgischen Eingriffen durch das Krankenhaus nach Aufnahme des Patienten bzw. der Patientin <p>Effektivität</p> <ul style="list-style-type: none"> • Ausmaß der funktionellen Gesundheit • Bettenbelegung im Krankenhaus durch PatientInnen, mit einem Alter von 75+, die mindestens zweimal aufgenommen wurden • Übernachtung nach ambulantem Eingriff • Wiederaufnahme ins Krankenhaus • Poststationäre ambulante psychiatrische Versorgung 	<i>(Health Quality & Safety Commission, 2012)</i>

EXPERTISE ERGEBNISQUALITÄTSINDIKATOREN IN DER STATIONÄREN PFLEGE

Land	Bereiche der Qualitätsmessung	Quelle
	<p>Gleichheit</p> <ul style="list-style-type: none"> • Verteilung aller Maßnahmen auf alle Bevölkerungsgruppen <p>Zugang</p> <ul style="list-style-type: none"> • % der berechtigten Bevölkerung mit Zervix-Screening • Impfungen für 2-jährige Kinder <p>Effizienz</p> <ul style="list-style-type: none"> • Gesundheitspflegekosten pro Kopf • % BIP an Gesundheitspflege • Fördermaßnahmen für Arbeitnehmerschaft • Krankenhaustage während der letzten sechs Lebensmonate 	
<p>Niederlande</p>	<p>Pflege (Behandlung/ Leben und Teilhabe)</p> <ul style="list-style-type: none"> • Erfahrungen mit Pflege (Behandlung), Leben und Teilhabe und deren Evaluation • Erfahrene Partizipation und Konsultation <p>Kommunikation und Information</p> <ul style="list-style-type: none"> • Erfahrene Behandlung • Erfahrene Information • Erfahrene Kommunikation und Möglichkeit, Pflegefachkräfte telefonisch zu erreichen <p>Physisches Wohlbefinden</p> <ul style="list-style-type: none"> • Erfahrungen mit physischer Pflege • Erfahrungen mit Mahlzeiten <p>Sicherheit</p> <ul style="list-style-type: none"> • Dekubitus • Unbeabsichtigter Gewichtsverlust • Stürze • medizinische Vorfälle • Psychopharmaka • Impfungen (Klienten, Mitarbeiter) 	<p><i>(ActiZ et al., 2007)</i></p>

Land	Bereiche der Qualitätsmessung	Quelle
	<ul style="list-style-type: none"> • Inkontinenz (Prävalenz, Diagnose) • Dauerkatheter • Problemverhalten • Fixierung • Erfahrene Professionalität und Sicherheit in der Durchführung der Pflege • Erfahrene Respektierung der Rechte im Hinblick auf Bewegungsfreiheit <p>Lebensbedingungen</p> <ul style="list-style-type: none"> • Erfahrener Lebenskomfort • Erfahrene Atmosphäre • Erfahrene Privatsphäre <p>Partizipation und soziale Gewandtheit</p> <ul style="list-style-type: none"> • Erfahrungen, wie Klienten ihren Tag verbringen und Partizipation gestalten • Erfahrene Unabhängigkeit/ Autonomie <p>Mentales Wohlbefinden</p> <ul style="list-style-type: none"> • Erfahrungen, die das mentale Wohlbefinden betreffen • Depression <p>Sicheres Leben/ sicherer Aufenthalt</p> <ul style="list-style-type: none"> • Erfahrene Sicherheit im häuslichen Bereich/ Lebensbereich • Erfahrene Zuverlässigkeit der MitarbeiterInnen • Anleitungen für Transferlifts <p>Kompetente MitarbeiterInnen</p> <ul style="list-style-type: none"> • Erfahrene Erreichbarkeit der MitarbeiterInnen • Erreichbarkeit der Pflegefachkräfte • Erreichbarkeit der Ärzte • Durchführung von risikoreichen Interventionen durch kompetente MitarbeiterInnen 	

EXPERTISE ERGEBNISQUALITÄTSINDIKATOREN IN DER STATIONÄREN PFLEGE

Land	Bereiche der Qualitätsmessung	Quelle
	Kohärenz in der Pflege <ul style="list-style-type: none"> • Kohärenz der Pflege 	
Norwegen	IPLOS Register enthält standardisierte Informationen über EmpfängerInnen von Pflege- und Betreuungsdienstleistungen; Informationen werden in sechs Bereiche unterteilt (persönliche Informationen, einmal solche, die sich ändern und noch einmal solche, die sich nicht ändern; Hilfebedarf; Dienste, die der Empfänger erhält; Diagnosen; spezielle Dienste) – enthält keine Informationen zu Qualitätsindikatoren für Pflege selbst	<i>(Helsedirektoratet, 2014a)</i>
Norwegen	[noch nicht in Pilotphase] <ul style="list-style-type: none"> • Arzt/ Ärztin in Stunden pro Woche je Einwohner • Physiotherapie pro Woche je Einwohner • Prozentualer Anteil Einzelzimmer • Anteil privater Bäder • Personal/ Kompetenz (Anteil der Vollzeitstellen, Berufsausbildung, krankheitsbedingte Fehlzeiten etc.) • Aufenthalt/ durchschnittliche Bettentage • Zufriedenheit der NutzerInnen • Ernährung • Arzneimittel • Gesundheitsförderung Mögliche Erweiterungen: <ul style="list-style-type: none"> • Körperliche Aktivität • Interaktion • Demenz • Psychische Gesundheit • Stürze • Schmerz • Infektionen 	<i>(Helsedirektoratet, 2013; Helsedirektoratet, 2014b)</i>

EXPERTISE ERGEBNISQUALITÄTSINDIKATOREN IN DER STATIONÄREN PFLEGE

Land	Bereiche der Qualitätsmessung	Quelle
	<ul style="list-style-type: none"> • Druckgeschwüre • Zähne 	
USA	<ul style="list-style-type: none"> • Stürze mit Verletzungen • Schmerzen • Dekubitus • Impfungen gegen Grippe • Impfungen gegen Pneumokokken • Infektionen des Urinaltrakts • Kontrollverlust über Blase • Blasenkatheter • Physische Fixierung • Hilfebedarf bei täglichen Aktivitäten • Gewichtsverlust • Depressive Symptomatik • Antipsychotische Medikation 	<p><i>(Centers for Medicare & Medicaid Services, 2013; Centers for Medicare & Medicaid Services, 2014)</i></p>
USA	<p>Folgende Maßzahlen werden empfohlen (Langzeitaufenthalt):</p> <ul style="list-style-type: none"> • Prozentualer Anteil der HeimbewohnerInnen, die eine saisonale Grippeimpfung erhalten haben • Prozentualer Anteil der HeimbewohnerInnen, die eine Pneumokokkenimpfung erhalten haben • Prozentualer Anteil der HeimbewohnerInnen, die eine Harnwegsinfektion bekommen haben • Prozentualer Anteil der HeimbewohnerInnen, die die Kontrolle über Darm oder Blase verloren haben • Prozentualer Anteil der HeimbewohnerInnen, die fixiert worden sind • Prozentualer Anteil der HeimbewohnerInnen, deren Hilfebedarf bei täglichen Aktivitäten gestiegen ist • Prozentualer Anteil der HeimbewohnerInnen, die unbeabsichtigt Ge- 	<p><i>(National Quality Forum, 2011)</i></p>

EXPERTISE ERGEBNISQUALITÄTSINDIKATOREN IN DER STATIONÄREN PFLEGE

Land	Bereiche der Qualitätsmessung	Quelle
	<p>wicht verloren haben</p> <ul style="list-style-type: none"> • Bewertung der Gesundheitsdienstleister – Pflegeheimsurvey • Prozentualer Anteil der HeimbewohnerInnen, die einen oder mehrere Stürze mit einer Verletzung erlebt haben • Prozentualer Anteil der HeimbewohnerInnen, die von mäßigen bis starken Schmerzen berichten • Prozentualer Anteil der HeimbewohnerInnen, die depressive Symptome zeigen • Bewertung der Gesundheitsdienstleister – Pflegeheimsurvey 	

Im Folgenden werden die Zielsetzungen, Entwicklungen, Erfahrungen und Perspektiven der pflegebezogenen Outcomemessungen mithilfe von Indikatoren beispielhaft zusammenfassend dargestellt.

Zielsetzung

Die Entwicklung und Anwendung von Indikatoren zur Überprüfung bzw. Bewertung der Pflegequalität soll auf unterschiedlichen Ebenen einen Nutzen bringen. Einen zentralen, übergreifenden Aspekt stellt hierbei die Qualitätsentwicklung dar, die auf verbesserte Pflegeergebnisse und die Einhaltung von Pflegestandards abzielt (Department of Health, 2012; Health Information and Quality Authority, 2009; National Aged Care Alliance, 2014; The Scottish Government, 2008), wenngleich auch betont wird, dass allein das Messen der Qualität nicht automatisch zu einer besseren Leistung führt (Health Information and Quality Authority, 2013).

Besonders hervorzuheben sind in diesem Kontext solche Indikatorensätze, die auf nationalen Datengrundlagen beruhen oder im überregionalen Kontext Anwendung finden, da diese für internationale bzw. nationale Vergleiche der einzelnen Einrichtungen genutzt werden können (Australian Institute of Health and Welfare, 2009; Health Information and Quality Authority, 2013). Durch diese Vergleiche sollen Erkenntnisse im Hinblick auf gute und schlechte Praxis abgeleitet und ein externes Benchmarking ermöglicht werden (Health Information and Quality Authority, 2013; The Scottish Government, 2008). Ein weiteres Ziel besteht darin, dass hierdurch Pflegeeinrichtungen ihre Ergebnisse in Relation zu denen anderer Pflegeeinrichtungen besser einschätzen und Stärken und Schwächen aufdecken können.

Daneben soll die Bewertung der Pflegequalität mithilfe von Indikatoren dazu beitragen, Transparenz zu schaffen (Health Information and Quality Authority, 2013; National Aged Care Alliance, 2014). Dies betrifft zum einen direkt die NutzerInnen sowie deren Angehörige. Weit verbreitet besteht die Erwartung, dass diese Gruppe durch eine öffentliche qualitätsbezogene Berichterstattung bei der Entscheidungsfindung unterstützt wird, wenn es darum geht, eine geeignete Pflegeeinrichtung zu finden (Department of Health, 2012; Health Information and Quality Authority, 2013; Health Service Executive, 2010; National Aged Care Alliance, 2014; National Quality Forum, 2011). Auf der anderen Seite können Pflegeeinrichtungen durch eine kontinuierliche Messung der Pflegequalität ihre Leistung transparent und nachvollziehbar machen bzw. Rechenschaft über geleistete Pflegemaßnahmen ablegen (Department of Health, 2012).

Einrichtungsintern können zudem die definierten und zu erreichenden Pflegeoutcomes von Pflegefachkräften zur Qualitätssicherung genutzt werden. Professionell Pflegenden erhalten durch diese einen patientenfokussierten und strukturierten Ansatz zur Umsetzung der gesetzlich definierten Richtlinien (Department of Health, 2012) bzw. es wird ein Bewusstsein zur eigenen Leistungsverbesserung geschaffen (Health Information and Quality Authority, 2013).

Im Wesentlichen zielt also der Einsatz von Indikatoren zur Messung der Pflegequalität auf Qualitätsentwicklungs- bzw. -sicherungsprozesse und Schaffung von Transparenz ab, was nicht zuletzt auch den Leistungsnutzern und –nutzerinnen zu Gute kommen soll.

Qualität der Indikatoren, Grundlage, Entwicklung

Qualitätsindikatoren werden häufig auf Grundlage von spezifischen Rahmenkonzepten (Australian Institute of Health and Welfare, 2009; Department of Health, 2012; National Quality Forum, 2011; The Scottish Government, 2008), bestehenden Indikatorensets wie RAI-MDS und den von der OECD entwickelten Indikatorensets (Australian Institute of Health and Welfare, 2009; Canadian Institute for Health Information, 2013; Canadian Nurses Association, 2008; National Quality Forum, 2011; The Scottish Government, 2008) sowie der bestehenden Gesetzgebung (Health Information and Quality Authority, 2009) entwickelt. Hierbei werden bei der Auswahl bzw. Entwicklung bestimmte Anforderungen an die Qualitätsindikatoren gestellt, die jedoch von Projekt zu Projekt unterschiedlich sind (Australian Institute of Health and Welfare, 2009; Health Information and Quality Authority, 2013; National Aged Care Alliance, 2014): Diese umfassen u. a. methodische Bestimmungen an Gütekriterien wie Reliabilität und Validität, denen Indikatoren gerecht werden müssen (Health Information and Quality Authority, 2013; National Aged Care Alliance, 2014).

Betrachtet man die Phasen der Indikatorenentwicklung, so fällt auf, dass sich einzelne Arbeitsschritte weitgehend wiederholen. So gehört u. a. die Definition der Zielgruppe und der (Qualitäts-)Dimensionen, die gemessen werden sollen, das Operationalisieren des Indikators, die Durchführung von Pilotphasen, das Hinzuziehen von ExpertInnen, LeistungserbringerInnen bzw. -nutzerInnen und öffentliche Konsultationen zum Standardprozedere (Australian Institute of Health and Welfare, 2009; Department of Health, 2012; Health Service Executive, 2010; National Institute for Health and Care Excellence, 2014; The Scottish Government, 2008).

Vor der Einführung von Indikatoren sind häufig Pilot- bzw. Evaluierungsphasen vorgesehen, diese sind de facto aber nicht immer Bestandteil des Entwicklungsprozesses. Dies mag damit zusammenhängen, dass oftmals bereits bestehende Indikatorensets genutzt werden, die sich vielfach bewährt haben und nach fest definierten methodischen Gesichtspunkten ausgewählt wurden (Canadian Institute for Health Information, 2013; Canadian Nurses Association, 2008; National Quality Forum, 2011). Wie bereits angedeutet, gibt es v.a. in den englischsprachigen Ländern eine längere Tradition, was die Nutzung von Qualitätsindikatoren für die pflegerische Outcomemessung betrifft. Hier hat sich regionen- und organisationsübergreifend vielerorts Inter-RAI-MDS etabliert, welches aufgrund seines jahrelangen und erprobten Einsatzes in vielen Ländern als Grundlage für weitere Qualitätsinitiativen dient.

Perspektiven

Zwar werden während der Indikatorenentwicklung unterschiedliche Perspektiven berücksichtigt, u. a. die der leistungserbringenden Pflegefachkräfte und der NutzerInnen, und der Fokus auf diesen Personenkreis und seinen Bedarf gelegt, jedoch scheinen weder bei den Qualitätsdimensionen, entlang derer die Outcomes gemessen werden, noch die Indikatoren selbst, sich dezidiert auf die Nutzerperspektive zu stützen – zumindest ist dieser Rückschluss aus den Dokumenten zu ziehen. Dies wird verschiedentlich als Defizit gewertet und ist ein Grund dafür, dass einzelne Organisationen wie die National Aged Care Alliance (2014) fordern, dass die Erfahrungen der HeimbewohnerInnen einen Platz in der Indikatorenentwicklung finden müssen. Eine Ausnahme stellen hier die niederländischen Qualitätsindikatoren dar

(ActiZ et al., 2007), die eine Involvierung und Partizipation der HeimbewohnerInnen, die über eine einfache Zufriedenheitsabfrage hinausreicht, in mehreren Qualitätsbereichen vorsehen.

4 Beantwortung der Fragen

4.1 Reichweite der Indikatoren

(Renate Stemmer & Jens Arnold)

4.1.1 Bildet das Indikatorenset von Wingenfeld die für eine vergleichende Qualitätsberichterstattung relevanten Pflegebereiche in der erforderlichen Themenbreite und -tiefe hinreichend ab? Auf welche von Wingenfeld et al. vorgeschlagenen Themenbereiche/Indikatoren kann verzichtet werden, welche sollten ergänzend berücksichtigt werden?

Gegenstand dieser Expertise sind die Themenbereiche 1-3 (Erhalt und Förderung von Selbstständigkeit, Schutz vor gesundheitlichen Schädigungen, Belastungen und Unterstützung bei spezifischen Bedarfslagen) aus dem von Wingenfeld et al. (2011) durchgeführten Projekt zur Messung von Ergebnisqualität in der stationären Altenhilfe. Schon in der Bearbeitung des Auftrages zur Entwicklung und Erprobung von Instrumenten zur Beurteilung der Ergebnisqualität in der stationären Altenhilfe des Bundesministeriums für Gesundheit sowie des Bundesministeriums für Familie, Senioren, Frauen und Jugend durch das Institut für Pflegewissenschaft an der Universität Bielefeld (IPW) und das Institut für Sozialforschung und Gesellschaftspolitik GmbH (ISG) wurde die Perspektive in Übereinstimmung mit der nationalen und internationalen Diskussion auf den Bereich des Erlebens der BewohnerInnen, der hier als ‚Lebensqualität‘ gefasst wurde, ausgeweitet.

Insofern besteht zwischen Wingenfeld et al. (2011) und den VerfasserInnen der vorliegenden Expertise Konsens, dass eine ausschließliche Erhebung von Indikatoren im gesundheitsbezogenen Bereich nicht ausreicht.

In diesem Kapitel wird die inhaltliche Diskussion geführt. Ob und inwieweit es zulässig ist, einen speziellen Indikator für einen Vergleich heranzuziehen, ist jedoch nicht zuletzt eine Frage der methodischen Güte. Diese beiden Perspektiven können nicht völlig unabhängig voneinander diskutiert werden. Wenn ein potentieller Indikator zwar inhaltlich wünschenswert ist, kommt er gleichwohl für die Outcomemessung nicht in Betracht, wenn er nicht den erforderlichen Kriterien entspricht.

Wingenfeld et al. (2011, 26) weisen folgende Auswahlkriterien für die Identifikation geeigneter Indikatoren aus:

- „Pflegesensitivität bzw. Einfluss der Einrichtung und ihrer Mitarbeiter/innen auf das Ergebnis, das mit dem jeweiligen Indikator erfasst werden soll (unter Einbeziehung der Ergebnisse wissenschaftlicher Forschung),
- Relevanz,
- Messbarkeit (Verfügbarkeit und Verlässlichkeit der zur Ergebniserfassung erforderlichen Informationen, Verfügbarkeit von Messverfahren/ -methoden),

- Abhängigkeit des Ergebnisses von der Bewohnerstruktur und Möglichkeiten, eine eventuelle Abhängigkeit über eine Risikoadjustierung bzw. eine Unterscheidung von Risikogruppen zu kompensieren (ebenfalls auf der Basis von Forschungsergebnissen),
- Anschlussfähigkeit an die Praxis (z. B. Verhältnis zwischen fachlichen Anforderungen bei der Messung und fachlichen Voraussetzungen auf Seiten der Einrichtungen),
- Zeitaufwand bzw. Verhältnis zwischen Aufwand und Nutzen.“

Diese Anforderungen an Indikatoren für die Outcomemessung decken sich weitgehend mit der nationalen und internationalen Diskussion zu diesem Thema, auch wenn z. B. das von ExpertInnen der Bundesgeschäftsstelle Qualitätssicherung (BQS) entwickelte QUALIFY-Instrument die Kategorien Relevanz, Wissenschaftlichkeit und Praktikabilität noch genauer spezifiziert. Allerdings finden sich auch Abweichungen. So fehlt der Hinweis auf die Berücksichtigung potentieller negativer Auswirkungen (Expertenkreis Qualitätsindikatoren für Nationale VersorgungsLeitlinien, 2009; National Aged Care Alliance, 2014).

Ein Hinweis noch zu dem Kriterium der ‚Anschlussfähigkeit an die Praxis‘, das sich im QUALIFY-Instrument als ‚Implementationsbarrieren‘ wiederfindet. Natürlich muss die Indikatorerhebung anschlussfähig sein. Gleichwohl liegt hier der Teufel im Detail, denn Anschlussfähigkeit kann und darf nicht bedeuten, dass die aktuelle Praxis zum einzigen Maßstab wird, sondern muss das Entwicklungspotential der Praxis mit in Rechnung stellen. Das von den Leistungsträgern und Leistungserbringern anvisierte Vorhaben der Verzahnung von interner Qualitätsentwicklung und externer Qualitätsprüfung verbunden mit der Qualitätsberichterstattung hat gegenüber der aktuellen Praxis eine neue Dimension, die Entwicklungsschritte in den Einrichtungen unerlässlich machen (s. Kapitel 4.2.2).

Themenauswahl

Welche Themenbereiche/ Indikatoren sollen und können nun in welcher Tiefe und Breite zur Qualitätsberichterstattung genutzt werden?

Grundsätzlich ist zu konstatieren, dass es nicht darum gehen kann, möglichst ‚alle‘ Themen mittels Indikatoren detailliert abzubilden, sondern dass Themen mit einer besonderen Relevanz zu identifiziert und in Indikatoren und Messinstrumente zu übersetzen sind. Auch wenn die britischen und niederländischen AutorInnen, die beide explizit diesen Ansatz verfolgen, zu einer unterschiedlichen Auswahl der Themen und Indikatoren kommen (ActiZ et al., 2007; Department of Health, 2010), ist dieses grundsätzliche Vorgehen nicht nur wegen des potentiellen Aufwandes unbedingt sinnvoll. Auch erfordert gerade die angestrebte Zusammenführung von interner Qualitätsentwicklung und externer Qualitätsprüfung ein Vorgehen, das den Einrichtungen ausreichend Spielraum lässt, um z. B. eigene Akzente zu verfolgen.

Die Identifikation der ‚relevanten Pflegebereiche‘ muss einerseits theoretisch und andererseits empirisch beantwortet werden. Outcomemessung hat primär auf Fragen der Sicherheit der BewohnerInnen fokussiert (Pringle & Doran, 2003). Doch immer mehr rückt die Bedeutung der Perspektive der BewohnerInnen selbst in den Blick. Diese wird offensiv sowohl in der deutschen als auch der europäischen Charta der Rechte hilfe- und pflegebedürftiger Menschen eingefordert. Beide Deklarationen fokussieren die Bereiche Selbstbestimmung und Hilfe zur Selbsthilfe; körperliche und seelische Unversehrtheit; Freiheit und Sicherheit;

Privatheit; Pflege, Betreuung und Behandlung; Information, Beratung und Aufklärung; Kommunikation, Wertschätzung und Teilhabe an der Gesellschaft; Religion, Kultur und Weltanschauung sowie palliative Begleitung, Sterben und Tod (Bundesministerium für Familie & Bundesministerium für Gesundheit, 2010; AGE Platform Europe, 2010).

Die europäische Charta der Rechte und Pflichten älterer hilfe- und pflegebedürftiger Menschen ergänzt die Aufzählung um das Recht auf Wiedergutmachung im Falle von Misshandlung, Gewalt und Vernachlässigung. Anders als in der deutschen Charta benennt die europäische Charta auch Pflichten der hilfebedürftigen Person wie die Pflicht des respektvollen Umgangs mit den sie umgebenden Menschen sowie der Übernahme von Verantwortung für die Zeit der Hilfe- und Pflegebedürftigkeit, z. B. durch die Regelung einer Betreuung (AGE Platform Europe, 2010).

Mit Blick auf die Qualitätsmessung ist es bemerkenswert, dass Themen, die im engeren Sinne mit Gesundheit und Sicherheit zu tun haben, nur in wenigen Artikeln Erwähnung finden. Dies betont noch einmal die Bedeutung der ‚weichen Faktoren‘, die in dem Entwurf von Wingenfeld und Engels den Bereichen 4 und 5 zugeordnet waren und nicht Gegenstand dieses Gutachtens sind.

Einleitend zur Beantwortung der Frage nach den relevanten Pflegebereichen für die Qualitätsberichterstattung ist die Zielgruppe der Qualitätsberichterstattung zu klären. Wem gegenüber soll die Qualität einer Einrichtung berichtet werden? Grundsätzlich kommen zwei Zielgruppen in Betracht. Es sind dies einerseits die VerbraucherInnen bzw. NutzerInnen und andererseits entsprechende Kontrollorgane. Die Gruppe der VerbraucherInnen ist ihrerseits nicht homogen, sondern lässt sich in Betroffene (potentielle BewohnerInnen und Angehörige) und Dienstleister (Krankenhaussozialdienst, Pflegestützpunkte, HausärztInnen etc.) differenzieren (s. Kapitel 4.5). Die Interessen dieser Zielgruppen sind nicht deckungsgleich. Zahlreiche Befunde weisen darauf hin, dass die Betroffenen eher an Informationen interessiert sind, die psychosoziale Aspekte abbilden, während die professionellen Dienstleister sowie die Kontrollorgane stärker gesundheitsbezogenen Daten Bedeutung zumessen.

Welche Ziele sollen mit der Qualitätsberichterstattung verfolgt werden? Neben dem Vergleich von Einrichtungen hinsichtlich ihrer Qualität ist die Information der NutzerInnen zur Unterstützung einer qualitätsinformierten Entscheidungsfindung ein vielfach verfolgtes Ziel. Auch die Zielsetzung, Pflege- und Versorgungsqualität durch die Schaffung kompetitiver Anreize zu verbessern, ist hier einzuordnen. Qualitätsziele und Qualitätskriterien sind wechselseitig aufeinander bezogen, d.h. mit spezifischen Kriterien können auch nur bestimmte Ziele erreicht werden (Weibler-Villalobos & Röhrig, 2010). Die Festlegung der Ziele ist nicht nur bestimmend für die inhaltliche Ausrichtung der Qualitätskriterien und –indikatoren, sondern auch für deren Gestaltung. Wenn eines der Ziele darin besteht, die Bewohnerperspektive explizit zu stärken, muss sich dies in der Ausgestaltung der Indikatoren niederschlagen (ActiZ et al., 2007; National Aged Care Alliance, 2014).

Die inhaltliche Relevanz von Themenbereichen/ Indikatoren steht aus fachlicher Sicht im Zusammenhang mit der Häufigkeit des Auftretens eines Ereignisses, dem Gefährdungspotential und/ oder einem professionellen Verständnis guter Pflege oder Versorgung. Die thematische Relevanz ist aus Verbrauchersicht aber vielfach ganz anders gelagert (Geraedts, Brechtel, Zöll, & Hermeling, 2012). Zudem ist der Nutzen der Qualitätsmessung mit Hilfe ausgewählter Indikatoren gegenüber einem potentiellen Schaden (z. B. bedingt durch hohen Aufwand oder

Fehlanreize mit Beeinflussung der Ressourcenverteilung) abzuwägen (Weibler-Villalobos & Röhrig, 2010).

Je nachdem, welche Kriterien, welche Indikatoren und welche Bewertungssystematik herangezogen werden, kann die Positionierung ein und derselben Einrichtung innerhalb eines Rankings sehr unterschiedlich ausfallen. Es ist also zwingend erforderlich, das Verfahren zur Auswahl der Kriterien und Indikatoren transparent zu machen (Weibler-Villalobos & Röhrig, 2010). Wie bereits ausgeführt sind diese Vorklärungen nur ansatzweise erfolgt (s. Kapitel 1.3).

Soweit möglich sollten Outcomeindikatoren aus bestehenden Leitlinien oder Standards abgeleitet werden. In Deutschland arbeitet das Deutsche Netzwerk für Qualitätsentwicklung in der Pflege (DNQP) in Kooperation mit dem Deutschen Pflegerat (DPR) seit 1999 an der Entwicklung evidenzbasierter Expertenstandards in der Pflege. Mit Relevanz für den Bereich der stationären Pflege wurden bis heute vom DNQP sieben Expertenstandards entwickelt und publiziert:

- Dekubitusprophylaxe in der Pflege
- Schmerzmanagement in der Pflege bei akuten Schmerzen
- Schmerzmanagement in der Pflege bei chronischen Schmerzen
- Sturzprophylaxe in der Pflege
- Förderung der Harnkontinenz in der Pflege
- Pflege von Menschen mit chronischen Wunden
- Ernährungsmanagement in der Pflege

Ein achter Expertenstandard, nun zum Thema „Erhaltung und Förderung der Mobilität“, wurde im März 2014 auf einer Fachkonferenz vorgestellt und konsentiert (DNQP, 2014a). Für die Auswahl von Themen für die Expertenstandards in der Pflege sind pflegeepidemiologische Daten leitend (DNQP, 2011b).

In inhaltlicher Hinsicht sind alle Expertenstandards dem Fokus Gesundheit und Sicherheit zuzuordnen. Die Themen Dekubitus, Schmerz, Sturz und Ernährung finden sich in den Empfehlungen von Wingenfeld et al. (2011) wieder.

Diskussion der von Wingenfeld et al. (2011) vorgeschlagenen Indikatoren

Bereich 1: Erhalt und Förderung der Selbständigkeit:

Wingenfeld et al. (2011) interpretieren Erhalt und Förderung von Selbständigkeit als

- Erhalt oder Verbesserung der Mobilität (Positionswechsel im Bett, stabile Sitzposition halten, Aufstehen aus sitzender Position/ Umsetzen, Fortbewegen innerhalb des Wohnbereiches, Treppensteigen),
- Selbständigkeitserhalt oder –verbesserung bei Alltagsverrichtungen (Körperpflege, An- und Auskleiden, Ernährung, Ausscheiden),

- Selbständigkeitserhalt oder –verbesserung bei der Gestaltung der Alltagskontakte (Tagesablauf gestalten und an Veränderungen anpassen, Ruhen und Schlafen, sich beschäftigen, in die Zukunft gerichtete Planungen vornehmen, Interaktion mit Personen im direkten Kontakt, Kontaktpflege zu Personen außerhalb des direkten Umfeldes) (Wingenfeld et al., 2011).

Die AutorInnen differenzieren bei dem Thema Mobilität und dem Selbständigkeitserhalt bei Alltagsverrichtungen jeweils nach den Gruppen:

- Bewohner, die keine oder nur geringe kognitive Einbußen aufweisen und
- Bewohner, die mindestens erhebliche kognitive Einbußen aufweisen.

Beim dritten Thema im Bereich 1, dem Selbständigkeitserhalt bei der Gestaltung der Alltagskontakte, wird diese Differenzierung nicht vorgenommen, stattdessen werden BewohnerInnen mit völligem / weitgehenden Fähigkeitsverlust im Bereich der kognitiven und kommunikativen Fähigkeiten aus der Berechnung ausgeschlossen (Wingenfeld et al., 2011).

Die Förderung von Selbständigkeit im körperlichen und sozialen Bereich gehört zu den Kernzielen pflegerischen Handelns. Eingeschränkte Mobilität bzw. Immobilität führt zu vielfältigen negativen Auswirkungen wie Dekubiti, Kontrakturen, Pneumonien, Schlafstörungen und Appetitlosigkeit (Bates-Jensen et al., 2004; Kottner, Dassen, & Lahmann, 2011). Eine deutsche Prävalenzerhebung zur Pflegeabhängigkeit beim Essen und Trinken (n= 3552) weist einen Anteil von 42,9% der PflegeheimbewohnerInnen auf, die überwiegend oder vollständig beim Essen und Trinken pflegeabhängig waren (Lahmann et al., 2012). Nach Schneekloth und v. Törne können sich 57% der BewohnerInnen von Alteneinrichtungen unmöglich allein und 31% nur mit Schwierigkeiten duschen oder waschen, 40% können nicht im Zimmer umhergehen und 16% können dies nur mit Schwierigkeiten. Immerhin 11% der BewohnerInnen sind im engeren Sinne bettlägerig und 8% verbringen täglich nicht mehr als 3-5 Stunden außerhalb des Bettes (Schneekloth & von Törne, 2007).

So ist die empirische Relevanz dieser Indikatoren nachdrücklich gegeben und findet sich die Erhebung des Grades der Selbständigkeit in vielen Indikatorensets. Die Items für die Messung von Selbständigkeit wechseln. Während Wingenfeld et al. (2011) Mobilität und Alltagsverrichtungen trennen, fasst das RAI-MDS, das als Datengrundlage für die Datenerhebung zur Qualitätsmessung in mehreren Ländern benutzt wird, Selbständigkeit als einen Bereich, der mit den Items Bettmobilität, Transfer, Essen und Toilettenbenutzung operationalisiert wird (RTI International, 2012). Der Grad der Selbständigkeit in den Aktivitäten des täglichen Lebens (ADL) bei BewohnerInnen in der Langzeitpflege hat sich als pflegesensitiv erwiesen (Doran et al., 2003). Die von Wingenfeld et al. (2011) vorgenommene Unterteilung in Mobilität und Selbständigkeit bei Alltagsverrichtungen erhöht den Grad der Differenziertheit der Einschätzung und kann positiv bewertet werden. Kritische Anfragen betreffen die Operationalisierung von Mobilität (Bensch, 2013), nicht jedoch das grundsätzliche Ansinnen, den Grad der Mobilität von BewohnerInnen in Pflegeheimen als eigenständiges Item in die Outcomemessung aufzunehmen.

Eine spezifische Unterteilung der BewohnerInnen auf der Ebene der Indikatoren nach dem Grad der kognitiven Leistungsfähigkeit findet sich allerdings in keinem der von uns recherchierten international vorgefundenen Indikatorensets. Stattdessen werden die damit verbun-

denen potentiellen Einflüsse in der Regel mittels multifaktorieller Risikoadjustierung, auf der Basis von Regressionsanalysen oder neuerdings auch hierarchischen Modellen, abgebildet (s. Kapitel 4.4.1).

Das dritte Feld der von Wingefeld et al. (2011) differenzierten Form der Selbständigkeit betrifft die Gestaltung der Alltagskontakte. Das Alten- oder Pflegeheim stellt für seine BewohnerInnen das Lebensumfeld dar. Die Einflussnahme auf den Tagesablauf und die Pflege sozialer Kontakte gehören in fundamentaler Weise zu einem selbstbestimmten Leben. Die Möglichkeiten zur Gestaltung des Tagesablaufes sind bei HeimbewohnerInnen oftmals jedoch eingeschränkt. Beispielsweise ist es der MUG IV-Erhebung zufolge nur 20% der BewohnerInnen mit Schwierigkeiten möglich zu telefonieren, 36% ist dies allein völlig unmöglich; für 55% der BewohnerInnen ist es unmöglich allein Besuche zu machen und 25% gelingt dies nur mit Schwierigkeiten (Schneekloth & von Törne, 2007). So ist auch dieses Themenfeld ohne Zweifel von hoher Relevanz.

Es bleibt aber die Frage, ob bzw. inwieweit sich für die Erhebung in diesem Feld die Fremdbeobachtung eignet. Die Gestaltung des Alltagslebens und der sozialen Kontakte wird u. a. operationalisiert durch den Grad der Selbständigkeit bei der Tagesgestaltung. Der Spielraum zur Gestaltung des Tagesablaufes könnte aber nicht nur durch das Vermögen der BewohnerInnen bedingt sein, sondern auch durch organisatorische Bedingungen der Einrichtung. Eine weitere Operationalisierung lautet ‚in die Zukunft gerichtet Planungen vornehmen‘. Nicht nur, dass unklar ist, wie hier ‚Zukunft‘ zu interpretieren ist und wie konkret die Planungen sein müssen. Es stellt sich auch die Frage, ob die Pflegekräfte³ in der Regel derartig dezidierte Kenntnisse über die BewohnerInnen haben. Wenn dem nicht der Fall ist, werden hier die Vorstellungen der Pflegekräfte über die Möglichkeiten der Alltagsgestaltung erhoben, die nicht zwingend deckungsgleich mit dem Erleben der BewohnerInnen sein müssen.

Wenn dieser Indikator erhalten bleibt, sollte er ergänzt werden um einen Abgleich mit einer Erhebung des Erlebens und der Erfahrungen der BewohnerInnen wie ihn das niederländische Steering Committee Responsible Care vorschlägt (ActiZ et al., 2007).

Bereich 2: Schutz vor gesundheitlichen Schädigungen

Der zweite von Wingefeld et al. (2011) vorgeschlagene Bereich zur Ergebnismessung in der stationären Pflege widmet sich dem Schutz vor gesundheitlichen Schädigungen. Identifiziert werden drei Themen, die jeweils mit zwei Indikatoren (in Abhängigkeit von Risikograd) formuliert sind:

Dekubitusentstehung

- bei Bewohnern mit *geringem* Dekubitusrisiko,
- bei Bewohnern mit *hohem* Dekubitusrisiko.

Stürze mit gravierenden Folgen

- bei Bewohnern, die keine oder nur geringe kognitive Einbußen aufweisen,

³ Gemeint sind AltenpflegerInnen, Gesundheits- und KrankenpflegerInnen, Gesundheits- und KinderkrankenpflegerInnen, aber auch Pflegepersonen mit einer einjährigen Ausbildung, die im Austausch bewohnerbezogene Kenntnisse zu diesem Indikator zusammentragen könnten.

- bei Bewohnern, die mindestens erhebliche kognitive Einbußen aufweisen.

Unbeabsichtigter Gewichtsverlust

- bei Bewohnern, die keine oder nur geringe kognitive Einbußen aufweisen,
- bei Bewohnern, die mindestens erhebliche kognitive Einbußen aufweisen (Wingenfeld et al., 2011).

Das Vorliegen eines Dekubitus ist ein Geschehen, das mit Schmerzen und weiteren Belastungen verbunden ist. Zwar zeigt sowohl die Prävalenz- als auch die Inzidenzrate je nach Erhebungsmethode, Setting und Land eine erhebliche Bandbreite. Unbestritten ist jedoch die grundsätzliche Relevanz dieses Themas. Seine Entstehung kann vielfach durch fachgerechte Pflege verhindert werden (DNQP, 2010a; Kottner et al., 2011). Der 3. Qualitätsbericht des MDS weist eine Dekubitusprävalenz von 4,4% auf (ohne dass hier allerdings eine Stadieneinteilung ausgewiesen ist). Der Dekubitus ist als Thema für die Outcomemessung in der stationären Pflege weit verbreitet. Es bestehen jedoch Unterschiede hinsichtlich der genauen Ausgestaltung des Indikators.

Im Hinblick auf die Zielsetzung, die Pflegequalität einer Einrichtung abzubilden, besteht weitgehende Übereinstimmung darin, die Dekubitusinzidenz (nicht die Prävalenz) zu erheben. Da die Erhebung eines Dekubitus Grad 1 (nicht wegdrückbare Hautrötung) nicht immer sicher erkannt wird, konzentriert sich die Outcomemessung in der Regel – wie auch bei Wingenfeld et al. (2011) – auf die Grade 2-4.

Wingenfeld et al. (2011) schlagen vor, die Dekubitusinzidenz sowohl bei BewohnerInnen mit hohem als auch mit niedrigem Dekubitusrisiko zu erheben. Hier stellt sich die Frage, ob eine Erhebung bei der Gruppe mit hohem Risiko nicht ausreichen würde. Das RAI-MDS (RTI International, 2012) empfiehlt ebenso wie das niederländische Steering Committee Responsible Care (ActiZ et al., 2007) nur die Erhebung der Daten in der Hochrisikogruppe. Wingenfeld et al. (2011) führen für die Erhebung in den beiden Risikogruppen keine spezifische Begründung auf. Vor dem Hintergrund jetzt schon überbordender Bürokratie und dem Anliegen des gezielten Ressourceneinsatzes scheint die Erhebung der Dekubitusinzidenz bei BewohnerInnen mit geringem Risiko verzichtbar.

Eine weitere Unterscheidung zeigt sich in der Differenziertheit des Indikators. Während Wingenfeld et al. (2011) ausschließlich das Vorhandensein eines Dekubitus Grad 2-4 erfragen, der in den letzten sechs Monaten in der Einrichtung entstanden ist, erhebt Health Quality Ontario (2013) Verschlechterungen eines vorliegenden Dekubitus. Die Überprüfung der Verschlechterung eines Druckgeschwürs scheint auf den ersten Blick sinnvoll, da sich mit dem Statusverlauf auch zwischenzeitlich erfolgte pflegfachliche Bemühungen abbilden sollten. Jedoch sollte sich dies auch in der Inzidenzrate widerspiegeln. Da der Erhebungsaufwand voraussichtlich höher, der zusätzliche Informationsgewinn aber begrenzt wäre, scheint es sinnvoll, es bei der Erhebung der Inzidenzrate zu belassen.

Der Sturz von BewohnerInnen ist ebenfalls ein zentrales Thema in der Langzeitpflege (DNQP, 2013). Eine Erhebung der Charité 2012 in deutschen Pflegeheimen kommt zu dem

Ergebnis, dass in den letzten 14 Tagen 7,8% der sturzgefährdeten BewohnerInnen tatsächlich gestürzt sind mit einer Spannweite der beteiligten Pflegeheime zwischen 0% - 52% (Lahmann et al., 2012). Die Häufigkeitsangaben liegen in einer schwedischen Studie bei ca. 40% oder mehr PflegeheimbewohnerInnen, die innerhalb eines halben Jahres stürzen. Diese Rate erhöht sich bei Menschen mit einer Demenzerkrankung auf ca. 60%. Ca. ein Drittel aller Stürze bei alten und / oder kranken Menschen führt zu einer Verletzung (Balzer, Junghans, Behncke, & Lühmann, 2013).

Sturz ist ein häufig genutztes Thema für die Outcomemessung. Der Sturz wird teilweise als Sturzereignis oder als Sturzereignis mit gravierenden Folgen als Outcomeindikator interpretiert. Wingenfeld et al. (2011, 52) haben sich für die Beschreibung des Indikators als Sturz mit gravierenden Folgen entschieden, der in den letzten 6 Monaten aufgetreten ist. Zu den gravierenden Folgen zählen „eine Fraktur und / oder eine behandlungsbedürftige Wunde und/oder andauernde Schmerzen und/oder ein erhöhter Hilfebedarf bei Alltagsverrichtungen“.

Health Quality Ontario erhebt die Sturzereignisse in den letzten 30 Tagen unabhängig von unmittelbar feststellbaren Folgen. Auch das niederländische Steering Committee hat sich für die Erhebung der Sturzereignisse entschieden, explizit unabhängig von evtl. Verletzungen (Health Quality Ontario, 2013; ActiZ et al., 2007). Die Center for Medicare and Medicaid nutzen als Vergleichsgröße demgegenüber den Indikator „Stürze mit gravierenden Folgen“ (Center for Medicare and Medicaid Services, 2012). Allerdings ist die Diskussion, ob es angesichts einer Vielzahl von Klassifikationen von Sturzereignissen nicht doch sinnvoll wäre, alle Stürze als Datengrundlage zu erheben, nicht abgeschlossen (National Quality Forum, 2011).

Nicht nur im Zusammenhang mit der Überarbeitung des Expertenstandards zur Sturzprophylaxe in der Pflege wurden die Herausforderungen diskutiert, die mit der Identifikation von Stürzen verbunden sind. Wann ist ein Ereignis als Sturz zu bewerten? Gilt auch ein abgefangener Sturz als Sturz, erfüllt auch das Herunterrollen von einer Matratze, wenn diese auf dem Boden liegt, schon den Tatbestand? Auch ist die Abgrenzung von Beinahe-Stürzen und tatsächlichen Stürzen nicht immer einfach (Balzer et al., 2013; Mueller & Karon, 2004).

Auf der Grundlage einer umfangreichen Literaturanalyse identifizieren Balzer et al. (2013) weit verbreitete gravierende Mängel bei der Dokumentation von Sturzereignissen, die auch durch zusätzliche Definitionen und Erläuterungen nicht ausgemerzt werden konnten. Vor diesem Hintergrund scheint es sinnvoll, für die Qualitätsberichterstattung dem Vorschlag Wingenfelds et al. (2011) zu folgen und nur Stürze mit gravierenden Folgen als Indikator für den Qualitätsvergleich heranzuziehen.

Ein unbeabsichtigter Gewichtsverlust kann auf Mangelernährung hinweisen. Der dritte Qualitätsbericht des MDS weist einen Anteil von 9,1% der PflegeheimbewohnerInnen mit einer relevanten Abnahme des Körpergewichts (mehr als 5% in 1–3 Monaten, mehr als 10% in 6 Monaten) auf (Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen (MDS), 2012). Die Ursachen für Mangelernährung sind komplex und stehen im Zusammenhang mit abnehmendem Appetit im Alter, einer Reduktion kognitiver Einsicht in eine ausreichende Nahrungsaufnahme und/ oder abnehmenden funktionalen Fähigkeiten sowie dem Angewiesensein auf Unterstützung. Ein vorliegendes Defizit kann nur schwer wieder ausge-

glichen werden (Bartholomeyczik et al., 2010; DNQP, 2013) . Als Ursachen für Ernährungsprobleme von PflegeheimbewohnerInnen weist die Studie der Charité u. a. funktionelle Probleme der Arme/ Hände (18,2%), Multimedikation (13,5%), Appetitlosigkeit (12,0%), Kauprobleme (8,8%), Schluckprobleme (8,8%) und Schmerzen (4,4%) aus (Lahmann et al., 2012).

Wingenfeld et al. (2011, 54) definieren den Indikator unbeabsichtigter Gewichtsverlust als „Anteil der Bewohner mit einer nicht intendierten Gewichtsabnahme von mehr als 10% ihres Körpergewichtes in den vergangenen sechs Monaten“. Das Thema Ernährung findet sich in zahlreichen Verfahren zur Qualitätsmessung in der Pflege. Mehrfach wird – wie von Wingenfeld et al. (2011) auch – der unbeabsichtigte Gewichtsverlust als Indikatordefinition herangezogen. Dies gilt für das US-amerikanische Nursing Home Compare und auch für das niederländische Responsible Care (Center for Medicare and Medicaid Services, 2012; ActiZ et al., 2007). Allerdings legt das niederländische Steering Committee absolute Werte fest (unbeabsichtigter Gewichtsverlust von mehr als 3 Kilo im letzten Monat oder mehr als 6 Kilo in den letzten 6 Monaten) (ActiZ et al., 2007). Zusätzlich wird auch der BMI als Outcome-Indikator genutzt (Australian Institute of Health and Welfare, 2009). Allerdings kann es schwierig werden, die Größe des Bewohners / der Bewohnerin z. B. im Falle von Rollstuhlpflichtigkeit, Bettimmobilität oder nach Amputationen an den unteren Extremitäten exakt zu bestimmen (Nygaard, 2008).

Gewichtskontrollen sind ein relativ einfaches Verfahren, um im Sinne eines Screenings einen ersten Hinweis auf ein relevantes Geschehen zu erfassen. In Übereinstimmung mit dem Expertenstandard zum Ernährungsmanagement (DNQP, 2010b) sollte allerdings nicht nur der unbeabsichtigte Gewichtsverlust nach 6 Monaten, sondern auch ein unbeabsichtigter Gewichtsverlust von 5% des Körpergewichtes innerhalb von 4 Wochen in die Definition aufgenommen werden.

Mit den Indikatoren zu Dekubitus, Stürze und unbeabsichtigtem Gewichtsverlust ist bei Wingenfeld et al. (2011) das Themenfeld der Sicherheit für BewohnerInnen abgeschlossen.

Im Abgleich mit den Themen der Expertenstandards bleiben noch die Themen Schmerz, chronische Wunden und Harnkontinenz.

Das Thema Schmerzmanagement weisen die AutorInnen dem Bereich der Unterstützung bei spezifischen Bedarfslagen zu. Die Diskussion dieses Bereiches erfolgt weiter unten.

Das Themenfeld ‚Pflege von Menschen mit chronischen Wunden‘ nimmt eine Sonderstellung ein. International sind viele Länder intensiv mit Fragen der Qualitätsmessung beschäftigt. Im Bereich der Outcomemessung für die Langzeitpflege finden sich im Kontext von Gesundheit und Sicherheit gehäuft folgende Themenbereiche in Australien (Australian Institute of Health and Welfare, 2009), Großbritannien, (Department of Health, 2012; Department of Health, 2010), Irland (Health Information and Quality Authority, 2009), Kanada (Canadian Institute for Health Information, 2013), Niederlande (ActiZ et al., 2007) Schottland (The Scottish Government, 2008) und den USA (Centers for Medicare & Medicaid Services, 2014):

- Dekubiti
- Stürze (mit und ohne Folgen)
- Medikation
- Schmerzen
- Lebensqualität/ Wohlbefinden

- Inkontinenz
- Funktionale Fähigkeiten
- Ernährung
- Fixierung

Ein Vergleich dieser international verbreitet für die Outcomemessung genutzten Themenfelder mit den vorliegenden Expertenstandards zeigt, dass sich ‚Wundversorgung‘ in keinem der recherchierten Indikatorensets als eigenständiges Thema wiederfindet. Eines der Probleme, das sich hier stellt, ist die Heterogenität. Eine Erhebung in niederländischen Pflegeheimen weist eine Prävalenz von 4,2% aus (n= 1514). Davon waren 46% Dekubiti. Die verbleibenden chronischen Wunden verteilten sich auf venös bedingte Ulcera Cruris, arteriell bedingte Ulcera Cruris, diabetische Gangrän, postoperative Wunden, Verletzungen, onkologische Wunden und weitere nicht identifizierte Ursachen (Rondas, Schols, Stobberingh, & Halfens, 2013). Der 3. Qualitätsbericht des MDS nach § 114a weist einen Anteil von 7,3% der BewohnerInnen mit chronischen Wunden aus bei einer Stichprobengröße von n= 61.985. Der Dekubitus macht einen Anteil von 59% an allen erfassten chronischen Wunden aus. Die verbleibenden 41% verteilen sich auf das Ulcus Cruris, das diabetische Fußsyndrom und sonstige chronische Wunden (Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen (MDS), 2012). Wingenfeld et al. (2011) bewerten die Versorgung von chronischen Wunden (wie Ulcus Cruris und diabetisches Gangrän) zudem als nicht pflegesensitiv, da die Wundversorgung ärztlicherseits angeordnet werden muss. Des Weiteren ist anzuerkennen, dass die Wundsituation häufig schwierig bleibt. Allerdings lassen sich oftmals Verbesserungen der Lebensqualität der Betroffenen erreichen (DNQP, 2009). Damit scheint es aus Sicht der Outcomemessung sinnvoll, dieses Thema ggf. aus der Perspektive der Lebensqualität aufzugreifen.

Im Abgleich zu den Themen der Expertenstandards bleibt noch die ‚Harnkontinenz‘. Das Deutsche Netzwerk für Qualitätsentwicklung in der Pflege (DNQP) identifizierte auch die Harnkontinenz als bedeutendes Thema für die Langzeitpflege und entwickelte einen Expertenstandard zur Kontinenzförderung (DNQP, 2014b). Eine deutsche Prävalenzerhebung, an der sich 47 Pflegeheime beteiligten und 3552 BewohnerInnen einbezogen wurden, weist einen Anteil von 67,4% mit Harninkontinenz aus (Lahmann et al., 2012). Das aktualisierte Literaturreview des DNQP (durchgeführt in 2013) weist eine deutlich verbesserte Studienlage incl. einem erhöhten Angebot an Leitlinien auf als noch in 2006. Diskutiert werden explizit kontinenzfördernde Maßnahmen bei Personen mit (erhöhtem) Pflegebedarf (DNQP, 2014b). Ein systematisches Review von Donald et al., deren Durchführung sich an den Recherchekriterien der Cochrane Collaboration orientiert und die Evidence der Effektivität des Einsatzes von Advanced Practice Nurses im Langzeitpflegebereich untersucht, belegt deren positive Auswirkung auf das Auftreten von Harninkontinenz im Sinne einer Reduktion der Kontinenzrate oder der Verlangsamung einer Verschlechterung (Donald et al., 2013).

Wingenfeld et al. (2011) ordnen das Thema Harnkontinenz in die Rubrik der begrenzt einsetzbaren Indikatoren ein. Dessen Relevanz für die Pflege wird auch von Wingenfeld et al. (2011) konstatiert. Allerdings stellen die AutorInnen die Beeinflussbarkeit dieses Indikators durch Pflegeleistungen in Frage. Health Quality Ontario nutzt vier Indikatoren für die öffentliche Qualitätsberichterstattung. Inkontinenz, definiert als der Prozentsatz an BewohnerInnen, deren Blasenkontrolle sich verschlechtert, gehört dazu (Health Quality Ontario, 2014a). Auch

das niederländische Steering Committee hat die Anzahl der KlientInnen mit Harninkontinenz (die auch jene mit einem Katheter umfassen) als Outcomeindikator aufgenommen (ActiZ et al., 2007).

Die von Wingefeld et al. (2011) durchgeführte Erprobung gibt Hinweise auf unzureichende einrichtungsbezogene Interventionen. Gefragt war nach dem „Anteil der BewohnerInnen, die unter einer Harninkontinenz leiden und bei denen ein Maßnahmenplan zum Erhalt und/oder der Verbesserung des Kontinenzstatus vorlag“ (Wingefeld et al. 2011, 238). Das Ergebnis zeigt eine Spannweite, die bei den beteiligten Einrichtungen in der Vollerhebung zwischen 38,2% und 97,4% lag, ein Ergebnis, das die AutorInnen selbst als ‚überraschend‘ bewerten.

Angesichts der hohen inhaltlichen Relevanz sollte die Harnkontinenz in den Kanon der Mess-themen aufgenommen werden. Nun ist anzuerkennen, dass die vorliegenden Studien zur Einflussmöglichkeit von Pflegekräften auf die Kontinenzrate nicht aus dem deutschsprachigen Raum kommen und die Rahmenbedingungen und Aufgabenprofile möglicherweise nicht vergleichbar sind. Die Kontinenzrate als Outcomeindikator sollte verfolgt werden. Angesichts der aktuellen Unsicherheiten bzgl. ihrer Beeinflussbarkeit sollte kurzfristig zumindest der von Wingefeld et al. (2011) bereits entwickelte Indikator, der zwar kein Outcome, sondern eine Leistung der Einrichtung, also ein Output abbildet, aufrechterhalten werden.

Bereich 3: Unterstützung bei spezifischen Bedarfslagen

In inhaltlicher und methodischer Art unterscheidet sich der Bereich 3 von den Bereichen 1-2. Aus Sicht der Outcomemessung besonders relevant ist der Charakter dieser Themen als Output.

Benannt werden von Wingefeld et al. (2011) im Bereich 3

- Integrationsgespräch für Bewohner nach dem Heimeinzug,
- Einsatz von Gurtfixierungen,
- Einschätzung von Verhaltensauffälligkeiten bei Bewohnern mit kognitiven Einbu-ßen und
- Schmerzmanagement (Schmerzeinschätzung/Information über Schmerz).

Anders als in den Bereichen 1 und 2, deren Themen jeweils ähnlichen Kategorien angehören, findet sich hier eine deutliche inhaltliche Heterogenität. Das Integrationsgespräch nach Heimeinzug und der Einsatz von Gurtfixierungen z. B. eint nur ihr formaler Charakter als Outputkriterium.

Ein Output reflektiert die Leistungen einer Einrichtung. Die Auswertung eines Outputindikators gibt keine Auskunft über den Nutzen dieser Leistung für die EndverbraucherInnen, hier also die BewohnerInnen. Die Dokumentation eines Integrationsgespräches nach dem Heimeinzug beispielsweise lässt keine Erkenntnis darüber zu, ob dieses Integrationsgespräch zum rechten Zeitpunkt stattgefunden hat, die für die BewohnerInnen hilfreichen Themen aufgegriffen wurden bzw. insgesamt den Eingewöhnungsprozess unterstützt und das Wohlbefinden gesteigert hat. Wingefeld et al. (2011) weisen selbst darauf hin, dass die Forschungslage zur Frage einer effektiven Unterstützung nach dem Heimeinzug weitgehend offen ist. Vorstellbar ist auch, dass die Wünsche und der Bedarf der neu eingezogenen Be-

wohnerInnen unterschiedlich sind, so dass die Abbildung einer spezifizierten Leistung vor diesem Hintergrund auch gar nicht sinnvoll ist.

Zusammenfassend handelt es sich bei dem Integrationsgespräch für BewohnerInnen nach dem Heimeinzug um einen Indikator, der eine Leistung erfasst, die in ihrer Ausgestaltung nicht evidencebasiert ist und dessen Interpretation im Hinblick auf die BewohnerInnen auf Vermutungen angewiesen ist.

Das Projekt zur Entwicklung und Erprobung von Instrumenten zur Beurteilung der Ergebnisqualität zielte darauf ab, wo immer möglich, die Beschreibung von Prozessen oder von Leistungsendpunkten zu verlassen und nach dem Outcome, also der Wirkung von Leistungen auf die BewohnerInnen, zu schauen. Das ist ein richtiger und sinnvoller Ansatz, der seit einigen Jahren weltweit verfolgt wird (s. Kapitel 3). Nicht nur das niederländische Steering Committee Responsible Care macht Mut, die Erfahrungen der BewohnerInnen verstärkt zu nutzen (ActiZ et al., 2007).

Die GutachterInnen gehen wie Wingenfeld et al. (2011) davon aus, dass es sich beim Heimeinzug um ein kritisches Ereignis handelt, das besondere Aufmerksamkeit verdient. Jedoch sollte hier nicht der Output, sondern das Outcome gemessen werden. Eine geeignete Operationalisierung könnte z. B. den Grad der Integration widerspiegeln, wäre aber noch genauer zu prüfen.

Freiheit ist ein hohes Gut, das gilt auch und gerade für Menschen, die hilfs- und pflegebedürftig sind. Häufig mit dem Argument der Sicherheit werden in Alten- und Pflegeheimen Fixierungen unterschiedlicher Art eingesetzt. Dazu gehören Gurtfixierungen, der Einsatz von Bettgitter und Stecktischen. Fixiert werden insbesondere Personen mit kognitiven Einschränkungen, die herausforderndes Verhalten, häufig verbunden mit Bewegungsdrang, zeigen. Die Hamburger Studie zu Prävalenz und Inzidenz freiheitsentziehender Maßnahmen (n=2367) kommt zu einem Anteil an BewohnerInnen mit mindestens einer freiheitsbeschränkenden Maßnahme am Stichtag zwischen 4% bis 59%. In vier Einrichtungen betrug die Prävalenz 4%-8%, in acht Einrichtungen 10%-20%, in sieben Einrichtungen 22%-30%, in drei Einrichtungen 30%-37%, in weiteren sieben Einrichtungen 41%-49% und in einer Einrichtung 59%“ (Köpke, Gerlach, Möhler, Haut, & Meyer G, 2012). Als freiheitsbeschränkend galt der Einsatz von Bettgitter, Gurten und Stecktischen. Auch international, so in Kanada, den Niederlanden und den USA, wird die Fixierung als Qualitätsindikator diskutiert bzw. eingesetzt (Health Quality Ontario, 2013; ActiZ et al., 2007; RTI International, 2012). Für Deutschland liegt seit 2012 eine Praxisleitlinie zu freiheitseinschränkenden Maßnahmen in der beruflichen Altenpflege vor (Köpke et al., 2012).

Die ethisch-rechtliche Dimension freiheitsbeschränkender Maßnahmen verbunden mit der Häufigkeit des Vorkommens verweisen auf die Relevanz dieses Indikators. Eine große Bandbreite an Maßnahmen kann zur Beruhigung vom Menschen mit herausforderndem Verhalten beitragen und fallweise den Einsatz von Fixierungen überflüssig machen. Auch die Praxisleitlinie weist die ‚Häufigkeit von freiheitsbeschränkenden Maßnahmen‘ als Ergebnisparameter aus (Köpke et al., 2012). So wird dieser Indikator grundsätzlich von den GutachterInnen gestützt. Es bleibt einzig noch die Frage, welche Definition von Fixierung genutzt werden soll. Wingenfeld et al. (2011) schlagen vor, nur Gurtfixierungen zu erheben. Der Einsatz von Bettgitter und Stecktischen bleibt unberücksichtigt. Die von Wingenfeld et al.

(2011) vorgetragene Begründung der Eindeutigkeit ist im Hinblick auf externe Berichterstattung überzeugend. Gleichwohl darf nicht außer Acht gelassen werden, dass auch andere Formen nicht nur eine Einschränkung der freien Beweglichkeit darstellen, sondern ihrerseits gefährdend sein können. So sind diese Formen der freiheitsbeschränkenden Maßnahmen im Rahmen der internen Qualitätsentwicklung zu erfassen, zu analysieren und ggf. Maßnahmen zu erfassen, die auf Abhilfe abzielen.

Verhaltensauffälligkeiten von Menschen mit kognitiven Einschränkungen, auch bezeichnet als herausforderndes Verhalten, stellen eine der Gründe für die Übersiedlung in eine stationäre Pflegeeinrichtung dar. Doch auch für die Einrichtungen stellen aggressive, abwehrende oder apathische Verhaltensweisen ebenso eine Herausforderung dar wie andauerndes Rufen oder ein hoher Bewegungsdrang. Der Bedeutungsunterschied zwischen dem Terminus ‚Verhaltensauffälligkeit‘, der die Ursache in die betroffene Person verlagert und dem Begriff ‚herausforderndes Verhalten‘, der darauf verweist, dass ein Verhalten sowohl für die betroffene Person als auch ihre Umgebung eine Herausforderung darstellt, ist konzeptionell bedeutsam, da es das Spektrum gezielter Interventionen erweitert.

Die Prävalenzerhebungen von Menschen mit herausforderndem Verhalten in Einrichtungen der stationären Altenhilfe zeigen – nicht zuletzt im Zusammenhang mit einer unterschiedlichen Definition dieses Phänomens – eine große Bandbreite. Die Zahlen bewegen sich zwischen einem Anteil von 30% – 80% (Bundesministerium für Gesundheit, 2006).

Die grundsätzliche Relevanz dieses Themas ist unbestritten. Vor diesem Hintergrund beauftragte das Bundesministerium für Gesundheit die Erstellung einer Rahmenempfehlung zum Umgang mit herausforderndem Verhalten bei Menschen mit Demenz in der stationären Altenhilfe. Die Expertengruppe betont die Bedeutung verstehender Diagnostik, personenzentrierter Konzepte und der Milieugestaltung. Gezielte Schulungen können dazu beitragen, die Verhaltensweisen der Menschen mit Demenz einzuordnen und einen professionellen Umgang zu etablieren, was letztlich das Ziel hat, den BewohnerInnen mit kognitiven Einschränkungen Angst zu nehmen und Sicherheit zu vermitteln (Bundesministerium für Gesundheit, 2006).

Der von Wingefeld et al. (2011) vorgeschlagene Indikator umfasst, ob für BewohnerInnen mit kognitiven und kommunikativen Einschränkungen eine instrumentengestützte Einschätzung von Verhaltensweisen vorliegt. Des Weiteren wird nach der Operationalisierung gefragt, die das einrichtungsindividuell genutzte Instrument aufweist.

Es stellt sich die Frage, welche Aussagekraft dieser Indikator in der hier konstruierten Form hat. Zwar wird erfasst, ob eine instrumentengestützte Einschätzung von Verhaltensweisen vorliegt, aus dem vorliegenden Konstrukt ist aber nicht ersichtlich, welche Form der Verhaltensänderung vorliegt, sondern nur welche Formen durch das einrichtungsindividuelle Instrument erfasst werden. Die Aussagekraft dieser Daten erschließt sich nicht.

In Übereinstimmung mit Wingefeld et al. (2011) wird eine Outcomemessung, z. B. die Erhebung der Rate der BewohnerInnen mit herausforderndem Verhalten in einer Einrichtung, nicht für sinnvoll erachtet. Nicht zuletzt wäre bei einem solchen Indikator die Gefahr von Nebenwirkungen in Form eines verstärkten Einsatzes von Psychopharmaka nicht von der Hand zu weisen. Orientiert an den Rahmenempfehlungen sollte aber über die Erfragung einer in-

strumentengestützten Einschätzung hinaus die Durchführung von Fallbesprechungen erhoben werden.

Studien zufolge liegt der Anteil an BewohnerInnen mit andauernden Schmerzen im Pflegeheim bei 40% – 85% (National Quality Forum, 2011). Das Vorhandensein von andauernden Schmerzen führt u. a. zu Einschränkungen in der Beweglichkeit, Schlafstörungen und psychischen Auswirkungen bis hin zu depressiven Verstimmungen (Burfield, Wan, Sole, & Cooper, 2012). Die Einschätzung von Schmerzen hinsichtlich Häufigkeit, Lokalisierung und Intensität gehört vom Grundsatz her zum pflegerischen Auftrag. Gerade die Einschätzung der Schmerzintensität bei Menschen mit einer Demenzerkrankung gestaltet sich aber als äußerst schwierig.

Wingefeld et al. (2011) schlagen als Indikator die Erhebung von Schmerzen vor, ohne akute und chronische Schmerzen klar zu differenzieren. Im Hinblick auf die Outcomemessung ist deren Differenzierung aber sinnvoll. Eine weitere Differenzierung sollte auf Grundlage der Fähigkeit der BewohnerInnen zur Selbstauskunft getroffen werden.

Die Durchführung einer Schmerzeinschätzung – wie von Wingefeld et al. (2011) vorgeschlagen – stellt einen wichtigen ersten Schritt dar, der aber unseres Erachtens nicht ausreicht.

Für den Bereich der akuten Schmerzen bei auskunftsfähigen BewohnerInnen gibt der Expertenstandard ein definiertes Schmerzniveau (in Abhängigkeit von Ruhe oder Bewegung) vor (DNQP, 2011a). Hierbei handelt es sich um ein Ergebnismaß, das sehr gut als Outcomeindikator ausformuliert werden könnte. Die Qualitätsmessung in der stationären Altenhilfe sollte dieses Outcome aufgreifen. Auch das National Quality Forum empfiehlt die Erhebung des selbst berichteten Schmerzlevels (National Quality Forum, 2011). Im Bereich der chronischen Schmerzen ist die Definition eines allgemein akzeptablen Schmerzniveaus demgegenüber kaum möglich. In Ergänzung zum Vorschlag von Wingefeld et al. (2011) sollte in Übereinstimmung mit dem Expertenstandard zu chronischen Schmerzen (DNQP, 2014c) jedoch nicht nur die Schmerzeinschätzung erfolgen, sondern darüber hinaus das individuelle Schmerzmaß ermittelt und ein individueller Maßnahmeplan daran ausgerichtet werden.

Mit den hier diskutierten Indikatoren sind zentrale, aber keineswegs alle möglichen Indikatoren benannt. Die Einführung der Outcomemessung in der stationären Altenhilfe ist jedoch nicht nur eine aufwändige Angelegenheit, sondern durchaus auch eine, die hohe methodische Anforderungen an die Einrichtungen selbst stellt, die ihrerseits nun zuständig für die Datenerhebung sind. Alle Erfahrungen weisen darauf hin, dass es sinnvoller ist, hier schrittweise vorzugehen. Vor diesem Hintergrund wird es als zweckmäßig erachtet, mit einer überschaubaren Anzahl von Indikatoren zu beginnen und die Outcomemessung sukzessive auszubauen.

4.1.2 Wenn relevante Themenbereiche nicht abgedeckt sind, wie könnten diese einbezogen werden, wenn eine Indikatorenbildung methodisch nicht valide möglich ist?

Die Messung von Outcomes stellt einen wichtigen Teil der Qualitätsmessung dar. Es ist unbedingt zu begrüßen, dass Anstrengungen unternommen werden, diesen Bereich auszubauen, zielen Outcomeindikatoren doch darauf ab, die Wirkung von erbrachten Leistungen auf der Ebene der BewohnerInnen abzubilden. Die Darstellung korrekter Leistungsprozesse (mittels Prozessindikatoren) oder die Abbildung der Häufigkeit einer Leistung (mittels Outputindikatoren) haben demgegenüber im Hinblick auf den Nutzen eine vergleichbar geringere Aussagekraft.

Gleichwohl gibt es verschiedene Gründe dafür, Qualitätsmessung nicht ausschließlich auf der Outcomeebene zu belassen.

1. Einige Themen oder Anliegen lassen sich prinzipiell auf der Outcomeebene nicht abbilden.
2. Einige Themen sind zwar grundsätzlich auf der Outcomeebene abbildbar, es liegen aber aktuell zu wenige Kenntnisse vor oder der Aufwand, der für die Datenerhebung betrieben werden müsste, übersteigt den Nutzen.
3. Die angestrebte Kombination von externer Qualitätsprüfung und interner Qualitätsentwicklung erfordert nicht zuletzt auf der Ebene der Einrichtungen Kenntnisse über Leistungsprozesse und Strukturen, um Entscheidungen auf der Steuerungsebene begründet treffen zu können.
4. Das Informationsinteresse der Menschen mit Pflegebedarf und ihrer Angehörigen liegt überwiegend im Bereich von Prozessen und Strukturen.

Zu 1:

Als ein Beispiel verweisen Wingefeld et al. (2011) in Übereinstimmung mit den GutachterInnen auf ethische und methodische Grenzen der Outcomemessung im Rahmen der Begleitung sterbender BewohnerInnen.

Die Outcomemessung ist in ihren Möglichkeiten ebenfalls limitiert, wenn es um die Möglichkeiten der Beurteilung der Leistungsqualität durch die BewohnerInnen und Angehörige geht. Das niederländische Steering Committee for Responsible Care beansprucht mit seinem Konzept einerseits so viele Themen wie möglich auf der Outcomeebene abzubilden, andererseits das Erleben der BewohnerInnen selbst zum Maßstab zu machen und das keineswegs nur separiert in Form von ‚Zufriedenheit‘ oder ‚Lebensqualität‘, sondern auch unmittelbar bezogen auf Leistungsbereiche wie Pflege und Wohnen (z.B. Erfahrungen zur Pflege in ihren psychosozialen und körperbezogenen Dimensionen, zu Teilhabe, zu Kommunikation, zu den Mahlzeiten, zur Verfügbarkeit von Pflegekräften und anderen MitarbeiterInnen, zur Wohnatmosphäre usw.) (ActiZ et al., 2007). Diese Kombination, einerseits die Messung der objektivierte Outcomedaten, andererseits die konsequente Erhebung der Erfahrung der BewohnerInnen scheint ausgesprochen sinnvoll, da sie die fachliche gesundheitsbezogene Ebene mit der subjektiven Erlebensebene im Hinblick auf die Qualität der Dienstleistung verbindet. Zwar ist zu bedenken, dass ein Teil der BewohnerInnen aufgrund kognitiver Einschränkungen nur

bedingt Auskunft geben kann, jedoch ist dieser Tatbestand kein Grund, nicht jene zu befragen, die in der Lage sind, ihre Erfahrungen mitzuteilen. Dies ist kein Plädoyer gegen die Erhebung von Lebensqualität als Outcomeindikator, sondern für den Weg, den BewohnerInnen konsequent mehr Gewicht bei der Beurteilung der Qualität der in den Einrichtungen der stationären Altenhilfe erbrachten Dienstleistung einzuräumen.

Zu 2:

Einige Themen sind von inhaltlicher Relevanz, aktuell liegen aber noch nicht ausreichend Kenntnisse oder Kompetenzen vor, als dass eine evidenzbasierte Indikatorentwicklung und/oder –erhebung schon möglich wäre. Soll dieses Themenfeld gleichwohl in die Qualitätsmessung einbezogen werden, ist ersatz- und übergangsweise ein Prozess- oder Outputindikator zu nutzen.

Dabei ist jeweils sehr sorgfältig abzuwägen, ob und inwieweit nicht doch die Outcomemessung möglich und sinnvoll wäre. Ein Beispiel stellt das Thema Kontraktur dar. Kontrakturen gehen mit Schmerzen, Bewegungseinschränkungen und insgesamt mit einer Verringerung der Lebensqualität einher. Bei der Erprobung der von Wingenfeld et al. (2011, 232) entwickelten Outcomeindikatoren zeigte sich eine Inzidenz von 10,6%. Dabei wurde von den erhebenden Pflegekräften das Vorliegen einer Kontraktur eher unterschätzt. Gleichwohl handelt es sich um ein Thema, das auch international nur selten als Outcomeindikator genutzt wird. Zudem gibt es nicht unerhebliche definitorische Unklarheiten (Gnass, Bartoszek, Thiesemann, & Meyer, 2010). Es ist zu begrüßen, dass Wingenfeld et al (2011) sich dieses Themas angenommen haben. Aufgrund von Herausforderungen in der Praxis, die insbesondere die pflegerische Kompetenz zur Einschätzung einer Kontraktur angeht, nehmen Wingenfeld et al. (2011) Abstand von der Kontrakturerhebung als Indikator für die vergleichende Qualitätsberichterstattung. Die aktuellen Unsicherheiten der Pflegekräfte zur Einschätzung einer Einschränkung in der Gelenkbeweglichkeit sollten keineswegs grundsätzlich dazu führen, die Zielsetzung, die Inzidenz von Kontrakturen zu erheben, aufzugeben. Jedoch sollte die Erhebung eines solchen Indikators gut vorbereitet sein. Für den Übergangszeitraum könnten ersatzweise Maßnahmen zur Förderung der Gelenkbeweglichkeit dargestellt werden.

Nicht zuletzt ist das Verhältnis von Aufwand und Nutzen zu berücksichtigen, ein Aspekt, der in den Überlegungen der Expertengruppe für die Outcomemessung zur externen Qualitätsprüfung medizinischer Versorgung nach jahrelanger Erfahrung in der Indikatorenentwicklung einen immer größeren Stellenwert einnimmt (AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH, 2013).

Zu 3:

Die Herausforderungen bei der Verzahnung von interner Qualitätsentwicklung und externer Qualitätskontrolle sowie –berichterstattung sind in den Kapiteln 1.3.2 und 4.2.2 ausführlich dargestellt und sollen hier nicht im Detail wiederholt werden. Die Kenntnis eines Outcomes gibt keinen unmittelbaren Hinweis auf das Zustandekommen dieses Ergebnisses, dies gilt sowohl für positive als auch für negative Ergebnisse. Die Zusammenhänge, die zu diesen Resultaten geführt haben, bleiben im Dunkeln. Das Wissen um Prozessaspekte stellt einen

Beitrag dar, um Outcomes zu interpretieren (Weibler-Villalobos & Röhrig, 2010). Für die Qualitätsberichterstattung – insbesondere an das Fachpublikum – zählen die Ergebnisse, für die Einrichtung ist es jedoch wichtig zu wissen, welche Prozesse und Strukturen im Zusammenhang mit dem Outcome wirksam geworden sind. So ist es ein wichtiger Faktor, dass Prozessindikatoren eher Handlungsoptionen eröffnen (AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH, 2010).

Zu 4:

Das Informationsinteresse der VerbraucherInnen, hier verstanden als die Menschen mit Pflegebedarf sowie ihren Angehörigen, liegen nur sehr nachrangig im Bereich der gesundheitsbezogenen Outcomes (s. ausführlich Kapitel 4.5). Im Mittelpunkt des Verbraucherinteresses liegen die personalen Bedingungen einer Einrichtung. Gefragt wird nach Eigenschaften, die sehr fundamental die Qualität der menschlichen Beziehung betreffen verbunden mit der Frage danach, ob Pflegekräfte und andere MitarbeiterInnen dann zur Verfügung stehen, wenn sie gebraucht werden.

Auch pflegefachlich und pflegewissenschaftlich betrachtet ist der Zusammenhang zwischen der Personalausstattung und der Pflegequalität, dem pflegerischen Outcome, nicht von der Hand zu weisen (Horn, Buerhaus, Bergstrom, & Smout, 2005; Kirkevoold & Engedal, 2008; Weibler-Villalobos & Röhrig, 2010). Zur Personalausstattung gehört die Frage der Qualifikation ebenso wie die Fluktuation. Verschiedene Studien weisen auf einen Zusammenhang von Personalausstattung und dem Grad der funktionellen Selbständigkeit der BewohnerInnen, Dekubitusinzidenz und unbeabsichtigtem Gewichtsverlust hin (Bostick, Rantz, Flesner, & Riggs, 2006).

Vor diesem Hintergrund haben verschiedene Systeme der Qualitätsberichterstattung wie das Nursing Home Compare (Center for Medicare and Medicaid Services, 2012) und das niederländische Responsible Care (ActiZ et al., 2007) die Personalausstattung bzw. die Verfügbarkeit von (Pflege-)Personal in den Katalog zur Berichterstattung aufgenommen.

Dieses Vorgehen wird von den GutachterInnen ausdrücklich begrüßt und auch für die Qualitätsberichterstattung in Deutschland empfohlen.

Im Hinblick auf die Fragestellung dieses Kapitels besteht zusammenfassend große Übereinstimmung, dass soweit möglich Outcomeindikatoren genutzt werden sollen, dass diese jedoch nicht ausreichen und durch Prozess- und Strukturindikatoren ergänzt werden müssen.

4.1.3 Bestimmte Personen werden aus der Bewertung ausgeschlossen. Sind die jeweils definierten Ausschlusskriterien (z. B. Schlaganfall oder akute Verschlechterung einer Gesundheitsstörung), die zum Ausschluss aus der Bewertung führen, akzeptabel oder führen sie zu einer systematischen Verzerrung der Qualitätsberichterstattung?

Insbesondere im Hinblick auf die Validität der Outcomemessungen und damit verbunden auch der Zuverlässigkeit des im Zuge der Qualitätsberichterstattung angestrebten Anbietervergleichs kommt der methodischen Qualität der Indikatordefinitionen eine zentrale Rolle zu. Die Ergebnisse können dabei nicht unwesentlich von den definierten Ausschlusskriterien abhängen, anhand derer bestimmte Personengruppen von der Bewertungssystematik ausgeschlossen werden. Die Ausschlüsse resultieren dabei in der Regel aus Fragen der sogenannten Indikatorevidenz, bei der es u. a. darum geht, ob tatsächlich belegbar ist, dass Pflegeeinrichtungen die Ergebnisse tatsächlich auch positiv beeinflussen können (Sangl, Saliba, Gifford, & Hittle, 2005; Weibler-Villalobos & Röhrig, 2010). So dürfte es beispielsweise ohne Rückgriff auf paranormale Kräfte oder sonstige Illusionen recht schwierig sein, die soziale Integration und Kontakthäufigkeit von Komapatienten und -patientinnen nachhaltig zu fördern. Letztendlich geht es um primär auf inhaltlich-definitiver Ebene zu entscheidende Fragen, zu welcher Bezugsgröße die Häufigkeit bestimmter Indikatorereignisse wie Stürze oder entstandene Dekubitusgeschwüre in Relation gesetzt werden kann. Da anhand dieser Relation das konkrete Ausmaß der Qualität bemessen wird, hat dies einen unmittelbaren Einfluss auf das Messergebnis. Der Ausschlusslogik liegen prinzipiell ähnliche Überlegungen zu Grunde, die auch für die Risikoadjustierung maßgeblich sind. Genauer gesagt können Ausschlüsse auch als basale Form der Risikoadjustierung aufgefasst werden (vgl. Kapitel 4.4). Diese Querverbindung wird übrigens im Hinblick auf unsere Empfehlungen noch ganz elementar. Was ist hier eigentlich konkret gemeint? Im vorliegenden Beispiel eines Qualitätsindikators „soziale Kontakte“ könnten bei Einschluss aller BewohnerInnen etwa Einrichtungen mit einem sehr hohen Anteil komatöser BewohnerInnen bei gleicher Pflegequalität allein schon aufgrund dieses „Handicaps“ verhältnismäßig weniger Erfolge erzielen als Einrichtungen, die grundsätzlich nur BewohnerInnen ohne gravierende Gesundheitseinschränkungen aufnehmen. Es ist unmittelbar einsichtig, dass eine brauchbare vergleichende Outcomemessung so nicht gewährleistet werden kann. Durch den Ausschluss bzw. die Adjustierung wird erst eine faire Vergleichsgrundlage hergestellt („risikostandardisierte Fallkonstellation“). Hätten alle Anbieter den gleichen Anteil komatöser BewohnerInnen, würde sich die Ausschlussfrage so übrigens gar nicht stellen, da die Relation in diesem Fall immer gleich bliebe.

Entsprechend werden abgesehen vom Indikator „Schmerzmanagement“ auch bei allen Wingenfeld-Indikatoren BewohnerInnen mit bestimmten Eigenschaften durch die Indikatoroperationalisierung ausgeschlossen. Einen zusammenfassenden Überblick über alle von Wingenfeld et al. (2011) definierten Ausschlüsse gibt Tabelle 5. In inhaltlicher Hinsicht erscheinen die vorgenommenen Ausschließungen weitestgehend plausibel und schlüssig. Auch wenn aufgrund der insbesondere in Detailfragen divergierenden Indikatoroperationalisierungen eine 1:1-Vergleichbarkeit nicht gegeben ist, finden sich insbesondere, was den Ausschluss von komatösen Bewohnern und Bewohnerinnen angeht, in der

vorliegenden Literatur oft analoge Kriterien, die punktuell natürlich abweichen können (vgl. etwa Arling, Kane, Lewis, & Mueller, 2005). So werden beispielsweise im Rahmen des Indikators Dekubitusentstehung mitunter auch Bewohner und Bewohnerinnen ausgeschlossen, bei denen das Druckliegegeschwür bereits außerhalb der Einrichtung zu Stande kam, z. B. im Krankenhaus (Abel et al., 2005). Das Grundproblem mit den Ausschlüssen liegt aus Sicht der GutachterInnen daher auch weniger in den inhaltlichen Begründungen und den verwendeten Dimensionen an sich, sondern ist vielmehr auf dem Hintergrund ihrer Messbarkeit angesiedelt. Damit eng verknüpft ist neben der Frage der Validität ihrer Dokumentation auch die Frage nach deren Transparenz.

Tabelle 5: Übersicht Ausschlusskriterien bei den Wingefeld-Indikatoren

Qualitätsbereich	Indikatorbereich	Ausgeschlossene BewohnerInnen
Bereich 1: Erhaltung und Förderung von Selbständigkeit	Mobilität	<ul style="list-style-type: none"> • Koma/Somnolenz • gravierendes Krankheitsereignis (<i>Fraktur, Apoplex, Herzinfarkt, OP an Gliedmaßen, akute Verschlechterung, Gesundheitsstörung mit erheblichem/dauerhaftem Einfluss auf Selbständigkeit</i>)
	Alltagsverrichtungen	<ul style="list-style-type: none"> • Koma/Somnolenz • gravierendes Krankheitsereignis (<i>Fraktur, Apoplex, Herzinfarkt, OP an Gliedmaßen, akute Verschlechterung, Gesundheitsstörung mit erheblichem/dauerhaftem Einfluss auf Selbständigkeit</i>)
	Gestaltung des Alltagslebens und sozialer Kontakte	<ul style="list-style-type: none"> • Koma/Somnolenz • gravierendes Krankheitsereignis (<i>Fraktur, Apoplex, Herzinfarkt, OP an Gliedmaßen, akute Verschlechterung, Gesundheitsstörung mit erheblichem/dauerhaftem Einfluss auf Selbständigkeit</i>) • völliger/weitgehender Fähigkeitsverlust im Bereich Kognition und Kommunikation
Bereich 2: Schutz vor gesundheitlichen Schädigungen und Belastungen	Dekubitusentstehung	<ul style="list-style-type: none"> • nachweisliche Ablehnung von prophylaktischen Maßnahmen (<i>Mobilisierung, Lagerung</i>)
	Stürze mit gravierenden Folgen	<ul style="list-style-type: none"> • komplett immobile Bewohner (<i>auch im Bett</i>)
	Unbeabsichtigter Gewichtsverlust	<ul style="list-style-type: none"> • medizinische Gründen ausschlaggebend (<i>Krebserkrank., Amputation, medikamentöse Aufschwemmung</i>)
Bereich 3: Unterstützung bei spezifischen Bedarfslagen	Integrationsgespräch nach dem Heimeinzug	<ul style="list-style-type: none"> • länger als sechs Monate in der Einrichtung
	Einsatz von Gurtfixierungen	<ul style="list-style-type: none"> • komatöse BewohnerInnen • keine/geringe kognitive Beeinträchtigungen (<i>NBA Modul 2</i>) • keine Beeinträchtigung Verhaltensweisen u. psych. Problemlagen (<i>NBA Modul 3</i>)
	Einschätzung von Verhaltensauffälligkeiten bei Bewohnern mit kognitiven Einbußen	<ul style="list-style-type: none"> • komatöse BewohnerInnen
	Schmerzmanagement	<ul style="list-style-type: none"> • keine Ausschlüsse

Je „weicher“ die Kriterien sind bzw. salopp gesagt, je mehr sie im Sinne von oberflächlichen „Pi-mal-Daumen“-Einschätzungen mit weiten Ermessensspielräumen gehandhabt werden können, desto mehr ist auch das Qualitätsberichterstattungssystem als solches zumindest dem Verdacht der Willkürlichkeit und fehlenden Objektivität ausgesetzt. Da die ausgeschlossenen Personen in den Outcome-Ergebnissen selbst nicht unmittelbar ersichtlich sind, also gewissermaßen unter den Tisch fallen, ist es aufgrund der unmittelbaren Ergebnisrelevanz durchaus möglich, durch eine „kreative“ Gestaltung einer im Detail nicht nachvollziehbaren Ausschlusspraxis die Qualitätsbewertung der Einrichtung sogar willentlich zu beeinflussen

bzw. unverblümter gesagt zu verfälschen. Hier mustergültig einige Beispiele für aufgrund unzureichender Objektivierbarkeit für Verzerrungen anfällige Ausschlussbereiche innerhalb des Indikatorensystems von Wingenfeld et al. (2011):

- In den Indikator „Dekubitusentstehung“ sollen keine BewohnerInnen einfließen, die prophylaktische Mobilisierungs- oder Lagerungsmaßnahmen ablehnen (Wingenfeld et al., 2011). Die Autoren und Autorinnen schränken aber selbst ein, dass bei annähernd der Hälfte der ausgeschlossenen BewohnerInnen aufgrund erheblicher kognitiver Einbußen nicht unmittelbar nachvollziehbar war, worauf sich diese Willenserklärung konkret stützte.
- Beim Indikator „Unbeabsichtigter Gewichtsverlust“ werden Personen ausgeschlossen, bei denen der Gewichtsverlust medizinische Ursachen hat. Ob eine medizinische Begründung vorlag oder nicht, wird lediglich anhand von freischriftlich dokumentierten Daten bemessen. Inwieweit im Pflegealltag in allen Fällen eine lückenlose schriftliche Dokumentation gewährleistet werden kann, bleibt eine offene Frage.

Natürlich könnte man vor diesem Hintergrund argumentieren, dass mit Blick auf die Objektivität der Ergebnisdokumentation sowieso regelmäßige unabhängige Prüfungen unerlässlich sind (vgl. auch Kapitel 4.2.3). Eine auf eher weichen Daten basierende Dokumentationspraxis entzieht sich aber auch hier weitestgehend den Kontrollmöglichkeiten. Soweit Ausschlüsse im Zuge der Indikatoroperationalisierung inhaltlich indiziert sind, empfehlen wir daher in der Dokumentation standardisierte und möglichst bewährte Testverfahren einzusetzen. Es ist uns an dieser Stelle durchaus bewusst, dass in der Pflegeforschung ein breiter Diskurs über die Güte von Skalen geführt wird, der teilweise zu kritischen Betrachtungen verschiedener Verfahren geführt hat. Trotzdem wäre es nicht sinnvoll, die Forderung nach testgütemäßig abgesicherten Verfahren aufzugeben. In Teilbereichen wird dies auch schon im Rahmen der Wingenfeld-Indikatoren versucht, indem die Autoren etwa beim Qualitätsindikator „Einsatz von Gurtfixierungen“ die Ausschlüsse anhand der beiden Module „Mobilität“ sowie „Kognitive und kommunikative Fähigkeiten“ des Neuen Begutachtungsassessments NBA (Wingenfeld, Büscher, & Gansweid, 2008) operationalisieren (vgl. Tabelle 5). Auch wenn hier zentrale test-theoretische Fragen noch nicht durchgehend geklärt sind (s. Kapitel 4.4.1), geht dieser Ansatz aber definitiv in die richtige Richtung. Auch bei anderen relevanten Ausschlüssen wie beispielsweise Koma sollte auf nachvollziehbare und transparente Verfahren zurückgegriffen werden. Zu prüfen wäre die Eignung der Glasgow Coma Skale. So enthält z. B. auch die in Kapitel 4.2.3 näher beschriebene CPS-Skala ein Koma-Screening (vgl. Paquay et al., 2007). Grundsätzlich wäre dazu zu raten, die Erhebung der Ausschlusskriterien fest in die Dokumentationssystematik zu integrieren.

Einige Autoren (vgl. Weibler-Villalobos & Röhrig, 2010; Nazir, Mueller, Perkins, & Arling, 2012) sehen Ausschlüsse von Risiken im Pflegebereich als grundlegend kritisch an und empfehlen entweder völlig oder zumindest bei bestimmten Risikogruppen wie bei Bewohnern bzw. Bewohnerinnen mit kognitiver Beeinträchtigung oder Demenz darauf zu verzichten. Neben den Fragen der Transparenz und Objektivierbarkeit ist allen Ausschlüssen ein elementares statistisches Problem inhärent, das sozusagen in der Natur der Sache liegt: Sie führen zu einer teils nicht unerheblichen Ausdünnung der Stichprobe und gehen damit unweigerlich mit der für kleinere Stichproben charakteristischen höheren Messfehleranfälligkeit einher (vgl. Arling, Lewis, Kane, Mueller, & Flood, 2007). Auch die Gutachter der vorliegenden Expertise

empfehlen in der Summe der umrissenen Einschränkungen, mit Ausschließungen möglichst sparsam umzugehen und nur dort vorzunehmen, wo es unerlässlich ist. Da es außerordentlich schwierig ist, hier indikatorbezogene Empfehlungen auszusprechen, bietet sich ein möglicher Ausweg, der sich sogar in mehrfacher Hinsicht als vorteilhaft erweisen könnte, auf einer ganz anderen Ebene an. Wie bereits einleitend beschrieben, steht die Diskussion rund um die Ausschlusskriterien in enger Beziehung zur Risikoadjustierung. In Kapitel 4.4 empfehlen wir den vorliegenden, aufgrund der Stratifizierungstechnik im Hinblick auf die Zahl potenziell abbildbarer Risikofaktoren stark limitierten Zugang durch eine regressionsbasierte multivariate Adjustierungstechnik zu ersetzen, die Art und Richtung des Einflusses, eine Vielzahl von Determinanten und deren Wechselwirkung auf die jeweiligen Outcome-Indikatoren berücksichtigen kann. In diesem Zusammenhang wäre es nur naheliegend, die hier definierten Ausschlüsse als weitere Risikofaktoren in die Modellbildung aufzunehmen. Die statistische Kontrolle würde dann über die Adjustierungstechnik erfolgen. Bei einer solchen Vorgehensweise könnten Ausschlüsse im klassischen Verständnis tatsächlich weitestgehend obsolet werden. Durch die adjustierten Kennwerte wäre eine Vergleichbarkeit der Einrichtungen trotzdem gewährleistet (vgl. Kapitel 4.4.1).

4.1.4 Fazit

Mit dem von Wingenfeld et al. (2011) vorgelegten Entwurf für Indikatoren zur Outcomemessung in der stationären Altenhilfe ist ein wichtiger Startpunkt gesetzt, der nun weiter verfolgt werden sollte.

Wenn Outcomemessung in der stationären Altenhilfe durchgeführt wird, dann gehören die Bereiche Selbständigkeit und Sicherheit sowie weitere Themen wie Fixierung und Schmerzmanagement in der Regel dazu. Betrachtet man jedoch die vorgeschlagene Reichweite und -tiefe der Wingenfeld-Indikatoren, dann zeigt sich Entwicklungsbedarf.

Die Indikatoren zu Mobilität und Selbständigkeitserhalt oder –verbesserung bei der Durchführung der Alltagsverrichtungen werden prinzipiell unterstützt. Die Erhebung des Selbständigkeitserhalts oder der -verbesserung bei der Gestaltung des Alltagskontakte allein als Fremdbeobachtung erscheint aber schwierig und sollte um den Abgleich mit der Erhebung des Erlebens der BewohnerInnen bezogen auf dieses Themenfeld ergänzt werden. Die Messung der Dekubitusinzidenz, der Stürze mit gravierenden Folgen und des unbeabsichtigten Gewichtsverlustes als Outcome entspricht prinzipiell dem state of the art. Einzig die Erhebung der Dekubitusinzidenz bei BewohnerInnen mit geringem Risiko scheint nicht zwingend erforderlich. Auch sollte der Gewichtsverlust nicht nur nach 6 Monaten, sondern schon nach 4 Wochen erkannt werden. Angesichts der hohen inhaltlichen Relevanz sollte die Harnkontinenz zusätzlich in den Kanon der Messthemen primär als Outputparameter aufgenommen werden. Im Bereich der spezifischen Bedarfslagen überzeugt der vorgeschlagene Indikator ‚Integrationsgespräch nach Heimeinzug‘ nicht, da er eine Leistung erfasst, die in ihrer Ausgestaltung nicht evidencebasiert ist und dessen Interpretation im Hinblick auf die BewohnerInnen auf Vermutungen angewiesen ist. Der Indikator zur Gurtfixierung wird in der vorgeschlagenen Form für sinnvoll erachtet. Anders sieht dies aus bei der Einschätzung von Verhaltensauffälligkeiten bei BewohnerInnen mit kognitiven Einbußen. Die grundsätzliche Relevanz des Indikators ist unbestritten, die Form der Operationalisierung erscheint aber nicht schlüssig. Zudem sollte dieser Outputparameter nicht allein das Vorhandensein einer instrumen-

tengestützten Einschätzung erfragen, sondern um die Durchführung von Fallbesprechungen erweitert werden. Der Indikator zum Schmerzmanagement sollte ausdifferenziert (akut/chronisch), im Bereich des akuten Schmerzes zu einem Outcomeindikator ausgebaut und im Bereich des chronischen Schmerzes um einen individuellen Maßnahmeplan ergänzt werden.

Es wird begrüßt, dass die Qualitätsmessung in der stationären Altenhilfe in Richtung auf Outcomemessung weiterentwickelt wird, diese ist jedoch durch Prozess- und Strukturindikatoren zu ergänzen.

In der von Wingefeld et al. (2011) vorgelegten Systematik sind Ausschlüsse letztendlich unumgänglich, da ein fairer Einrichtungsvergleich hier sonst nicht zu gewährleisten wäre. Der Ausschluss etwa von Komapatienten, deren Situation bei bestimmten Outcomes durch pflegerisches Handeln nicht beeinflusst werden kann, stellt in diesem Zusammenhang daher auch eine Form der Risikoadjustierung dar. Da die Notwendigkeit von Ausschlüssen in der Literatur generell kritisch hinterfragt wird, sollten diese nur in begründeten Fällen vorgenommen werden und dann am besten über ein standardisiertes Testverfahren in transparenter und valider Weise dokumentiert werden. Auf Ausschlüsse komplett verzichtet werden könnte im Rahmen der von uns vorgeschlagenen alternativen Risikoadjustierungsmethode. Im Rahmen dieses multivariaten Ansatzes könnten die Ausschlusskriterien einfach als zusätzliche Risikofaktoren einfließen und kontrolliert werden.

4.2 Datengrundlage/Externe Qualitätsprüfung

(Renate Stemmer)

4.2.1 Sind für eine Bewertung der Indikatoren neben der Pflegedokumentation weitere Datenquellen notwendig?

Die Rohdaten für gesundheitsbezogene Outcomindikatoren müssen an eine Auswertungsstelle geliefert, dort auf Plausibilität geprüft und berechnet werden.

Wie ist nun die Ausgangslage hinsichtlich der zu liefernden Rohdaten einzuschätzen? Mit Ausnahme der Daten der Pflegekassen liegen bewohnerbezogene Daten in Deutschland ausschließlich einrichtungs- bzw. trägerbezogen vor. Auf dieser Ebene stellt die Pflegedokumentation die wesentliche Quelle für die Datenerhebung zur Messung der gesundheitsbezogenen Outcomeindikatoren dar.

Aber verschiedene Herausforderungen sind nicht zu übersehen. So bietet der Markt eine Vielzahl von uneinheitlichen Pflegedokumentationssystemen an, die von den Einrichtungen frei ausgewählt werden können. Diese sind papierbasiert oder elektronisch, sehen eine frei Formulierung der Pflegeprobleme und –interventionen vor, enthalten hausinterne Standards und/ oder nutzen eine Fachsprache wie ENP (European Nursing Care Pathways). Zusammenfassend sind die Dokumentationssysteme höchst unterschiedlich. Zudem ist der Umfang der Dokumentationspflichten immer umfangreicher und sind die Dokumentationsunterlagen immer unübersichtlicher geworden, so dass unter dem Stichwort ‚Bürokratieabbau‘ mittlerweile mehrere Initiativen auf Bundes- und Landesebene gestartet wurden mit dem Ziel, den Dokumentationsaufwand in der stationären und ambulanten Pflege zu reduzieren.

Diese Vielfalt stellt im Hinblick auf die erforderliche Datenqualität (s. Kapitel 4.2.3) eine erhebliche Hürde dar. Aus Sicht des Managements der Outcomemessung ist eine einheitliche edv-basierte Dokumentation zumindest der bewohnerbezogenen Assessments nicht nur äußerst wünschenswert, sondern unerlässlich. Beispielhaft sei auf die US-amerikanischen Pflegeheime verwiesen, die Medicare-Leistungen beziehen und verpflichtend computerbasiert das RAI bzw. MDS einsetzen. Aus diesem System können unmittelbar wesentliche Daten für die Outcomemessung generiert werden.

Die Situation in Deutschland ist grundlegend anders. Hier wäre es derzeit erforderlich, die Daten für die Outcomemessung aus den individuellen höchst unterschiedlichen Pflegedokumentationen zu extrahieren. Einerseits sind die vorliegenden Daten in keiner Weise aufbereitet oder austauschfähig. Andererseits ist zu vermuten, dass aktuell grundsätzlich Daten für den Bereich 2, dem Schutz vor gesundheitlichen Schädigungen und Belastungen sowie teilweise für den Bereich 3 erhoben werden. Dezidierte Angaben z.B. zum Selbständigkeitserhalt bei der Gestaltung des Alltagslebens (Bereich 1) der Wingefeld-Indikatoren werden schon seltener in den Dokumentationen zu finden sein.

Zusammenfassend ist es kaum vorstellbar, dass es gelingen kann, die erforderlichen Daten in der gewünschten Differenziertheit zur Verfügung zu stellen ohne die Pflegedokumentation grundsätzlich umzustellen. Erforderlich sind ja nicht nur Daten für die unmittelbaren Indikatoren, wie die Dokumentation eines Dekubitus, sondern auch für die Risikoadjustierung. Des Weiteren wäre der Übertragungsprozess höchst aufwändig und fehleranfällig. Der Start der

Outcomemessung auf Grundlage der aktuellen Dokumentationssituation bedeutet einen erheblichen Mehraufwand für die Einrichtungen mit einem absehbar schlechten Ergebnis, was die Qualität der Rohdaten und damit auch die Interpretationsfähigkeit der ausgewerteten Outcomes angeht, der so nicht empfohlen werden kann.

Es bedarf einer weitgehenden Umstellung.

Erforderlich ist

- die verpflichtende Einführung einer edv-gestützten Pflegedokumentation,
- die Sicherstellung der Kompatibilität der Schnittstelle zur Übermittlungssoftware,
- die Herstellung eines Konsenses bzgl. eines einheitlichen Dateiformates (z.B. Health Level 7), das den Austausch der erbetenen Daten ermöglicht,
- die Integration der erforderlichen Items für die Outcomemessung in die Dokumentationssoftware.

Bereits heute nutzen ca. 45% der stationären Altenhilfeeinrichtungen eine edv-gestützte Pflegedokumentation, ca. 10% planen eine entsprechende Umstellung (Meißner & Althammer, 2012). Zwar ist hiermit noch nicht klar, wie genau die EDV genutzt wird, ob beispielsweise mit mobilen Endgeräten gearbeitet wird oder ob die Daten im Stützpunkt nachträglich eingetragen werden, gleichwohl gibt es hier in etwa der Hälfte der Einrichtungen eine Ressource, auf die aufgebaut werden kann.

Jenseits der Dokumentationsgrundlage bleibt bei der Nutzung von Daten aus der Pflegedokumentation die Frage nach Abweichungen im Prozess der Datenerhebung durch die Pflegepersonen aufgrund der – trotz Schulung – unterschiedlichen Interpretationen der einzelnen Items (Mor, 2005). Zudem gibt es Grund zur Vermutung, dass Pflegeheime mit einer höheren Qualität aufmerksamer sind, wenn es darum geht, Pflegeprobleme zu entdecken und zu dokumentieren. Einrichtungen mit schlechter geschulten Pflegekräften nehmen Pflegeprobleme evtl. nicht wahr, dokumentieren diese entsprechend auch nicht und haben dann in der Auswertung der Outcomedaten entsprechend niedrigere Angaben zu unerwünschten Ereignissen (Castle & Ferguson, 2010). Es ist also parallel zur Herstellung der technischen Voraussetzung für die Outcomemessung auch die Kompetenz zur Datenerhebung zu schulen.

Aus Sicht der GutachterInnen geht es zusammenfassend nicht darum, weitere Datenquellen neben der Pflegedokumentation für die Datenerhebung für die Outcomemessung heranzuziehen, sondern die Pflegedokumentation in den Stand zu versetzen, die Daten zu liefern, die für die Outcomemessung benötigt werden.

4.2.2 Wie kann erreicht werden, dass die Daten des internen Qualitätsmanagements flächendeckend eine zuverlässige und valide Datengrundlage liefern?

Vorweg bedarf die Datenerhebung für die externe Qualitätsprüfung eines reliablen und validen Instrumentariums. Dieses zu entwickeln ist ein herausfordernder Prozess (Mor, 2006; Mor, Angelelli, Gifford, Morris, & Moore, 2003).

Doch auch wenn dieses Instrumentarium vorliegt, ist eine zuverlässige Datenerhebung noch nicht gewährleistet. Wie in Kapitel 4.2.2. bereits ausgeführt, ist die Zielsetzung von interner Qualitätsentwicklung und externer Qualitätsprüfung auch hinsichtlich der Datenerhebung keineswegs deckungsgleich mit Konsequenzen für die Auswahl der Daten und die Qualität der Datenerhebung.

Primäres Ziel des internen Qualitätsmanagements ist die Optimierung der Versorgungsleistungen einer spezifischen Einrichtung, das Vorgehen ist prospektiv und bezogen auf die Mitverantwortung aller MitarbeiterInnen. Maßnahmen dieses Bereiches können die Einrichtung von Qualitätszirkeln, Fallbesprechungen und Pflegevisiten mit Ergebnisbesprechungen sein, die letztlich jeweils darauf abzielen, einen gemeinsamen Lernprozess anzustoßen (Ollenschläger, 2007; Gebert & Kneubühler, 2003).

Verschiedene Ansätze zielen darauf ab, die Verzahnung interner Qualitätsentwicklung und externer Qualitätsüberprüfung sowie –berichterstattung zu unterstützen. Health Quality Ontario bietet den Einrichtungen mit einem Qualitätskompass Orientierung an. Der Qualitätskompass identifiziert nicht nur prioritäre Bereiche, sondern dieser stellt den Einrichtungen umfangreiche Informationen, Strategien und Hilfen zur Verfügung, um die interne Qualitätsentwicklung im Hinblick auf diese empfohlenen Prioritäten zu gestalten. Dieses Instrument benennt die empfohlene Praxis, Ideen zur Veränderung, Zielsetzungen und Maße, Instrumente und Ressourcen, die der Qualitätsentwicklung dienen (Health Quality Ontario, 2014b).

Auch das britische Department of Health gibt im Essence of Care eine ähnlich umfangreiche Handlungsempfehlung, wie ein internes Qualitätsmanagement zu betreiben ist, sodass es für ein externes Benchmarking genutzt werden kann. Hier wird ein Prozess aus neun Schritten beschrieben. Der Prozess beginnt mit der Prioritätensetzung und führt die Einrichtungen über eine ausführliche Erläuterung der klassischen Entwicklungsschritte durch den Qualitätsentwicklungsprozess. Jeder Schritt ist mit zahlreichen Ideen und Beispielen unterlegt. Die zentralen inhaltlichen Themen, die als Fokus des Essence of Care gelten, sind benannt. Jedoch sind diese nicht erschöpfend und Qualitätsentwicklung sollte sich keineswegs nur auf die ausgewiesenen Themen beziehen (Department of Health, 2010).

Diese methodischen Hinweise sollen unterstützend wirken. Das Besondere ist jedoch der Einsatz der Indikatoren und den damit verbundenen Daten für unterschiedliche Zwecke. Das niederländische Steering Committee weist darauf hin, dass es ausschlaggebend für den Erfolg dieses Ansatzes ist, ob es gelingt eine Balance zwischen der internen Nutzung und der Nutzung für externe Prüfung herzustellen (ActiZ et al., 2007).

In dem niederländischen und dem britischen Beispiel wird den Einrichtungen ein komplettes System zur Verfügung gestellt, das zentrale Themen identifiziert, diese in Indikatoren über-

setzt und operationalisierte Erhebungsverfahren zur Verfügung stellt. Die Daten werden zentral ausgewertet und den Einrichtungen für den eigenen Qualitätsentwicklungsprozess zur Verfügung gestellt, jenseits der Frage der externen Qualitätsprüfung und Qualitätsberichterstattung.

Dies kann als Unterstützung erlebt und bewertet, es kann aber auch als aufoktroiert wahrgenommen werden. Entscheidend ist, wie Qualitätsentwicklung und Qualitätsmessung verstanden werden. So kann der Input von außen als Chance zur Verbesserung der eigenen Leistungen genutzt werden oder es ist ein weiterer Auftrag, der abzuarbeiten ist (Dellefield, 2007). Die Qualitätsentwicklung ist eng mit den organisatorischen Bedingungen verbunden. Es zeigt sich, dass eher diejenigen Einrichtungen gegenüber Fragen der Qualitätsentwicklung aufgeschlossen sind, die grundsätzlich offen für Innovationen sind und eine Kultur des Teamworks leben (Mor, 2005).

Oftmals stellt ein Top-Down-Vorgehen bei der Implementierung von einrichtungsbezogenen Erhebungen von Qualitätsindikatoren ein Hindernis dar, das in mangelnder Akzeptanz mündet oder Neuerung werden unhinterfragt übernommen und der Effekt für die Einrichtung ist gering. Zudem findet sich auf der Ebene der Pflegekräfte häufig die Annahme, dass die durchzuführende Messung sowohl für die Pflegepraxis als auch für die pflegebedürftigen Menschen nicht relevant sei (Dunckley, Aspinall, Addington-Hall, Hughes, & Higginson, 2005).

Für den Erfolg dieser Kombination von interner Qualitätsentwicklung und externer Qualitätsmessung wird es entscheidend sein, ob es einer Einrichtung gelingt, das externe System in eigene Zielsetzungen zur Qualitätsentwicklung zu integrieren. Dazu gehört, alle betroffenen Abteilungen und Disziplinen am Prozess der Qualitätsentwicklung unter Nutzung der externen Vorgaben zu beteiligen und auch interne und externe Stakeholder der Einrichtung einzubeziehen. Die Benennung eines/einer Qualitätsbeauftragten ist unabdingbar. Sämtliches Personal der Einrichtung muss im Hinblick auf die Qualitätsentwicklung geschult werden. Auch sollten bestehende Ressourcen bestmöglich genutzt werden. Um das Personal für den Prozess der Entwicklung zu motivieren, sind Wertschätzung und ein Anreiz- und Belohnungssystem unterstützend. Zur Mitarbeitermotivation gehört ein zeitnahes Feedback, welches bislang nicht immer selbstverständlich ist. Eine besondere Herausforderung stellt die Aufrechterhaltung der Verfahrensweisen, die zu einer Qualitätsverbesserung geführt haben, dar. So zeigen Studien, dass oftmals nach erfolgreicher Implementierung die routinemäßigen Prozesse wieder aufgenommen werden (Compas, Hopkins, & Townsley, 2008; Gebert & Kneubühler, 2003; Ollenschläger, 2007).

Viele Einrichtungen in Deutschland orientieren sich bei ihren Bemühungen um interne Qualitätsentwicklung an den Vorgaben der Qualitätsprüfungsrichtlinien (QPR). Diese werden zu Grundlagen für Qualitätshandbücher. Soweit sinnvoll werden diese vielfach mit träger- oder einrichtungsspezifischen Standards hinterlegt bzw. im Falle der Expertenstandards mit entsprechenden Anpassungen der Expertenstandards an Einrichtungsspezifika. Daneben finden sich interne Audits, die oftmals den Charakter einer Vorbereitung auf eine MDK-Prüfung haben. Die QPR benennt ganz überwiegend Struktur- und Prozesskriterien. Auch jene Prüfkriterien, die als Ergebniskriterien ausgewiesen sind, erfüllen den Anspruch an ein Outcome (s. Kapitel 1.3) kaum, sondern spiegeln ein verkürztes Verständnis wider, das sich ebenfalls in den ‚Gemeinsamen Grundsätzen‘ findet (Stemmer & Arnold, 2012). Zwar werden Ereignis-

se wie Dekubiti, Fixierungen, Kontrakturen dokumentiert und summiert. Eine dezidierte Datenerhebung für eine Outcomemessung ist in deutschen Alten- und Pflegeheimen aktuell aber nur im Ausnahmefall gegeben.

Die Orientierung an der QPR stellt in gewissem Umfang jedoch faktisch schon eine Verbindung von internem Qualitätsmanagement und externer Qualitätsprüfung dar. Hier sollte im Sinne der oben skizzierten britischen und niederländischen Beispiele angeknüpft werden. Nicht zuletzt bedeutet dies ein entsprechendes Instrumentarium aufzubauen und den Einrichtungen zur Verfügung zu stellen.

Erhebung und Übermittlung der Daten

Gesetzt den Fall, ein entsprechendes methodisch abgesichertes Instrumentarium steht zur Verfügung und die Verzahnung von interner Qualitätsentwicklung und externer Qualitätsprüfung gelingt grundsätzlich, ist damit eine zuverlässige Datengrundlage noch nicht sichergestellt.

Outcomemessung stützt sich auf Daten, die von den Einrichtungen zur Verfügung gestellt werden. Auch wenn die verwendeten Instrumente eine akzeptable Reliabilität aufweisen, ist es dennoch möglich, dass die Datenerhebung vor Ort erhebliche Differenzen in der Datenqualität aufweist (Mor, 2005). Selbst ein Instrument wie das RAI, dessen Entwicklung bereits seit über 25 Jahren verfolgt wird, weist an manchen Stellen noch eine mangelnde Konsistenz und Eindeutigkeit auf (Dellefield, 2007). So ist beispielsweise für die Datenerhebung des Minimum Data Sets (MDS) ein zeitlicher Umfang von unter zwei bis über sechs Stunden nachgewiesen. Das vorgegebene Eingangsassessment benötigt im Mittelwert 4,2 Stunden und die vierteljährliche Erhebung im Mittel 1,8 Stunden. Diese große zeitliche Bandbreite bei der Erhebung der notwendigen Daten verweist auf eine große Variabilität des Verständnisses und der Umsetzung (Straker & Bailer, 2008).

Ein Beispiel für unterschiedliche Interpretationen im Rahmen der Erhebung von Qualitätsdaten stellt auch die erste Phase nach der Einführung der Pflegetransparenzvereinbarung dar. Hier zeigten sich deutliche Unterschiede bei der Beurteilung zwischen den vom rheinland-pfälzischen Medizinischen Dienst vergebenen Noten und dem bundesweiten Notendurchschnitt. Als Gründe für die Beurteilerdifferenzen wurden zum einen Verfahrensunsicherheiten und zum anderen mangelnde Eindeutigkeit und breiter Interpretationsraum der Ergebnisse der Qualitätsmessung identifiziert (Weidner, Laag, & Brühl, 2011).

Auch wenn z. B. Assessmentinstrumente mit ausreichender Reliabilität vorliegen, ist noch nicht gesichert, dass diese für den Einsatz durch alle Berufsgruppen geeignet sind. So erfolgt beispielsweise die Einschätzung des Ernährungszustandes in nahezu allen Fällen durch Pflegefachkräfte. Die bestehenden Screeninginstrumente wurden jedoch häufig nicht für die Anwendung in der Pflege entwickelt. Folglich entsteht ein erhöhter Bedarf an Fort- und Weiterbildungen und an multiprofessioneller Zusammenarbeit (Bartholomeyczik et al., 2010).

Dies knüpft an die Frage der Organisation des Vorgehens bei der Datenerhebung an. Dieses muss eindeutig definiert sein. Drei zentrale Fragen spielen hierbei eine Rolle. (1) Wie sollen die Daten erhoben werden? (2) Wer soll die Daten erheben? (3) Wie sollen die Daten dokumentiert werden? (Dellefield, 2007). Als hilfreich wird eine multiprofessionelle Kooperation erachtet. Die Daten einer Studie in Ohio zeigen auf, dass 97% der befragten Einrichtungen

bei der Datenerhebung für das Resident Assessment Instrument (RAI) interdisziplinär zusammenarbeiten (Straker & Bailer, 2008).

Jenseits der methodischen gibt es Herausforderungen auf der praktischen Ebene. Der zusätzliche Zeitaufwand, den die Pflegekräfte durch die Erhebung haben, stellt ein Problem dar (Dunckley et al., 2005). Auch mangelnde Kenntnis im Umgang mit dem Erhebungsinstrumentarium kann sich negativ auf die Datenqualität auswirken (Mor, 2005).

Zur Sicherstellung einer ausreichenden Datenqualität ist ein strukturiertes Vorgehen unerlässlich. Straker & Bailer (2008) zeigen ein dezidiertes Verfahren zur Sicherstellung der Datenqualität anhand des RAI/MDS auf, in dem Verantwortlichkeiten für die Datenerhebung ebenso festgelegt werden wie der Zeitpunkt der Datenerfassung und das Procedere der Datenübermittlung.

Auch eine leichte Handhabbarkeit des Verfahrens, eine nicht allzu häufige Datenerhebung und die Nutzung von Routinedaten begünstigen eine Implementierung. Der Nutzen der Erhebung sollte zudem den MitarbeiterInnen und den BewohnerInnen sichtbar gemacht und die Ergebnisse sollten für alle transparent dargestellt werden (Dunckley et al., 2005).

Ein weiterer fördernder Faktor für die Datenerhebung in ausreichender Qualität ist die Benennung eines verantwortlichen Koordinators bzw. einer verantwortlichen Koordinatorin vor Ort (Dellefield, 2007; Dunckley et al., 2005). Eine Erhebung in Ohio (n=125) zeigt auf, dass 90% der teilnehmenden Einrichtungen angeben, mindestens einen bzw. eine MDS-Assessment-KoordinatorIn mit dem Stellenumfang einer Vollzeitstelle zu haben. Hierbei sind die eingesetzten Zeitressourcen abhängig von der Größe der Einrichtung. So setzen 14 der 40 Einrichtungen mit mehr als 101 Bewohnerplätzen drei oder mehr Vollzeitkoordinatoren und –koordinatorinnen ein. Diese haben die Aufgabe, die RAI-Assessments zu vervollständigen und die Erhebung zu koordinieren und zu kontrollieren (Straker & Bailer, 2008).

Die Übermittlung der Daten muss elektronisch erfolgen. Es ist ein Dokumentationssystem verpflichtend einzuführen, das einerseits der Pflegedokumentation dient und zugleich der Datenerhebung für das Qualitätsmanagement und der externen Qualitätsprüfung und –berichterstattung. Das RAI ist hier beispielhaft einzuordnen (Mor, 2005).

Zum Standard gehört ein Verfahren zur Sicherung der Datenqualität auf Seiten der Auswertungsstelle. Dazu gehören Plausibilitätsprüfungen durch die Auswertungsstelle, ggf. Nachfrage und Korrektur sowie ein stichprobenartiger Abgleich der Daten vor Ort (AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH, 2013; Straker & Bailer, 2008).

4.2.3 Können mit den von Wingefeld et al. vorgeschlagenen Plausibilitätsprüfungen im Rahmen externer Qualitätsprüfungen die Daten über die Qualität der Pflege verifiziert werden

Externe Qualitätskontrollen finden derzeit mittels Vor-Ort-Begehung durch die Mitglieder des MDK statt. Mit dem Ziel die Prozessqualität zu überprüfen, werden die Pflegedokumentationen eingesehen, MitarbeiterInnen der Einrichtungen befragt und BewohnerInnen sowohl interviewt als auch in Augenschein genommen (MDS & GKV, 2014).

Diese Form der Überprüfung erlaubt einen Einblick in die Durchführung von Handlungsweisen, den aktuellen Gesundheitsstatus der ausgewählten BewohnerInnen und nicht zuletzt in die Sorgfalt, mit der die Dokumente geführt worden sind. Die Vor-Ort-Begehung gibt ebenfalls Aufschluss über die Atmosphäre einer Einrichtung. Prozessqualität thematisiert die Qualität, mit der eine Handlung durchgeführt wird. Zur Qualitätssicherung werden Standards entwickelt, die den idealen Ablauf dieses Handlungsprozesses beschreiben. Situationsbedingt kann von diesem Standard abgewichen werden. Dieses wiederum ist in der individuellen Pflegedokumentation ersichtlich. Im Rahmen der aktuellen Qualitätsprüfungen durch den MDK wird die Güte der aktuell beschriebenen Prozesse eruiert unter der Annahme, dass diese auch den regelhaft umgesetzten Prozessen entsprechen. Die Datenerhebung liegt bei diesem Verfahren – mit Ausnahme der Strukturdaten, die von den Einrichtungen zur Verfügung gestellt werden, – in den Händen der Prüfbehörden.

Mit dem Beginn der Outcomemessung kommt es zu gravierenden Veränderungen. Bei der Qualitätskontrolle mittels Outcomemessung geht die Aufgabe der Datenerhebung in die Hände der Einrichtungen über. Diese müssen nun für eine zuverlässige Datenerhebung und Datenübermittlung sorgen. Die Qualität der Rohdaten ist von allergrößter Bedeutung. Es muss das Ziel sein, zu möglichst gesicherten und plausiblen Angaben zu kommen. Im Krankenhausbereich gibt es seit 2005 auch in Deutschland bereits Erfahrungen mit ergebnisorientierter externer Qualitätsberichterstattung. Im Prozess der Sicherstellung einer ausreichenden Datenqualität kommt neben den Altenhilfeeinrichtungen der Auswertungsstelle (hier erscheint es sinnvoll, dass ein unabhängiges Institut diese Aufgabe übernimmt) eine zentrale Bedeutung zu. Die Aufgabe der Auswertungsstelle geht über die reine Durchführung von Rechenoperationen deutlich hinaus. Es ist ihre Aufgabe, den Einrichtungen vielfältige Unterstützung zu geben, damit diese in die Lage versetzt werden, korrektes Datenmaterial zu liefern. Dazu gehört die Federführung bei der Erstellung einer (durch eine Softwarefirma produzierten) geeigneten Übermittlungssoftware ebenso wie die Erstellung eines Manuals, in dem technische Anforderungen an Datenerfassung und –übermittlung, Schlüsseldefinitionen und Exportformate spezifiziert werden. Sowohl die Software als auch das Manual sind kontinuierlich anzupassen und fortzuentwickeln (AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH, 2013).

Diese Form der Unterstützung ist unerlässlich, die Erfahrung zeigt aber, dass damit die gewünschte Datenqualität meist noch nicht zufriedenstellend erreicht wird. Es sind zusätzliche Maßnahmen erforderlich. Sowohl die Organisation der Datenerhebung und –übermittlung zur Outcomemessung mittels des Minimum Data Sets (MDS) in den USA (Straker & Bailer, 2008) als auch das Messmanagement des AQUA-Institutes, das die Auswertung für die externe Qualitätsprüfung im Krankenhaus durchführt, arbeiten mit mehrfachen Plausibilitätsprü-

fungen und Rückkopplungsschleifen, bevor die Daten dann in den Instituten zur Auswertung kommen.

Erst wenn diese Auswertungen vorliegen, kommt die externe Qualitätskontrolle auf der inhaltlichen Ebene zum Zuge. Die externe Qualitätssicherung im Krankenhausbereich sieht bei auffälligen Daten, d.h. sowohl unbefriedigenden als auch besonders guten Ergebnisse, den sog. Strukturierten Dialog vor. In einem stufenweisen Vorgehen, das zeitnah nach dem Vorliegen der Ergebnisse eingeleitet wird, wird den Auffälligkeiten nachgegangen. Die erste Stufe besteht in einem Hinweis. Hierzu kommt es, wenn die Schwere der Abweichung als eher gering eingestuft wird und/ oder die Einrichtung ein erstes Mal auffällig geworden ist. Auf der zweiten Stufe wird von der Einrichtung eine Stellungnahme eingefordert. Wenn diese Stellungnahme schlüssig ist und akzeptiert wird, endet der Strukturierte Dialog. Andernfalls folgt der dritte Schritt, der in einer Vor-Ort-Begehung besteht und versucht, Ursachen für Qualitätsmängel zu klären. Angestrebt wird eine gemeinsam von EinrichtungsvertreterInnen und GutachterInnen getragene Einschätzung der bedingenden Zusammenhänge, die in einer schriftlichen Zielvereinbarung mündet. Nicht kooperierende Einrichtungen werden den Aufsichtsbehörden gemeldet (AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH, 2013).

Der Blick auf die Verfahrensweise im Krankenhaus verdeutlicht, dass Qualitätsprüfung auf der Grundlage von Outcomemessungen immer retrospektiv agiert.

Gesetzt den Fall, dass weiterhin – anders als im Bereich der Krankenhäuser – stichprobenartig und regelmäßig Einrichtungen der stationären Altenhilfe von GutachterInnen des MDK begangen werden, könnten diese nun auf die bereits vorliegenden Outcomedaten zurückgreifen und diese zum Ausgang für das Gespräch zwischen ihnen und den VertreterInnen der Einrichtung machen. Wie in Kapitel 4.1.2 ausgeführt, wird es notwendig sein, darüber hinaus themenbezogen Aspekte der Struktur- und Prozessqualität im Portfolio der Qualitätsmessung in der stationären Altenhilfe zu belassen.

4.2.4 Fazit

Die anvisierte dreifache Nutzung der Daten zur internen Qualitätsentwicklung, zur externen Qualitätskontrolle und zur Qualitätsberichterstattung hat Konsequenzen auf drei Ebenen.

Die Pflegedokumentation wird nun zum unmittelbaren Datenlieferanten für die Outcomemessung. Die aktuelle Vielfalt der Dokumentationssysteme und -verfahren lässt kaum eine Datenerhebung zu, die den Anforderungen an die Datenqualität für die Outcomemessung gerecht wird. Die Systematik der Pflegedokumentation ist äußerst heterogen und die Anschlussfähigkeit an ein Übermittlungssystem in keiner Weise gesichert. Wenn die aktuell breit gefächerten Dokumentationssysteme für die Datengrundlage genutzt werden sollen, hat dies umfangreiche Doppeldokumentationen verbunden mit hoher Fehleranfälligkeit zur Folge. In der Konsequenz ist eine edv-gestützte Pflegedokumentation verpflichtend einzuführen, die Items für die Datenerhebung in das Dokumentationssystem zu integrieren und die Schnittstelle für die Datenübermittlung zu definieren.

Wenn die dreifache Nutzung der Daten auch aus Sicht der Einrichtungen gewinnbringend gelingen soll, darf diese nicht nur die Aufgabe haben, für eine ausreichend Qualität der Rohdaten zu sorgen, sondern muss die Ergebnisse der Outcomemessung für die eigene Quali-

tätsentwicklung nutzen. Dazu sind die Auswertungsdaten den Einrichtungen zur Verfügung zu stellen. Darüber hinaus ist zu empfehlen, dass die Einrichtungen bei der Integration weitere Unterstützung erfahren. Dazu könnten die Benennung prioritärer Bereiche gehören, sowie das Angebot von Informationen, Strategien und Hilfen, um die interne Qualitätsentwicklung im Hinblick auf diese empfohlenen Prioritäten zu gestalten.

Auf der Ebene der externen Qualitätskontrolle kommt nun der Auswertungsstelle in enger Zusammenarbeit mit den Einrichtungen eine wichtige Funktion bei der Sicherung einer ausreichenden Qualität der Daten und damit der Verlässlichkeit der Ergebnisse der Outcomemessung zu.

Im Rahmen von Vor-Ort-Begehungen des MDKs können Outcomes nicht unmittelbar überprüft werden, aber sehr wohl im Sinne von Plausibilitätsprüfung zum Ausgangspunkt des Gespräches werden.

4.3 Bewertungssystematik

(Jens Arnold)

4.3.1 Stellen die von Wingenfeld et al. gegebenen Empfehlungen zur Bewertungssystematik eine Form der Bewertung von Indikatoren/Indikatorenbereichen dar, die eine Differenzierung von Einrichtungen ermöglicht und sind die vorgenommenen Setzungen ausreichend fachlich begründet?

Von Wingenfeld et al. (2011) werden sehr konkrete Vorschläge für eine Bewertungssystematik unterbreitet, mit deren Hilfe die von den Autoren zur standardisierten Dokumentation empfohlenen Ergebnisqualitätsindikatoren im Sinne eines Benchmarkings für Anbietervergleiche herangezogen werden können. Auf den ersten Blick wirkt die vorgelegte Systematik gut durchdacht und in sich schlüssig. Um die an die vorliegende Expertise gerichtete Frage aber zielgerichtet beantworten zu können, muss die vorgeschlagene Systematik differenziert betrachtet werden, da die Bewertungen teils indicatorspezifisch erfolgen und zudem auf mehreren Betrachtungsebenen angesiedelt sind.

Wingenfeld et al. (2011) haben einerseits eine Bewertungssystematik für die Einrichtungsbeurteilung in Bezug auf die 15 von ihnen definierten Einzelindikatoren entwickelt, darauf aufbauend aber zusätzlich auch noch eine Systematik für eine übergreifende Bewertung der drei aggregierten Qualitätsbereiche vorgelegt. Dabei werden dann die fünf Indikatoren des Qualitätsbereichs 1, die sechs des Qualitätsbereichs 2 sowie die vier des Qualitätsbereichs 3 jeweils zusammenfassend bzw. aggregiert betrachtet. Bei der Diskussion der Bewertungssystematik sowie der vorgenommenen Setzungen müssen beide Bewertungssystematiken, also sowohl die Indikatorbezogene als auch die Indikatorübergreifende, im Hinblick auf ihre fachliche Fundierung sowie Differenzierungsfähigkeit von Einrichtungsdaten getrennt voneinander betrachtet werden. Es ist ferner notwendig, sich einen Überblick über die Einzelheiten der Vorgehensweisen, die eine Bewertung von Einrichtungen ermöglichen sollen, zu verschaffen. Abschließend wird ein Gesamtfazit gezogen.

Indikatorbezogene Bewertungssystematik

Einen zusammenfassenden Überblick über die zentralen Elemente der Wingenfeld-Bewertungssystematik auf Einzelindikatorebene gibt Tabelle 6. Verbindendes Element der indikatorenbezogenen Bewertungssystematik ist das einheitliche Bewertungsraaster, das die Einrichtungen in drei Klassen nach überdurchschnittlichem, durchschnittlichem und unterdurchschnittlichem Abschneiden innerhalb der jeweiligen Indikatoren einteilt. Die Frage nach der Sinnhaftigkeit dieser Dreiteilung sei an dieser Stelle noch zurückgestellt. Eine zentrale methodologische Frage ist zunächst, wie genau diese Einteilung zustande kommt. Insgesamt kommen hier drei unterschiedliche Verfahren zum Einsatz, anhand derer die beiden Trennwerte bestimmt werden, die notwendig sind, um Einrichtungen mit maximal möglichen Indikatorausprägungen zwischen 0 und 100 Prozent, einem von drei Qualitätsniveaus, zuweisen zu können:

- Bei den Indikatoren des Qualitätsbereichs 1 erfolgt die Trennung einheitlich anhand des oberen und unteren Quartils („**Qualitätszuweisungsmethode**“ (**QZM**) 1). Dabei werden anhand der Verteilung der vorliegenden empirischen Daten die 25 % der Einrichtungen mit dem besten Abschneiden und die 25 % der Einrichtungen mit dem vergleichsweise schlechtesten Abschneiden identifiziert. Im Fall des Qualitätsbereichs 1 sind das die Einrichtungen mit den höchsten und niedrigsten ermittelten Anteilen mit Erhalt bzw. Verbesserung auf den Indikatoren Mobilität, Alltagsverrichtungen oder Gestaltung des Alltagslebens und sozialer Kontakte. Da sich hier bei jedem Indikator anhand der in den erhobenen Daten konkret zu beobachtenden Verteilungen orientiert wird, ergeben sich jeweils unterschiedliche Schwellenwerte, anhand derer die konkrete Trennung zwischen den drei Qualitätsklassen vollzogen wird (s. Tabelle 6). Beim Indikator „Erhalt bzw. Verbesserung der Mobilität bei Menschen mit hohen kognitiven Einbußen“ werden Einrichtungen mit einem Anteil von mehr als 53,8 % an Bewohnern und Bewohnerinnen, bei denen dies erreicht wurde, als überdurchschnittlich bewertet. Ist der Anteil kleiner als 26,2 %, wird der Anbieter als unterdurchschnittlich bewertet, zwischen 26,2 % und 53,8 % als durchschnittlich. Bei geringerem Risiko verschieben sich diese Werte in der Regel deutlich nach oben, da es hier sowohl qualitativ guten wie schlechten Anbietern gleichermaßen in höherem Umfang gelingt, das Outcome zu erreichen. Im Falle der Mobilität wären dies dann etwa mindestens 80,0 % Kriterienerfüllung für das Prädikat „überdurchschnittlich“ und kleiner 50,1 % für eine Klassifikation als unterdurchschnittlicher Anbieter (vgl. Tabelle 6). Der Abstand zwischen den beiden Schwellenwerten wird bei QZM 1 als Interquartilsabstand (IQR) bezeichnet und gibt Aufschluss über die statistische Diskriminierungsfähigkeit des jeweiligen Indikators (siehe Abschnitt „Differenzierung von Einrichtungen“ weiter unten).
- Im Qualitätsbereich 2 erfolgt die Trennung der Qualitätsklassen mit einer geringfügig abweichenden Methode („**Qualitätszuweisungsmethode**“ (**QZM**) 2). Bei den sechs hier definierten Indikatoren wird analog zu QZM 1 ebenfalls auf die konkrete Verteilung bzw. „Norm“ der Indikatorausprägungen in der Untersuchungsstichprobe zurückgegriffen. Die Trennwerte werden hier allerdings nicht auf rangbasierten Quantilen, sondern unter Bezugnahme auf Mittelwertstatistiken definiert, genauer als das 0,5 bzw. 1,5fache der durchschnittlichen Auftretenshäufigkeit des Indikators in der Gesamtstichprobe aller Anbieter. Da gemäß der Indikatoroperationalisierungen des Qualitätsbereichs 2 eine positive bzw. hohe Anbieterqualität durch Absenz von schädigenden Einflüssen und Belastungen definiert ist, wird eine Einrichtung hier als überdurchschnittlich bewertet, wenn ihr Anteilswert unterhalb des 0,5fachen Gesamtdurchschnittswerts liegt. Wenn er über dem 1,5fachen des Gesamtdurchschnitts liegt, wird das Qualitätsniveau als unterdurchschnittlich eingestuft.

Tabelle 6: Übersicht Bewertungssystematik bei den Wingefeld-Indikatoren

Qualitätsbereich	Indikator	Bewertungsraster Einrichtungen	Trennwerte für die Klassen	Empirische Schwellenwerte
Bereich 1: Erhaltung und Förderung von Selbständigkeit	Mobilität (geringe kogn. Einbußen)	3 Klassen - überdurchschnittlich - durchschnittlich - unterdurchschnittlich <i>Positive Bewertung bezogen auf Anteile der Einrichtungen mit Erhalt bzw. Verbesserung auf dem Indikator</i>	oberes Quartil unteres Quartil	> 80,0% IQR: < 50,1% 29,9%
	Mobilität (hohe kogn. Einbußen)		oberes Quartil unteres Quartil	> 53,8% IQR: < 26,2% 27,6%
	Alltagsverrichtungen (geringe kogn. Einbußen)		oberes Quartil unteres Quartil	> 93,3% IQR: < 68,5% 24,8%
	Alltagsverrichtungen (hohe kogn. Einbußen)		oberes Quartil unteres Quartil	> 66,7% IQR: < 43,0% 23,7%
	Gestaltung des Alltagslebens und sozialer Kontakte		oberes Quartil unteres Quartil	> 75,0% IQR: < 50,1% 24,9%
Bereich 2: Schutz vor gesundheitlichen Schädigungen und Belastungen	Dekubitusentstehung (niedriges Risiko)	3 Klassen - überdurchschnittlich - durchschnittlich - unterdurchschnittlich <i>Positive Bewertung bezogen auf geringeren Anteil in Bezug zum durchschnittlichen Auftreten in der Gesamtstichprobe</i>	0,5facher 1,5facher ... Durchschnittswert	< 1,1% Diff.: > 3,2% 2,1%
	Dekubitusentstehung (hohes Risiko)		0,5facher 1,5facher ... Durchschnittswert	< 4,5% Diff.: > 13,4% 8,9%
	Stürze mit gravierenden Folgen (geringe kogn. Einbußen)		0,5facher 1,5facher ... Durchschnittswert	< 4,1% Diff.: > 12,3% 8,2%
	Stürze mit gravierenden Folgen (hohe kogn. Einbußen)		0,5facher 1,5facher ... Durchschnittswert	< 6,5% Diff.: > 19,4% 12,9%
	Unbeabsichtigter Gewichtsverlust (geringe kogn. Einbußen)		0,5facher 1,5facher ... Durchschnittswert	< 1,9% Diff.: > 5,6% 3,7%
	Unbeabsichtigter Gewichtsverlust (hohe kogn. Einbußen)		0,5facher 1,5facher ... Durchschnittswert	< 3,8% Diff.: > 11,4% 7,6%
Bereich 3: Unterstützung bei spezifischen Bedarfslagen	Integrationsgespräch nach dem Heimeinzug	3 Klassen - überdurchschnittlich - durchschnittlich - unterdurchschnittlich	> 75% < 50% ... der Bewohner	keine, kriterienbasierte Vorgaben der Autoren
	Einsatz von Gurtfixierungen		0,5facher 1,5facher ... Durchschnittswert analog Bereich 2	< 2,3% Diff.: > 7,9% 5,6%
	Einschätzung von Verhaltensauffälligkeiten bei Bewohnern mit kognitiven Einbußen		> 75% < 50% ... der Bewohner	keine, kriterienbasierte Vorgaben der Autoren
	Schmerzmanagement		> 90% < 70% ... der Bewohner	keine, kriterienbasierte Vorgaben der Autoren

Anmerkungen: IQR = Interquartilsabstand

- Die Bewertungssystematik im Qualitätsbereich 3 ist heterogen. Beim Indikator kommt ebenfalls die durchschnittswertbasierte QZM 2 zum Einsatz. Die Trennwerte aller anderen Indikatoren basieren dahingegen nicht auf den empirischen Verteilungsmustern in den erhobenen Daten, sondern auf absoluten Setzungen der Autoren, die aus Leitlinien abgeleitet werden bzw. als, aus Expertenurteilen abgeleitete, sog. Kriteriumsstandards gelten („**Qualitätszuweisungsmethode“ (QZM) 3**). Als Begründung für diese Vorgehensweise wird von Wingenfeld et al. (2011) darauf verwiesen, dass bei diesen Indikatoren ein Großteil der Einrichtungen die entsprechenden Anforderungen nicht erfüllt. Zum Beispiel sei es allgemein noch keine gängige Praxis, Integrationsgespräche nach dem Heimeinzug durchzuführen. Als weitere Gründe werden angeführt, dass notwendige Einschätzungen nicht in ausreichendem Umfang vorgenommen werden (Indikator Verhaltensauffälligkeiten) oder wie etwa beim Schmerzmanagement ein derart sensibler Qualitätsindikatorbereich vorläge, dass besonders hohe Qualitätsanforderungen sicherzustellen seien.

Wie sind die einzelnen Vorgehensweisen nun zu bewerten? Bei eingehender Betrachtung ergeben sich schnell einige kritische Fragen im Hinblick auf die vorgenommenen methodischen Setzungen, zu denen wir Optimierungsvorschläge unterbreiten möchten:

- Es leuchtet zunächst nur bedingt ein, warum Wingenfeld et al. (2011) überhaupt zwischen den Qualitätszuweisungsmethoden 1 und 2 unterscheiden. Aus Sicht der Gutachter stellt die im Rahmen von QZM 1 beschriebene **Klassifikationsmethode auf der Grundlage von rangbasierten Quartilen** (wie im vorliegenden Fall) bzw. Quantilen (im verallgemeinerten Fall) eine in der Statistik übliche und bewährte Technik dar. Insbesondere im Qualitätsbereich 2 werden davon abweichend die Schwellenwerte mit Abweichungen von den Mittelwerten der Gesamtstichprobe (0,5facher bzw. 1,5facher Abstand zum Durchschnittswert der Gesamtstichprobe) definiert. Rein rechnerisch gesehen, hätten aus den vorliegenden Daten auch hier problemlos Quartile ermittelbar sein müssen. Ein zwingender inhaltlicher Grund, warum hier darauf verzichtet wurde, erschließt sich den Gutachtern an dieser Stelle nicht. Allein der Verweis darauf, dass aufgrund der spezifischen Indikatordefinition in diesem Qualitätsbereich eine hohe Qualität als Nicht-Auftreten von unerwünschten Ereignissen wie Stürzen o. Ä. operationalisiert ist, rechtfertigt dies in keiner Weise. Etwa unterdurchschnittliche Qualität als die Zugehörigkeit zu den 25 % der Einrichtungen zu definieren, bei denen das negative Indikatorereignis am vergleichsweise häufigsten auftritt, wäre in der Rezeption durch statistisch eher unbewanderte Anwender mindestens genauso intuitiv nachvollziehbar gewesen. Unsere Empfehlung lautet daher, bei allen Bewertungsklassifikationen, die eine Einordnung in eine Referenzstichprobe beinhalten sollen, stringent auf einen rangbasierten Ansatz, etwa analog zu den im Qualitätsbereich 1 eingesetzten Quartilen, zurückzugreifen.

Welche Gründe sprechen überhaupt für die Verwendung rangbasierter Klassifikationen? Das Hauptargument liegt sicher darin, dass Mittelwertstatistiken, die auf metrisch skalierten Daten basieren, sehr anfällig für Ausreißer sind (Stengel, Bhandari, & Hanson, 2011). D.h. einzelne Einrichtungen mit extremen Indikatorausprägungen können schnell die ganze Bewertungssystematik verzerren. Rangbasierte Verteilungsparameter wie etwa der Median (der übrigens gleichsam auch das zweite Quartil darstellt) sind hier deutlich

robuster und daher weniger fehleranfällig. Ein weiteres Argument ist, dass entsprechende Auswertungen mit Hilfe der sehr bewährten Boxplots (Tukey, 1977) sehr leicht auch anschaulich visualisiert werden können. Boxplots eignen sich ferner hervorragend dazu, die besagten Ausreißer in der Qualitätsberichterstattung überhaupt erst zu identifizieren. Als Ausreißer werden hier alle Werte bzw. Einrichtungen definiert, die außerhalb des 1,5fachen („milde Ausreißer“) oder 3fachen Interquartilsabstands liegen („extreme Ausreißer“).

Letztendlich gibt es in der Statistik keine allgemeingültigen Vorschriften oder Patentrezepte wie eine Klassifikation idealiter vorzunehmen ist. Dies ist immer vor dem Hintergrund der vorliegenden Daten abzuwägen. Insofern sind subjektive Ermessensspielräume hier unvermeidlich. Nichtsdestotrotz gibt es gebräuchlichere und weniger gebräuchliche Ansätze. Für Schwellenwertintervalle rund um Mittelwerte gibt es beispielsweise keine gängigen Konventionen. Eine Setzung als 0,5facher bzw. 1,5facher Abstand zum Mittelwert weist aufgrund ihrer stärkeren Abhängigkeit von der konkreten Verteilungsform einen höheren Grad an Willkürlichkeit als rangbasierte Klassifikationen auf. Neben der Einheitlichkeit des Vorgehens spricht auch dies wiederum für die letztgenannte Methode.

- Auf die kriteriums-basierte Qualitätszuweisungsmethode QZM 3 und die globale Abgrenzung zu den eng miteinander „verwandten“ norm- bzw. referenzstichprobenorientierten QZM 1 und QZM 2 wird in Kapitel 4.3.2 noch ausführlicher eingegangen.
- Im Rahmen der Darstellung der Bewertungssystematik im Qualitätsbereich 2 beschreiben Wingenfeld et al. (2011) einen zusätzlichen „Schutzmechanismus“, der verhindern soll, dass kleine Einrichtungen negative Bewertungen aufgrund von Zufallsschwankungen erhalten: Ist ein Anteilswert lediglich durch einen Einzelfall zustande gekommen, kann die Einrichtung nicht als unterdurchschnittlich bewertet werden. Prinzipiell ist es sehr löblich, solche Überlegungen in die Bewertungssystematik einfließen zu lassen. Die Umsetzung erscheint hier allerdings eher als verkomplizierendes Element. In der im Kontext der Erstellung der vorliegenden Expertise recherchierten Literatur wurden derartige Probleme normalerweise über die Definition von Mindeststichprobengrößen in Angriff genommen. Dabei werden Indikatoren ab dem Unterschreiten einer bestimmten Bewohnerzahl nicht mehr berechnet (vgl. Ausführungen zur Risikoadjustierung in Kapitel 4.4.1).

Übergreifende Bewertungssystematik der Qualitätsbereiche

Neben den indikatorspezifischen Qualitätsbewertungsvorschlägen beschreiben Wingenfeld et al. (2011) noch eine zusätzliche Bewertungssystematik, die es auf Ebene der drei von den Autoren dargestellten Qualitätsbereichen ermöglichen soll, Qualitätsvergleiche zwischen Anbietern auch auf indikatorübergreifender Ebene vorzunehmen. Einen Überblick über diese Bewertungssystematik gibt Tabelle 7.

Die übergreifende Bewertungssystematik knüpft an die indikatorspezifischen Bewertungen an. Dabei wird aus dem Abschneiden auf den Indikatoren des jeweiligen Qualitätsbereichs ein Gesamtpunktwert ermittelt, wobei ein überdurchschnittliches Abschneiden bei der Summenbildung mit einem Pluspunkt und ein unterdurchschnittliches Abschneiden mit einem Minuspunkt gerechnet werden. Ein durchschnittliches Abschneiden auf einem Indikator wird mit null Punkten gewertet. Die Spannweite der Gesamtpunktwerte richtet sich demnach nach der Anzahl der jeweils definierten Indikatoren. Da dem Qualitätsbereich 2 beispielsweise sechs

Indikatoren zugewiesen wurden, können die Gesamtpunktwerte hier von -6 bis +6 reichen. Um eine Vergleichsbasis für Einrichtungen generieren zu können, bei denen nicht alle Einzelindikatoren vorliegen, können die Qualitätsniveaus darüber hinaus noch in Form von Mittelwerten dargestellt werden (s. Tabelle 7). Dies erscheint grundsätzlich sinnvoll, damit an dieser Stelle nicht zu viele Anbieter sozusagen aus formalen Gründen vom Vergleich ausgeschlossen werden.

Tabelle 7: Übergeordnete Bewertungssystematik der Wingenfeld-Qualitätsbereiche

Qualitätsbereich	Bewertungsraster (4 Klassen)	Zählweise der Ergebnisse Einzelindikatoren	Gesamtpunktwert	Mittlerer Punktwert
Bereich 1: Erhaltung und Förderung von Selbständigkeit	herausragend	+1 wenn überdurchschnittl. 0 wenn durchschnittl. -1 wenn unterdurchschnittl.	3 bis 5	0,6 bis 1,0
	überdurchschnittlich		2	0,3 bis 0,5
	durchschnittlich		-1 bis 1	-0,2 bis 0,2
	unterdurchschnittlich		-2 bis -5	-1,0 bis -0,3
Bereich 2: Schutz vor gesundheitlichen Schädigungen und Belastungen	herausragend	+1 wenn überdurchschnittl. 0 wenn durchschnittl. -1 wenn unterdurchschnittl.	5 bis 6	0,8 bis 1,0
	überdurchschnittlich		3 bis 4	0,5 bis 0,7
	durchschnittlich		0 bis 2	0,0 bis 0,4
	unterdurchschnittlich		-6 bis -1	-0,1 bis -1,0
Bereich 3: Unterstützung bei spezifischen Bedarfslagen	Ziele vollständig erreicht	+1 wenn überdurchschnittl. 0 wenn durchschnittl. -1 wenn unterdurchschnittl.	4	1,0
	Ziele überwiegend erreicht		2 bis 3	0,5 bis 0,9
	Ziele teilweise erreicht		-1 bis 1	0,4 bis -0,4
	Ziele nicht erreicht		-2 bis -4	-0,5 bis -1,0

Nichtsdestotrotz ist aus Sicht der Gutachter die übergreifende Bewertungssystematik in einigen zentralen Aspekten kritisch zu bewerten:

- Im Vergleich zur Bewertung der Einzelindikatoren fällt zunächst unmittelbar ins Auge, dass hier eine **abweichende Klassifikation der Qualitätsniveaus** verwendet wird (s. Tabelle 7). Werden auf Indikatorebene die Einrichtungen in drei Stufen als über-, unter- und durchschnittlich bewertet, so kommt in der übergreifenden Systematik mit „herausragend“ noch eine weitere Stufe hinzu. Zwar begrüßen die Autoren der vorliegenden Expertise grundsätzlich eine differenziertere Abstufung der Klassifikationen (s. auch Kapitel „Differenzierung von Einrichtungen“), ein zwingender Grund für eine Erweiterung der Abstufung wird an dieser Stelle aber nicht dargelegt. Insgesamt erscheint die Bewertungssystematik so etwas instingent. Diskussionswürdig ist in diesem Zusammenhang ebenfalls, warum keine symmetrische Stufung verwendet wurde, die dann auch den weit unterdurchschnittlichen Bereich miteinschließt. Auf diese Weise könnte in der Außenwahrnehmung der fälschliche Eindruck entstehen, dass das Bewertungssystem präferiert positive Qualität darstellen (und ggf. negative Qualität eher kaschieren) möchte.
- Bei Gegenüberstellung der in Tabelle 7 für die einzelnen Qualitätsniveaus ausgewiesenen mittleren Punktwerte fällt auf, dass diese sich von Qualitätsbereich zu Qualitätsbereich unterscheiden. D.h. je nach Qualitätsbereich ist ein verhältnismäßig anderer Schnitt

von guten Einzelindikatorbewertungen notwendig, um auch insgesamt als „überdurchschnittlicher“ Anbieter eingestuft zu werden. Dies gilt vice versa natürlich analog auch im negativen Qualitätsbereich. Aus den im Abschlussbericht von Wingenfeld et al. (2011) ausgewiesenen Ergebnissen kann geschlossen werden, dass mit Hilfe der **unterschiedlichen Bewertungsmaßstäbe** versucht wurde, die Gesamtbewertung der Einrichtungen annähernd normalverteilt zu gestalten, wobei die meisten Anbieter dann in den durchschnittlichen Bereich fallen (vgl. Wingenfeld et al., 2011, 106). Da es sich hier um den Versuch handelt, die Bewertung an den in der Referenzstichprobe empirisch ermittelten Sachverhalten auszurichten, können hier gleichsam die Argumente für und wider einer bezugsnorm- oder kriterienbezogenen Konzeption der Bewertungssystematik angelegt werden (vgl. Kapitel 4.3.2). Wie in dem entsprechenden Abschnitt deutlich wird, favorisieren wir insgesamt eher einen kriterienbezogenen Ansatz. Dabei würden vergleichbare Punktmittelwerte in allen Qualitätsbereichen gleiche Bewertungen nach sich ziehen. Aber auch davon losgelöst muss hinterfragt werden, ob es zwingend notwendig ist, dass Daten, die bereits auf Einzelindikatorebene „normiert“ bewertet wurden, auf der übergreifenden Ebene dann nochmals normiert werden müssen. Diese Frage ist auch im Hinblick auf die Transparenz und Nachvollziehbarkeit der Bewertungen relevant. Ferner ist zu berücksichtigen, dass eine zu große Zahl von durchschnittlichen Einrichtungen insgesamt wenig Anreize für die Initiierung von Qualitätsentwicklungsprozessen setzt.

- Eine elementarere Fragestellung tangiert die **Sinnhaftigkeit einer übergreifenden bzw. summatorischen Betrachtung von Indikatoren** als solche: Damit die Ergebnisse von Einzelindikatoren in einen Gesamtwert verrechnet werden dürfen, ist es notwendig, dass sich diese auf zumindest ähnliche Sachverhalte beziehen. In inhaltlicher Hinsicht ist die von Wingenfeld et al. (2011) vorgenommene Zuordnung der Indikatoren in die drei Qualitätsbereiche zwar durchaus augenscheinlich, die Ähnlichkeit sollte sich aber idealerweise auch auf empirischer Ebene in Form von signifikanten Korrelationen zeigen (Mor, 2006; Weibler-Villalobos & Röhrig, 2010). Konkrete Hinweise auf die Zusammenhänge zwischen den Wingenfeld-Indikatoren liegen leider nicht vor. Entsprechend kann die Gültigkeit bzw. faktorielle Validität der Dimensionalität der vorgenommenen Zuordnungen testtheoretisch nicht verifiziert werden. Einige der Zuweisungen werden durch die vorliegenden empirischen Befunde gestützt. So zeigen etwa Bates-Jensen und MacLean (2007) sowie Castle und Ferguson (2010), dass Dekubitusentstehung oftmals mit Mangelernährung bzw. dem Risiko für unbeabsichtigten Gewichtsverlust in Zusammenhang steht. Mor, Angelelli, Gifford, Morris und Moore (2003) verweisen aber darauf, dass viele Studien oft nur eine geringe Korrelation zwischen den verschiedenen Qualitätsmaßen nachweisen. Anbieter, die in einem Bereich eine Kumulation von unerwünschten Outcomes haben, könnten in anderen Bereichen durchaus veritabel abschneiden. Dies widerspreche der intuitiv-heuristischen Annahme, dass gute Pflege auch durchgängig gute Ergebnisse nach sich ziehen müsse.

In der bilanzierenden Einschätzung der dargelegten Diskussionspunkte und offenen Fragen möchten wir vom Einsatz der übergreifenden Bewertungssystematik insgesamt abraten. Dies sollte vor dem Hintergrund der Wingenfeld-Indikatoren insbesondere so lange gelten, wie die faktorielle Validität der drei Qualitätsbereiche und ihrer Indikatorzuordnungen nicht schlüssig belegt ist. Auch wenn die Zuordnungen in inhaltlich-thematischer Hinsicht plausibel erschei-

nen, wäre es sonst durchaus denkbar, dass hier die sprichwörtlichen Äpfel und Birnen zu einem Fruchtquark im wahrsten Sinne des Wortes vermengt werden.

Auch losgelöst von der von Wingenfeld et al. (2011) vorgelegten Bewertungssystematik sehen wir im Kontext der Outcomebestimmung in der stationären Pflege die Verwendung von indikatorübergreifenden Gesamtscores im Rahmen des Qualitätsbenchmarks von Einrichtungen als kritisch an. Diese Einschätzung liegt im Wesentlichen darin begründet, dass zur Verrechnung der Indikatoren nicht nur relevant ist, dass die Indikatoren in irgendeiner Form zusammenhängen und somit eine inhaltlichkohärente übergeordnete Dimension beschreiben, sondern sich darüber hinaus auch noch die Frage der Gewichtung der Indikatoren stellt, die aufgrund des Fehlens geeigneter Maßstäbe in der Regel im Dunklen bleibt bzw. nicht trivial beantwortet werden kann (vgl. Wingenfeld et al., 2011, 104). Wie auch bei der Wingenfeld-Bewertungssystematik hat daher in Ermangelung an Alternativen eine Gesamtbewertung in der Regel einen additiven, gleichgewichteten und kompensatorischen Charakter. Dies bedeutet einerseits, dass in der Verrechnung allen Indikatoren die gleiche Relevanz zukommt, im Qualitätsbereich 3 also beispielsweise Integrationsgespräche die gleiche Bedeutung haben wie der Einsatz von Gurtfixierungen oder das Schmerzmanagement. Könnte man sich damit im Wingenfeld-Beispiel bei den Indikatoren der Qualitätsbereiche 1 und 2, die in fachlicher Sicht in höherem Maße gleichrangig erscheinen, noch einigermaßen anfreunden, treibt die kompensatorische Logik die nächsten Sorgenfalten auf die Stirn. Diese besagt nichts anderes, als dass schlechtes Abschneiden auf einer Bewertungsdimension durch ein besseres Abschneiden auf einer anderen aufgefangen werden kann. Auch daraus lassen sich interessante Beispiele konstruieren: Soll es möglich sein, dass ein Anbieter mit unterdurchschnittlicher Qualität bei der Entstehung gravierender gesundheitsschädigender Dekubitusgeschwüre dies durch ein überdurchschnittliches Abschneiden bei der Vermeidung von unbeabsichtigtem Gewichtsverlust kompensieren kann? Oder am Ende gar durch den verstärkten Einsatz von Gurtfixierungen, mit niedrigeren Sturzraten ausgestattet, dann in der Endbewertung ggf. als „überdurchschnittlich“ dastehen dürfen? Diese Beispiele sind natürlich bewusst provokant formuliert, um die Fallstricke einer Gesamtbewertung von Indikatoren deutlich zu machen. Wie kann aber alternativ vorgegangen werden, wenn unseren Empfehlungen folgend auf übergreifende Bewertungen verzichtet und nur Einzelindikatoren in die Qualitätsberichterstattung einfließen?

Insbesondere im Hinblick auf die Verbraucherperspektive (vgl. Kapitel 4.5) sollte bedacht werden, dass die Anzahl der für den Einrichtungsvergleich bereitstehenden Indikatoren in einem überschaubaren Rahmen bleibt. Mehr als ca. drei bis maximal sechs Qualitätsindikatoren zu veröffentlichen würde hier wahrscheinlich sehr schnell zu einer Überfrachtung und Überforderung führen. Für die Verbraucherseite wäre es daher notwendig, eine Auswahl zentraler bzw. besonders relevanter Qualitätsindikatoren zu treffen. Ein Beispiel einer aus unserer Sicht gelungenen Umsetzung dieses Zugangs ist der „Quality Compass“ in Ontario, Kanada (Health Quality Ontario, 2014c). Für die Stakeholder auf professioneller Ebene sollten natürlich alle fachlich relevanten Benchmarks in möglichst umfassender Breite abgebildet werden.

Zwar erfüllen natürlich auch die drei Qualitätsbereiche prinzipiell die Anforderung einer komprimierten Verbraucherinformation, eine wichtige Frage in diesem Zusammenhang ist aber auch, ob sich die VerbraucherInnen ohne umschweifige Erläuterungen unter weitläufigen Begriffen wie „Erhalt und Förderung von Selbständigkeit“, „Schutz vor gesundheitlichen

Schädigungen und Belastungen“ oder insbesondere unter „Unterstützung bei spezifischen Bedarfslagen“ wirklich intuitiv erschließen können, was hier genau gemeint ist bzw. welche Aspekte hier nicht eingeschlossen sind.

Man könnte in diesem Zusammenhang ebenfalls argumentieren, dass für VerbraucherInnen speziell eine einzelne übergreifende Gesamtbewertung, im Sinne etwa des Gesamturteils der Stiftung Warentest, hoch attraktiv sein könnte. Wenn sich dies unter fachlichen Gesichtspunkten allerdings nicht in valider Weise realisieren lässt, hat ein solcher Wert keine Aussagekraft und würde dem Verbraucher bzw. der Verbraucherin letztendlich allenfalls den Anschein von Nützlichkeit und Bedeutung vortäuschen. Eine aus Sicht der Autoren der vorliegenden Expertise überaus spannende Idee wurde in diesem Zusammenhang bei der „Minnesota Nursing Home report card“ realisiert (siehe „<http://nhreportcard.dhs.mn.gov>“). Die VerbraucherInnen können bei der Suche auf der Webseite selber Prioritäten setzen, welche der dort veröffentlichten Qualitätsdimension für sie persönlich besonders relevant sind. Diese subjektive Gewichtung könnte natürlich problemlos auch für die Bildung eines individuellen Gesamtscores verwendet werden. Das Problem der allgemeingültig nur schwierig zu umreisenden Bewertungsmaßstäbe wird hier also den individuellen Präferenzen der jeweiligen Adressatengruppen überlassen. Das ist ein sehr eleganter und „kundenorientierter“ Weg.

Differenzierung von Einrichtungen

Die zentrale Frage, ob durch die vorgelegte Bewertungssystematik eine hinreichende Differenzierung von Einrichtungen ermöglicht wird, muss von zwei Seiten beleuchtet werden: Global betrachtet mit der Frage nach der Zweckdienlichkeit des vorgeschlagenen Bewertungsrasters und im Detail im Hinblick auf die Differenzierungsfähigkeit der definierten Qualitätsindikatoren auf Grundlage ihrer statistisch-empirischen Eigenschaften.

Als zentrales Bewertungsraster für die Einrichtungen schlagen Wingefeld et al. (2011) drei Stufen vor: Entsprechend werden die Anbieter als überdurchschnittlich, durchschnittlich und unterdurchschnittlich klassifiziert. Auf der indikatorübergreifenden Ebene kommt mit dem Qualitätsniveau „herausragend“ noch eine weitere Stufe hinzu. Da diese asymmetrische Art der Klassifizierung schon im vorausgegangenen Abschnitt kritisch betrachtet wurde und zudem aus Sicht der Gutachter die Validität von Gesamtbewertungen grundsätzlich in Frage zu stellen ist, möchten wir uns an dieser Stelle auf das dreistufige Bewertungssystem fokussieren.

Insgesamt erscheint das vorgelegte Bewertungsraster für eine angemessene Differenzierung von Einrichtungen bzw. Anbietern als zu grob. Bei der quartilsbasierten Zuweisung der Einrichtungen in die drei Klassen entfallen 50 % der Einrichtungen in die Durchschnittskategorie. Je 25 % werden als über- oder unterdurchschnittlich eingestuft (dies gilt für QZM 1, die Verteilungen dürften bei QZM 2 aber sehr wahrscheinlich ähnlich aussehen). Die Wahrscheinlichkeit am Ende im Durchschnittsbereich zu liegen, ist statistisch gesehen also am größten. Wenn sich zu viele Anbieter im statistischen „Niemandland“ bzw. qualitativen Indifferenzbereich tummeln, ist dies weder für VerbraucherInnen noch die professionelle Ebene besonders informativ, um Unterschiede zwischen den Einrichtungen erkennen zu können, noch dürfte es den Anbietern selbst einen großen Anreiz geben, die Qualität ihrer Leistungen zu verbessern. Das Argument von Wingefeld et al. (2011), dass der Anreiz zu qualitativen Verbesserungen bei durchschnittlichem Abschneiden besonderes hoch sei, kön-

nen wir so nicht teilen. So finden auch Grando, Rantz und Maas (2007) in ihrer Studie, dass Pflegeheime sogar rein von den absoluten Zahlen her betrachtet eher kritische Outcome-Ergebnisse als völlig unproblematisch empfunden haben, wenn diese sich im Rahmen des nationalen Durchschnitts bewegt haben. Aus unserer Sicht braucht es daher eine differenziertere Bewertungssystematik, um auch feinstufigere Ziele setzen zu können. Auch Weibler-Villalobos und Röhrig (2010) fordern pro Indikator mindestens vier Bewertungsstufen, um Einrichtungen, welche die Qualitätsanforderungen kaum erfüllen können, von denen unterscheiden zu können, die sie fast erreichen. Zur Optimierung der Bewertungssystematik von Wingenfeld et al. (2011) könnten auf methodischer Ebene folgende Ansätze in Betracht gezogen werden:

- Reduzierung der Anzahl der Einrichtungen, die in die Durchschnittskategorie fallen. Dies könnte in der bestehenden Systematik schon dadurch erreicht werden, indem statt Quartile beispielsweise Terzile zur Bestimmung der Trennwerte für die drei Qualitätsniveaus Verwendung fänden. Dann würden je ein Drittel der Einrichtungen als über-, unter- oder durchschnittlich eingestuft. Das Grundproblem der zu groben Rasterung bliebe dann aber weiterhin bestehen. Zum Beispiel würden hier vermehrt auch faktisch eher mittelmäßige Anbieter in den überdurchschnittlichen Bereich „hineinrutschen“.
- Für ein feiner gestuftes Raster mit höherer Differenzierungsfähigkeit müssen weitere Trennwerte eingeführt werden. Würde beispielsweise im Rahmen der quartilsbasierten QZM 1 auch noch der Median als Trennwert berücksichtigt, entstünde ein vierstufiges Bewertungsraster. Da dies allerdings keinen genuine Durchschnittsbereich mehr umfassen würde, wäre ein solches System eher unpraktikabel, da es in der Rezeption durch Dritte vermutlich als defizient wahrgenommen werden würde.
- Im Rahmen einer Trennung auf der Basis von Quintilen mit vier Trennwerten könnten Anbieter in fünf 20 % Klassen eingeordnet werden. Auch unabhängig von der Methode, mit der es erstellt wird, stellt ein derart fünfstufiges Bewertungssystem aus Sicht der Gutachter den bestmöglichen Kompromiss dar und ist auch entsprechend weit verbreitet (siehe nächster Absatz).

Bei Sichtung der internationalen Literatur zur Bewertung von Qualitätsmessungen und insbesondere bei den bereits seit längerem erprobten Qualitätsberichterstattungssystemen in den Vereinigten Staaten, sind feiner gestufte Bewertungen in der Regel der Standard. Dies gilt insbesondere für fünfstufige Bewertungssysteme. Beispielhaft sei an dieser Stelle auf das „Nursing Home Compare Five-Star-Quality Rating System“ (Wagner, McDonald, & Castle, 2012) oder die „Minnesota NH report card“ (Nazir et al., 2012) verwiesen. Bei der Fünf-Sterne-Bewertung gelten folgende Qualitätszuweisungen: Weit überdurchschnittliche Anbieter erhalten fünf Sterne, überdurchschnittliche vier, durchschnittliche drei, unterdurchschnittliche zwei sowie weit unterdurchschnittliche einen Stern (vgl. Wagner et al., 2012). Grundsätzlich erscheint ein solches Bewertungsmodell ohne Einschränkungen auch auf den deutschsprachigen Raum übertragbar. Für das Fünf-Sterne-System spricht ferner, dass es sich durch seine Ähnlichkeit etwa mit den bei Hotels üblichen Sternevergaben auch verbraucherseitig intuitiv erschließt, analog etwa zu Schulnoten o. ä. (s. Kapitel 4.5). Für die Akzeptanz eines solchen Systems seitens der Anbieter könnte ferner hilfreich sein, dass eine geringe Anzahl von Hotelsternen deutlich weniger negativ besetzt ist als bspw. eine „mangelhafte“ Schulnote. In der Studie von Nazir et al. (2012) konnte zudem gezeigt werden, dass im Zuge

einer Akkreditierungskampagne das Fünf-Sterne-System hinreichend änderungssensitiv war, um Qualitätsverbesserungen bei den untersuchten Anbietern aufzeigen zu können.

Wingenfeld et al. (2011) lehnen ein fünfstufiges Bewertungssystem mit Hinweis auf die durch Stichprobenbedingte Zufallsschwankungen bedingte Gefahr zunehmend ungenaueren Bewertungen ab. Grundsätzlich ist es richtig, dass ein differenzierteres Bewertungssystem messfehleranfälliger wird. Um hier ein klareres Bild zu bekommen, sollten die Einzelindikatoren nochmals etwas genauer in den Blick genommen werden. Dies leitet zu unserem zweiten zentralen Diskussionspunkt über, der Differenzierungsfähigkeit der Qualitätsindikatoren bzgl. ihrer statistisch-empirischen Eigenschaften.

Im Hinblick auf die empirische Differenzierungsfähigkeit von Qualitätsindikatoren kann als grobe Faustregel gelten, dass diese bei sonst vergleichbaren Rahmenbedingungen umso besser differenzieren, je mehr Variabilität sie aufweisen. D.h. je breiter das Qualitätsspektrum ist bzw. je vielschichtiger und häufiger die Outcomes sind, die durch einen Indikator beschrieben werden, desto eher können feiner gestufte Qualitätsunterschiede valide abgebildet werden. Andersherum nimmt die Messfehleranfälligkeit (höherer Standardfehler) zu, je seltener und je weniger variabel ein Outcome ist (Castle & Ferguson, 2010; Mor et al., 2003). In den vorliegenden Daten kann die Variabilität der Wingenfeld-Indikatoren anhand der für die Zuordnung zu den drei Qualitätsklassen maßgeblichen empirischen Schwellenwerte abgeschätzt werden (s. Tabelle 6). Im Rahmen der beschriebenen QSM 1 eignet sich hierfür der Interquartilsabstand, der in der Statistik neben der Standardabweichung ein sehr gebräuchliches Dispersionsmaß ist. Bei den Indikatoren des Qualitätsbereichs 2 sowie dem Indikator „Einsatz von Gurtfixierungen“, bei dem die durchschnittswertorientierte QZM 2 angewendet wird, kann zumindest die prozentuale Differenz zwischen den beiden Schwellenwerten herangezogen werden (s. Tabelle 6). Diese ist zwar insbesondere in Unkenntnis, ob etwa eine rechtsschiefe, linksschiefe, Gleich- oder Normalverteilung vorliegt, kein „echtes“ Variabilitätsmaß, ist hier in Ermangelung an besseren Alternativen aber zumindest für eine grobe Abschätzung der Variabilität brauchbar. Leider gibt es in der Statistik keine wirklich gängigen Konventionen, ab wann eine vorliegende Verteilung eine hinreichende Streuung aufweist. Als grober Richtwert kann hier gelten, dass die Streuung mindestens etwa 10 % der maximalen Spannweite betragen sollte. Vor diesem Hintergrund kann im vorliegenden Kontext den Wingenfeld-Indikatoren des Qualitätsbereichs 1 (Förderung und Erhalt von Selbständigkeit) eine hinreichende Differenzierungsfähigkeit attestiert werden. Deutlich kritischer sind allerdings die Indikatoren des Qualitätsbereichs 2 zu beurteilen, die insgesamt deutlich weniger variabel scheinen. Besonders kritisch sind dabei die Niedrigrisikooptimalisierungen der Indikatoren Dekubitusentstehung und Gewichtsverlust (s. Tabelle 6). Beide beschreiben analog zum Indikator „Gurtfixierungen“ aus Qualitätsbereich 3, der ebenfalls eine geringe Variabilität aufzuweisen scheint (vgl. Schnelle et al., 2004), Dimensionen mit allgemein niedriger Inzidenz. Darunter fallen auch alle sog. „Sentinel events“, also seltene Ereignisse, die aber ein hohes Gefährdungspotenzial für die BewohnerInnen haben (vgl. Faust, 2004). Wie gesagt ist es in solchen Bereichen grundsätzlich schwierig, zu differenzierenden Einschätzungen zu kommen. Ein weiterer Indikator mit eher unzureichender Diskriminationsfähigkeit stammt ebenfalls aus Qualitätsbereich 3: Bei den „Integrationsgesprächen nach dem Heimeinzug“ ist deren praktische Seltenheit sogar der explizite Grund, warum Wingenfeld et al. (2011) von den sonst überwiegenden referenzstichprobenorientierten Klassifikationsmethoden (QZM 1 und 2) hier zugunsten eines kriterienbezogenen Expertenstandard (QZM 3) ab-

weichen. Da die Verwendung von Indikatoren mit mangelhafter Unterscheidungsfähigkeit im Rahmen eines Qualitätsbenchmarkings natürlich grundsätzlich fragwürdig erscheint, könnte sich in dieser Hinsicht, zumindest für die beiden Niedrigrisikovarianten der Dekubitus- und Gewichtsverlust-Indikatoren, eine Lösung auch im Kontext der von den Gutachtern im Rahmen der Risikoadjustierung vorgeschlagenen alternativen multivariaten Adjustierungsmethode abzeichnen, die eine Trennung nach hohem und niedrigem Risiko obsolet machen würde und sich dann vermutlich auch positiv auf die Diskriminationsfähigkeit der entsprechenden Indikatorbereiche auswirken würde (s. Kapitel 4.4).

4.3.2 Gibt es Alternativen, z. B. die ergänzende Berücksichtigung von zu erfüllenden Mindestniveaus, und wie werden diese bewertet?

Handelt es sich bei den Qualitätszuweisungsmethoden QZM 1 und 2 (s. Kapitel 4.3.1) um „normorientierte“ Vorgehensweisen, bei denen das Bezugssystem für den Anbietervergleich auf Grundlage einer empirisch ermittelten Referenzstichprobe definiert wird, anhand derer sich das individuelle Abschneiden eines Anbieters mit dem Abschneiden aller Anbieter vergleichen lässt (Wilson & Twiss, 1999), stellt die Qualitätszuweisungsmethode 3 ein sog. kriteriums-basiertes Vorgehen dar, bei dem die Schwellenwerte nicht relativ in Bezug auf eine Vergleichsstichprobe, sondern anhand von Experten festgelegten Kriteriumsstandards definiert werden. Eine Variante des kriteriums-basierten Ansatzes stellt die Festlegung von Mindestniveaus dar (siehe unten). Da bei 12 der 15 von Wingefeld et al. (2011) entwickelten Qualitätsindikatoren normorientiert vorgegangen wird (vgl. Tabelle 6), überwiegt diese Methode beim vorliegenden Ansatz zwar deutlich, da die Autoren aber auf beide denkbaren grundlegenden Verfahrensspielarten zur Bestimmung der spezifischen Trennwerte der Qualitätsniveaus im Rahmen eines Einrichtungsvergleichs zurückgreifen, kann folglich nicht von der einen Alternative gesprochen werden. Die spezifischen Vor- und Nachteile der beiden Verfahren werden daher im Folgenden gegeneinander abgewogen, um auf dieser Grundlage eine Empfehlung aussprechen zu können, welches Verfahren für die zukünftige Weiterentwicklung des Wingefeld-Ansatzes stärker in den Fokus rücken sollte.

Ein normorientiertes Benchmarkingverfahren ist in der Regel dann indiziert, wenn es keinen etablierten (Kriteriums-)Standard zur Bemessung der Anbieterqualität gibt, der allgemein akzeptiert ist (vgl. Mor et al., 2003). Aus dieser Perspektive erscheint es nur folgerichtig, dass Wingefeld et al. (2011) im Rahmen ihrer Erstentwicklung eines Ergebnisqualitätsindikatorsets für das deutsche Gesundheitswesen, bei dem empirische Orientierungspunkte weitestgehend fehlen bzw. die Übertragbarkeit der Studienergebnisse aus anderen Ländern nur schwer abschätzbar ist, diesen Ansatz in den Mittelpunkt stellen. Trotzdem hat der normorientierte Ansatz einige gravierende Nachteile. Diese resultieren vornehmlich aus der Abhängigkeit von den konkret vorgefundenen Verteilungsmustern der Daten:

- Die Orientierung an einer „Norm“ ist im Kontext von Qualitätsvergleichen immer dann kritisch, wenn das Qualitätsniveau insgesamt niedrig ausfällt. Da hier alle Qualitätsbeurteilungsstufen, die in der Bewertungssystematik vorkommen können, immer gleichmäßig über die vorgefundene Spannweite der tatsächlich beobachteten Qualitätsniveaus verteilt werden, gibt es völlig unabhängig davon, wo der Spitzenwert im Einzelnen absolut gesehen liegt, immer Einrichtungen, die als überdurchschnittlich bewertet werden. Ein plastisches Beispiel, wo dies problematisch wäre, findet sich beim Wingefeld-Indikator „Integ-

rationsgespräche nach Heimeinzug“. Da diese insgesamt kaum eine gängige Praxis darstellen, könnten bei normorientierter Betrachtung auch Anbieter als überdurchschnittlich bewertet werden, die unregelmäßig und selten Integrationsgespräche durchführen. Entsprechend weichen die Autoren bei diesem Indikator folgerichtig auf einen Kriteriumsstandard zur Qualitätsbewertung aus. Salopp gesagt wird im Extremfall auf diese Weise dann Schlendrian mit Schlendrian verglichen. So fordern etwa O'Reilly, Courtney, Edwards und Hassall (2011), dass sich Bewertungen nicht an durchschnittlicher Pflegequalität, sondern möglichst an „Exzellenz-Kriterien“ ausrichten sollte.

- Das Bezugssystemproblem kann nicht nur dazu führen, dass bei allgemein niedriger Qualität Anbieter gewissermaßen zu Unrecht als herausragend eingestuft werden, sondern umgekehrt bei insgesamt hoher Qualität in der Referenzstichprobe dazu führen, dass auch ein an sich schon ordentliches Qualitätsniveau als unterdurchschnittlich bewertet wird. Wenn Einrichtungen trotz adäquater Pflege schlecht abschneiden, kann dies die Akzeptanz des Qualitätsbewertungssystems erheblich beschädigen (Grando, Rantz, & Maas, 2007).
- Im Hinblick auf die Differenzierung von Einrichtungen erscheint zunächst positiv, dass beim normorientierten Ansatz alle Qualitätsbewertungsstufen gleichmäßig besetzt werden. Insbesondere bei seltenen Outcomes wie etwa den bereits erwähnten „Sentinel events“ oder Indikatoren mit vergleichsweise geringer Variabilität wie etwa der Niedrigrisikovariante des Indikators Dekubitusentstehung in der Wingenfeld-Systematik (vgl. auch Abschnitt „Differenzierung von Einrichtungen“ oben), kann dies aber schnell eine Pseudo-Diskriminationsfähigkeit suggerieren. Die in diesen Fällen zwangsläufig höhere Fehlerstreuung macht eine zuverlässige Einordnung in die jeweiligen Qualitätsstufen dann nahezu unmöglich.
- Zugunsten der normorientierten Methode wird häufig argumentiert, dass sich bei insgesamt steigender Qualität keine Einrichtung auf ihrem Status-Quo ausruhen kann, da sonst ein Abrutschen in den Qualitätsbewertungen droht. Dieses Argument ist prinzipiell durchaus stichhaltig und vor dem Hintergrund des zugrundeliegenden dynamisch-prozesshaften Verständnisses von Qualitätsmanagement im Sinne eines stetigen Weiterentwicklungsprozesses („der Weg ist das Ziel“) grundsätzlich auch nur zu begrüßen. Problematisch wird es aber auch hier, wenn sich die Qualität bei allen Einrichtungen in gleichmäßigem Ausmaß und Tempo weiterentwickelt. Dies würde unweigerlich dazu führen, dass praktisch durchaus fruchtende Optimierungsanstrengungen nicht mit einem besseren Qualitätsrating „belohnt“ werden würden. Ein Bewertungssystem, das Entwicklungen nicht abbilden kann, hätte langfristig wohl ebenfalls nur eine begrenzte Akzeptanz.

Eine normbasierte Bewertungssystematik kann dann zielführend sein, wenn die Referenzstichprobe das gesamte Qualitätsspektrum abdeckt und annähernd eine Normalverteilung aufweist. Da davon abweichende Verteilungsmuster und die damit ggf. verbundenen statistischen Einschränkungen im praktischen Einsatz in der Regel wohl weder Verständnis noch Berücksichtigung finden dürften, sind die darauf fußenden Qualitätsbewertungen im Grunde nicht seriös interpretierbar.

Eine ernstzunehmende Alternative dazu bietet das kriteriums-basierte Vorgehen, bei dem die Schwellenwerte für die Qualitätsniveaus durch Expertenurteile festgelegt werden. Die einfachste Variante dieses Vorgehens besteht in der Definition von Minimalstandards, bei denen

qualitative Mindestniveaus erreicht werden müssen. Diese werden in der Literatur allerdings oftmals kritisch gesehen. Als zentrale Argumentationslinien gegen den Einsatz von Mindeststandards wird ins Felde geführt, dass diese am Ende in der Praxis dann oft als Maximalstandards fungieren, insbesondere wenn es um die Zuweisung finanzieller Ressourcen ginge sowie imageschädigende Diskussionen der Art „Welcher Prozentsatz an Dekubitusgeschwüren ist akzeptabel?“ hervorbrächten (ActiZ et al., 2007). Diese Einwände sind sicher berechtigt, wenn ausschließlich (einfache) Mindeststandards erstellt werden. Nach Vorstellung der Autoren der vorliegenden Expertise müssten aber mindestens zwei Trennwertkriterien definiert werden: ein Grundstandard, der die Grenze für gerade noch akzeptable Qualität darstellt sowie darüber hinaus noch ein transparenter, an guter Qualität orientierter, „Zielstandard“. Auf diese Weise ist es leichter ein, für die Differenzierung von Einrichtungen notwendiges, möglichst feinstufiges Bewertungsraster anzulegen, das nach unseren Vorstellungen auch hier fünfstufig sein sollte (siehe Diskussion im Abschnitt „Differenzierung von Einrichtungen“). Dabei wäre es weder notwendig noch in der konkreten Umsetzung besonders zielführend und praktikabel, wenn alle dafür notwendigen Schwellenwerte durch Expertenübereinkunft festgelegt werden würden. Primär kommt es darauf an, die Eckpunkte für gute und schlechte Qualität, also Grund- und Zielstandard, zu umreißen. Eine Binnendifferenzierung der Klassengrenzen kann dann auch auf empirischem Weg vorgenommen werden. Durch den Zielstandard ist ferner eine Stoßrichtung für die Initiierung von Qualitätsentwicklungsprozessen gegeben. Dabei sind gleichsam „Verwechslungen“ mit den Mindestniveaus ausgeschlossen.

Die Argumentation, dass die Bemühungen um Kriteriumsstandards in ideologisch-kontraproduktiven Grabenkämpfen rund um Prozentzahlen münden würden, greift definitiv zu kurz. Jede Profession, die dauerhaft ein ernstzunehmendes eigenständiges fachliches Gewicht für sich in Anspruch nehmen möchte, muss irgendwann Standards formulieren. Dies gilt sowohl für die Struktur- und Prozessebene als eben auch für die Ergebnisqualitätsebene. Dabei muss es nicht notwendigerweise nur um Prozentzahlen gehen. Ein Ergebnis dieses Diskurses kann eben auch sein, dass bestimmte Zwangsmaßnahmen wie Gurtfixierungen unter keinen denkbaren pflegerischen Settings tolerabel sein sollten (Mor et al., 2011). Dabei kann durch Standards natürlich auch sichergestellt werden, dass für Indikatoren, die aus fachlicher Sicht als besonders relevant und sensibel eingestuft werden, besonders hohe Qualitätsanforderungen gelten (wie etwa von Wingefeld et al. (2011) für den Indikator „Schmerzmanagement“ vorgeschlagen). Nur durch Kriterienvorgaben kann letztendlich gewährleistet werden, dass garantiert gute Qualität und ein hoher Grad an Fachlichkeit, im Sinne der höchstmöglichen Annäherung an die Exzellenz, die Zielmarken des Qualitätsdialogs sind und nicht der sprichwörtliche graue Durchschnitt. Grundsätzlich würden wir es daher präferieren, dass auch die von Wingefeld et al. (2011) vorgeschlagene Bewertungssystematik in diese Richtung weiterentwickelt werden würde. Bei den Indikatoren Schmerzmanagement, Verhaltensauffälligkeiten sowie Integrationsgespräche ist dies ja bereits in vielversprechender Weise realisiert.

Wie kann eine Bestimmung von Kriteriumsstandards auf der Basis von Expertenurteilen aber sinnvollerweise ablaufen? Bei Durchsicht der Literatur fällt schnell auf, dass es dazu nur wenige Informationen gibt. Der Ansatz ist aktuell leider noch nicht weit verbreitet. Dies mag damit zusammenhängen, dass die Festlegung von festen Kriterien bzw. Standards eine anspruchsvolle Zielsetzung ist, die keineswegs trivial oder gar „per Order di Mufti“ zu realisieren

ist. Aufgrund der großen Vorteile lohnt es sich aus unserer Sicht aber sehr, diesen Weg trotz aller potenziellen Hemmnisse einzuschlagen. Ein mustergültiges Vorgehen wird von O'Reilly, Courtney, Edwards und Hassall (2011) im Rahmen der Umsetzung des ResCareQA („Residential Care Quality Assessment“) in Australien beschrieben. Die Standards wurden hierbei in folgenden Schritten definiert (vgl. O'Reilly, Courtney, Edwards, & Hassall, 2011):

1. Bildung eines Expertengremiums aus Forschern, Leitungen und Fachkräften.
2. Übermittlung und Darstellung der empirischen Grundlagen (Verteilungen und ranggeordnete Ergebnisse der einzelnen Qualitätsindikatoren) als Grundinformation für die Experten.
3. Festlegung von oberen und unteren Schwellenwerten für die Fragen, wann Pflegeergebnisse nicht mehr akzeptabel sind und wann exzellente Qualität beginnt. Dies findet auf der Basis von Delphi-Prozessen (Ammon, 2009) statt. Dabei handelt es sich um eine mehrstufig rückgekoppelte Befragungstechnik, die darauf abzielt, einen Expertenkonsens durch Ausschaltung von potenziell negativen Einflüssen gruppendynamischer Prozesse (z. B. eine „Wertediktatur“ durch dominante Persönlichkeiten) zu erreichen. In einer ersten Befragungsrunde schätzen die Experten die Fragen unabhängig von der Gruppe auf Grundlage der Daten, aktueller Forschungstrends und eigener Erfahrungen ein. Die Ergebnisse werden dann in der Regel anonym an einen „Mediator“ übermittelt, der diese zusammenfassend auswertet (bspw. die mittleren Experten-Schwellenwerte für jeden Indikator).
4. In weiteren Runden (im vorliegenden Beispiel drei) wird die zusammengefasste „Gruppenmeinung“ an die Experten zurückgespiegelt. Die einzelnen Schwellenwerte werden dann so lange adjustiert, bis die Streuung der Einschätzung in der Gruppe soweit minimiert ist, dass ein akzeptabler Grad an Übereinstimmung erreicht ist.

Eine entscheidende Bedeutung kommt dabei der Besetzung des Expertengremiums zu. Es ist für die Akzeptanz der Arbeitsergebnisse wichtig, dass alle zentralen Stakeholder und Interessensgruppen möglichst egalitär vertreten sind. Dabei muss aber im Hinblick auf die Konsensfähigkeit und zeitliche Dimension der operativen Entscheidungsfähigkeit auch darauf geachtet werden, dass das Gremium nicht zu umfangreich besetzt wird. Die Besetzung mit sechs Experten wie in der Beispielstudie (O'Reilly et al., 2011) tangiert hier aber auch das unterste noch akzeptable Limit schon deutlich.

Ein weiteres, aus Sicht der Autoren der vorliegenden Expertise unverzichtbares Element für die adäquate Umsetzung eines kriterienbezogenen Bewertungssystems ist die Einbeziehung empirischer Daten. Es sollte also immer induktiv und nicht rein deduktiv vorgegangen werden. Oder anders ausgedrückt: Standards sollten nicht aus dem „luftleeren Raum“ entstehen, da die Konsensfindung sonst insbesondere durch ideologische Barrieren erschwert werden kann. Wenn beispielsweise Gurtfixierungen aus fachlich-theoretischer Sicht abzulehnen sind, diese aber überall gängige Praxis wären, wäre die Kluft zwischen Anspruch und Wirklichkeit so groß, dass keine realistische Zielsetzung formuliert werden könnte, die in einem produktiven Qualitätsoptimierungsprozess münden könnte. In diesem Fall wäre es besser Zielmarken zu setzen, den Anteil an Gurtfixierungen in Anlehnung an die relativ gesehen besseren Anbieter sukzessive zu erniedrigen.

Durch ein datengestütztes Vorgehen im Rahmen der Delphi-Prozesse könnten letztlich die Stärken beider Bezugssysteme, sowohl des normorientierten als auch des kriterienorientierten, miteinander kombiniert werden. Damit die empirischen Daten richtig interpretiert werden, sollten auch Statistiker bzw. Personen mit statistischer Expertise in das Expertengremium aufgenommen werden. Auch bei der expertengeleiteten Festlegung von Qualitätsniveaus muss darauf geachtet werden, dass die Setzungen eine hinreichende Differenzierungsfähigkeit gewährleisten. Wenn die Trennwerte zu eng beisammen liegen oder so bemessen sind, dass am Ende alle Anbieter ausschließlich im guten bis sehr guten Bereich liegen, ist der Informationsgehalt der Bewertungssystematik und damit auch der praktische Nutzen für die VerbraucherInnen (s. Kapitel 4.5) gering.

Ein datengeleitetes Vorgehen ist schlussendlich auch vor dem Hintergrund wichtig, dass die Kriteriumsstandards keine in Stein gemeißelten und für alle Ewigkeit gültigen Werte darstellen sollten. Auch im Zuge des kriteriumsbezogenen Ansatzes ist es wichtig, die Festlegungen z. B. in Synchronisierung mit den Ergebnissen von längerfristigen Qualitätsentwicklungsprozessen oder etwa geänderten gesetzlichen Rahmenbedingungen, in regelmäßigen Abständen einer Neubewertung und ggf. Neufestlegung zu unterziehen (vgl. Rantz et al., 2009). In Anbetracht des damit verbundenen Aufwandes wären hier Zeitintervalle in der Größenordnung von 5-10 Jahren praktikabel.

4.3.3 Fazit

Für eine differenzierte Einteilung von Einrichtungen erscheint die von Wingenfeld et al. (2011) vorgelegte dreistufige Bewertungssystematik als zu grob. Die AutorInnen der vorliegenden Expertise empfehlen analog zum internationalen Standard ein fünfstufiges Bewertungsraster, das in weit überdurchschnittliche, überdurchschnittliche, durchschnittliche, unterdurchschnittliche und weit unterdurchschnittliche Einrichtungen unterteilt. Die Differenzierungsfähigkeit wird allerdings nicht nur durch die Feingliedrigkeit der Bewertungsstufen, sondern auch durch die tatsächlich zu beobachtende Variabilität der Messwerte beeinflusst. In diesem Kontext erweisen sich insbesondere einige der für BewohnerInnen mit niedrigen Risikoausprägungen vorgesehenen Varianten der Wingenfeld-Indikatoren als ungeeignet. Die von uns vorgeschlagene alternative Risikoadjustierungsmethode, die ohne derartige Unterscheidungen auskommt, könnte hier allerdings Abhilfe schaffen. Auch andere von Wingenfeld et al. (2011) vorgenommene Setzungen können nicht durchgehend überzeugen. Von der Verwendung indikatorübergreifender Gesamtwerte zum Vergleich von Einrichtungen, raten wir auf der Grundlage des vorliegenden Forschungsstandes ab. Damit am Ende nicht sprichwörtlich „Schlendrian mit Schlendrian“ verglichen wird, sollte die Qualitätsbewertungen nicht ausschließlich am Anbieterdurchschnitt bemessen werden. Wir würden daher dringend empfehlen zukünftig verstärkt expertenbasierte Kriteriumsstandards zu formulieren. Damit sich bessere Qualität auch lohnen kann, sollte auf dem Hintergrund der Initiierung von ziel führenden Qualitätsentwicklungsprozessen, aber nicht ausschließlich auf Mindestniveaus zurückgegriffen werden. Die Kriteriumsstandards sollten mindestens zwei Orientierungsmaßstäbe setzen, einen Grundstandard, der noch akzeptable Qualität umreißt und einen Zielstandard, ab dem höherwertige „gute“ pflegerische Qualität beginnt. Eine fortlaufende Nachjustierung wäre allerdings auch hier unerlässlich.

4.4 Risikoadjustierung

(Jens Arnold)

4.4.1 Sind die der Risikoadjustierung zugrunde gelegten Kriterien (Items aus dem NBA) angemessen, um eine valide Risikoadjustierung zu ermöglichen? Gibt es Hinweise oder Vorschläge zu Alternativen oder zur Weiterentwicklung?

Hintergrund Risikoadjustierung

Unter Risikoadjustierung wird im Kontext der Outcomemessung ein Verfahren verstanden, das die Ergebnisse von unterschiedlichen Anbietern vergleichbar macht. Durch die Risikoadjustierung wird den unterschiedlichen Zielgruppen und Bewohnerzusammensetzungen („Case-Mix“) Rechnung getragen (vgl. Li, Schnelle, Spector, Glance, & Mukamel, 2010). Es ist unmittelbar evident, dass pflegerische Ergebnisse neben den professionellen Dienstleistungen von individuellen Risiken der BewohnerInnen abhängen: Je mehr physische und kognitive Beeinträchtigungen vorliegen, umso kränker, älter, gebrechlicher und je höher die Komorbidität, desto größer wird die Herausforderung, ein gutes Outcome erreichen zu können. Mukamel et al. (2004) weisen völlig zu Recht darauf hin, dass Anbieter nicht mit einer schlechteren Qualitätsbewertung „bestraft“ werden dürfen, nur weil sie sich beispielsweise auf Zielgruppen mit höherem Risiko spezialisiert haben. Es kann ebenfalls nicht Sinn und Zweck sein, dass Anbieter BewohnerInnen mit niedrigem Risiko bevorzugen, nur um damit eine bessere Qualitätseinstufung zu begünstigen (Konetzka, Polsky, & Werner, 2013).

Durch die Risikoadjustierung wird dafür gesorgt, dass nicht sprichwörtlich Äpfel mit Birnen verglichen werden. Die Qualität der pflegerischen Prozesse darf dabei natürlich nicht mitadjustiert bzw. kontrolliert werden. Es geht bei der Adjustierung letztlich darum zu gewährleisten, dass sich ausschließlich die professionelle Leistung in den Outcomes widerspiegelt und alle anderen Faktoren, die von den Anbietern nicht beeinflussbar sind, methodisch kontrolliert werden (vgl. Phillips, Chen, & Sherman, 2008). Dieser „Kunstgriff“ ist keineswegs trivial. Da aber nur auf diese Weise ein fairer Anbietervergleich möglich ist (vgl. auch Kapitel 4.3 und Kapitel 4.5), müssen nach Meinung der Gutachter Überlegungen zu einer angemessenen Risikoadjustierung immer ein elementarer Bestandteil der Betrachtung von outcomebasierten Qualitätsindikatoren sein. Damit soll aber nicht gesagt sein, dass eine Risikoadjustierung bei jedem Indikator zwingend notwendig sein muss. Nach Mor et al. (2011) sollte beispielsweise grundsätzlich keine Risikoadjustierung auf den Indikator Gurtfixierungen angewendet werden, da aus Sicht der Autoren keine spezifischen Eigenschaften von Bewohnern bzw. Bewohnerinnen vorstellbar seien, unter denen diese zwingend indiziert wären.

Die Risikoadjustierung ist jeweils individuell unter Einbeziehung von Forschungsbefunden und inhaltlichen Überlegungen abzuwägen. Mitunter können für einzelne Indikatoren auch unterschiedliche Adjustierungsmethoden angemessen sein (vgl. Horn et al., 2005). Aufgrund der elementaren Bedeutung der Risikoadjustierung werden folgerichtig auch von Wingenfeld et al. (2011) Vorschläge zur Adjustierung der von ihnen beschriebenen Qualitätsindikatoren unterbreitet. Diese sollen an dieser Stelle diskutiert werden.

Überblick Adjustierungsverfahren und Einordnung des Wingenfeld-Ansatzes

Um die von Wingenfeld et al. (2011) vorgeschlagenen Risikoadjustierungskriterien einordnen zu können, ist zunächst ein Überblick über die gebräuchlichen Adjustierungsverfahren erforderlich. In der Literatur werden vorwiegend vier Verfahren beschrieben: Stratifikation, multiple Regression sowie hierarchische lineare Modelle (vgl. AQUA 2010):

- **Stratifizierung** ist eine gängige und einfache Form der Risikoadjustierung. Dabei werden Risikogruppen gebildet, die sog. „Strata“ bzw. Schichten, die dann separat getrennt voneinander betrachtet werden. Das Verfahren ist allerdings nur praktikabel, solange die Zahl der zum Vergleich herangezogenen Schichten innerhalb der Risikofaktoren und Faktorenkombinationen überschaubar bleibt.
- Bei der **multiplen Regression**, die bei dichotomen Qualitätsindikatoren als ein logistisches und bei stetigen Qualitätsindikatoren als ein lineares statistisches Regressionsmodell formuliert wird, besteht dieser Nachteil nicht. Hier kann der Einfluss von beliebig vielen bewohnerspezifischen Risikofaktoren auf die jeweiligen Qualitätsindikatoren in die risikoadjustierten Einrichtungsvergleiche einbezogen werden. Neben dem Ausmaß des Einflusses einzelner Faktoren könnten dabei theoretisch gesehen sogar Wechselwirkungen berücksichtigt werden (vgl. Methode von Cox & Wermuth, 1994). Letzteres ist im Moment aber noch keine gängige Praxis.
- Bei den **hierarchischen linearen Modellen** handelt es sich genau wie bei der multiplen Regression ebenfalls um ein multivariates Verfahren, das zunehmend in den Fokus des fachlichen Diskurses rückt und in der Bewertung immer mehr als „Königsweg“ angesehen wird. Bei diesem Ansatz kann zusätzlich die Varianz der Outcomes innerhalb und zwischen den Einrichtungen sowie die Wechselbeziehungen der Beobachtungen (sog. Mehr-Ebenen-Struktur der Daten) in die Adjustierung einbezogen werden. Die Realisierung ist allerdings an hohe statistisch-methodische Anforderungen und Kompetenzen gebunden (s. Diskussion und Empfehlungen weiter unten).
- Eine im extremen Kontrast zu den hierarchischen Modellen stehende, überaus basale Form des Umgangs mit Risiken, die u. a. in einer Studie von Arling et al. (2007) besprochen wird, ist die **Ausschließung** (auch „risikostandardisierte Fallkonstellation“). Dabei werden überhaupt nur Fälle mit vergleichbaren Risiken betrachtet. Ausgeschlossen werden beispielsweise Ergebnisse, die nicht von den Anbietern beeinflusst werden können wie z. B. bereits bei Aufnahme vorhandene Druckgeschwüre, tödliche Erkrankungen oder Koma (Sangl et al., 2005). Fälle mit abweichenden Risiken fallen hier sprichwörtlich unter den Tisch. Dies kann mitunter eine erhebliche Reduzierung der Fallzahl mit sich führen. Die den Ausschließungen inhärente Problematik wurde bereits in Kapitel 4.1.3 dargelegt. Einige Autoren (z. B. Weibler-Villalobos & Röhrig, 2010) sehen Ausschlüsse von Risiken im Pflegebereich grundsätzlich kritisch. Die Frage ist hier insbesondere, wie gewährleistet werden kann, dass die inhaltlich und argumentativ i. d. R. durchaus nachvollziehbaren Ausschlüsse von Personengruppen in der praktischen Umsetzung so messgenau gestaltet werden können, dass sie nicht den Anschein von Willkürlichkeit erwecken. Obwohl die Ausschließung in der Literatur als Risikoadjustierungsmethode genannt wird, werden hier in der Regel auch Fragen der Indikatoroperationalisierung tangiert, bei denen es neben der Frage nach dem

Handling von Risiken auch stärker darum gehen kann, welche Bereiche durch Pflege überhaupt beeinflussbar sind.

Tabelle 8: Übersicht Risikoadjustierungskriterien bei den Wingenfeld-Indikatoren

Qualitätsbereich	Indikatorbereich	Risikoadjustierungskriterium	Auswirkung der Adjustierung
Bereich 1: Erhaltung und Förderung von Selbständigkeit	Mobilität	Stratifizierung nach NBA-Modul „Kognitive und Kommunikative Fähigkeiten“	2 separate Indikatoren für Bewohner mit niedrigen und erheblichen kognitiven Einbußen
	Alltagsverrichtungen	Stratifizierung nach NBA-Modul „Kognitive und Kommunikative Fähigkeiten“	2 separate Indikatoren für Bewohner mit niedrigen und erheblichen kognitiven Einbußen
	Gestaltung des Alltagslebens und sozialer Kontakte	Keine Risikoadjustierung	-
Bereich 2: Schutz vor gesundheitlichen Schädigungen und Belastungen	Dekubitusentstehung	Stratifizierung nach dem NBA-Item „Positionswechsel im Bett überwiegend oder komplett selbständig durchführen“	2 separate Indikatoren für Bewohner mit niedrigen und hohem Dekubitusrisiko
	Stürze mit gravierenden Folgen	Stratifizierung nach NBA-Modul „Kognitive und Kommunikative Fähigkeiten“	2 separate Indikatoren für Bewohner mit niedrigen und erheblichen kognitiven Einbußen
	Unbeabsichtigter Gewichtsverlust	Stratifizierung nach NBA-Modul „Kognitive und Kommunikative Fähigkeiten“	2 separate Indikatoren für Bewohner mit niedrigen und erheblichen kognitiven Einbußen
Bereich 3: Unterstützung bei spezifischen Bedarfslagen	Integrationsgespräch nach dem Heimeinzug	Keine Risikoadjustierung	-
	Einsatz von Gurtfixierungen	Keine Risikoadjustierung	-
	Einschätzung von Verhaltensauffälligkeiten bei Bewohnern mit kognitiven Einbußen	Ausschlussverfahren: nur Bewohner mit min. erheblichen Einbußen NBA-Modul „Kognitive und Kommunikative Fähigkeiten“	-
	Schmerzmanagement	Keine Risikoadjustierung	-

Der Frage nach einer validen Risikoadjustierung kommt in der Bewertung von Einzelindikatoren und Indikatorensets eine entscheidende Schlüsselstellung zu. Dies gilt insbesondere, wenn ein Einrichtungsvergleich bzw. Benchmarking angestrebt wird. Eine unzureichende Risikoadjustierung kann letztlich den praktischen Nutzen eines jeden in der Erhebung noch

so sensitiven, trennscharfen, messgenauen und validen Indikators nahezu vollständig zunichtemachen.

Wie wurde die Risikoadjustierung im Rahmen der Wingenfeld-Indikatoren umgesetzt? Tabelle 8 gibt einen Überblick. Im Hinblick auf die Empfehlungen zur Weiterentwicklung unterscheiden wir in der folgenden Diskussion ferner zwischen dem gewählten Risikoadjustierungsverfahren als solchem und den im Rahmen dieses Verfahrens im einzelnen verwendeten Methoden zur Bestimmung bzw. Messung der von den Autoren zur Adjustierung ausgewählten Risikofaktoren.

Wie Tabelle 8 zu entnehmen ist, greifen Wingenfeld et al. (2011) überwiegend auf die Stratifizierungsmethode zurück. Bei einem einzelnen Indikator „Verhaltensauffälligkeiten von Bewohnern mit kognitiven Einbußen“ werden Personen mit nicht mindestens erheblichen Einbußen ausgeschlossen. Auf eine Risikoadjustierung verzichtet wird bei den meisten Indikatoren des Qualitätsbereichs 3 „Unterstützung bei spezifischen Bedarfslagen“ sowie bei dem Indikator „Gestaltung des Alltagslebens und sozialer Kontakte“. Die von Wingenfeld et al. (2011) vorgenommene Stratifizierung bildet jeweils zwei Risikogruppen, die dann separat voneinander betrachtet werden. Dies fließt unmittelbar mit in die Operationalisierung der Indikatoren mit ein, da für jede Risikogruppe ein eigener Indikator definiert wird. Für den Bereich Mobilität beispielsweise ein Indikator für BewohnerInnen mit keinen bis geringen kognitiven Einbußen und ein Indikator für BewohnerInnen mit erheblichen Einbußen. Charakteristisch für die hier vorgenommene Stratifizierung ist ferner, dass pro Indikatorbereich (s. Tabelle 8) jeweils nur ein einzelner Risikofaktor und keine Kombinationen von unterschiedlichen Risikofaktoren berücksichtigt werden. Abgesehen vom Indikator Dekubitus, dessen Risiko anhand der Fähigkeit der Bewohner und Bewohnerinnen zu Positionswechseln im Bett adjustiert wird, erfolgt die Stratifizierung bei allen anderen Bereichen allein anhand des Risikofaktors „Grad der kognitiven Einbußen“.

Bereits in dieser kurzen Beschreibung werden wesentliche Limitierungen des Stratifizierungsansatzes deutlich:

Zum einen unterliegt die Kategorisierung der Risikofaktoren in zwei dichotome Schichten immer einer gewissen Willkür. Der Cut-Off-Punkt, an dem die BewohnerInnen beispielsweise anhand des Abschneidens auf der NBA-Skala „Kognitive und Kommunikative Fähigkeiten“ in die beiden Gruppen geringe vs. erhebliche kognitive Einbußen eingeteilt werden, ist mehr inhaltlich als empirisch begründbar. Genauso gut hätte auch eine andere Aufteilung, etwa in drei Risikogruppen mit geringer, mittlerer und hoher kognitiver Beeinträchtigung, gewählt werden können. Im Grunde müssten hier immer Sensitivität und Spezifität der Kategorisierungen betrachtet werden. Auch Wu, Miller, Lapane, Roy und Mor (2005b) finden, dass unterschiedliche Cut-Off-Werte eine große Auswirkung auf die Reliabilität und Validität dichotomisierter Indikatoren haben können. Der Grad der kognitiven Beeinträchtigung ist in Wirklichkeit ein Kontinuum und lässt sich weniger gut in Schubladen fassen. Da die Daten in der Erfassung bereits kontinuierlich erhoben werden, birgt die nachträgliche Kategorisierung aus statistisch-methodischer Sicht einen unerwünschten Informationsverlust in sich. Bei der multiplen Regression dahingegen ist eine Kategorisierung von stetigen Einflussgrößen nicht notwendig. Der Informationsverlust muss aus verfahrenstechnischen Gründen bei der Stratifizierung in Kauf genommen werden. Eine zu feinteilige Risikogruppenbildung würde schnell zu Unübersichtlichkeit und Intransparenz führen. Für unser Beispiel einer Dreiteilung der kognitiven Einbußen hätten in der Wingenfeld-Systematik beispielsweise schon fünf zu-

sätzliche Indikatoren angelegt werden müssen. Würde man noch weitere Risikofaktoren oder gar Kombinationen (vgl. nächster Absatz) hinzunehmen, wäre eine überschaubare Anzahl schnell überschritten. Vor dem Hintergrund der Stratifizierungsmethode erscheint daher die von Wingefeld et al. (2011) gewählte Dichotomisierung nachvollziehbar und sinnvoll.

Ein weiteres, aus Sicht der Gutachter noch gravierenderes Grundproblem der Stratifizierung ist, dass es bei diesem Ansatz nahezu unmöglich ist, alle relevanten Risiko- bzw. Einflussfaktoren für die einzelnen Outcomes zu berücksichtigen. Eine Beschränkung auf wesentliche (am besten singuläre) Risikofaktoren ist auch hier wiederum der Übersichtlichkeit geschuldet. Wenn beispielsweise unabhängig von der kognitiven Leistungsfähigkeit auch noch das Geschlecht, bestimmte Vorerkrankungen und das Alter eine Rolle spielen würden, müssten theoretisch die Stratifizierungsschichten durch Kombination aller Faktoren gebildet werden. Selbst gesetzt den Fall, dass alle genannten Risikofaktoren dichotomisiert vorliegen und nur zwei Vorerkrankungen relevante Risiken darstellen, ergäben sich im Beispiel für einen einzelnen Indikatorbereich statt zwei Schichten (für die kognitiven Einbußen) beim Zutreffen aller kombinatorischen Möglichkeiten dann schon zehn Schichten, für die Indikatoren definiert werden müssten (vgl. Mukamel et al., 2008).

Die Durchsicht der Literatur lässt in der Gesamtschau den Schluss zu, dass multifaktorielle Einflüsse auf Outcome-Indikatoren eher die Regel als die Ausnahme zu sein scheinen. An dieser Stelle sollen beispielhaft für ausgesuchte Indikatorbereiche, bei denen Ähnlichkeiten zu den Wingefeld-Indikatoren bestehen, ausgewählte Befunde dargestellt werden. Es geht uns an dieser Stelle insbesondere auch darum deutlich zu machen, dass neben der kognitiven Leistungsfähigkeit auch noch andere Faktoren eine Rolle spielen. Da in den einzelnen Studien jeweils nur eine unterschiedliche und eingeschränkte Auswahl von Risikofaktoren einbezogen wurde, darf aus dem Fehlen bestimmter Faktoren in der Aufzählung nicht auf deren Irrelevanz geschlossen werden:

- Für Qualitätsindikator **Mobilitätsverlust** wurden in der Studie von Berlowitz et al. (2005) folgende Risikofaktoren ermittelt: Alter, geringerer ADL-Score („Activities of daily living“/Alltagsverrichtungen), Dekubitus der Stufe 2 oder höher, Vorhandensein von Verhaltensauffälligkeiten und terminale Erkrankungen. Es sei an dieser Stelle darauf hingewiesen, dass die kognitive Leistungsfähigkeit von Berlowitz et al. (2005) nicht in die Modellbildung einbezogen werden konnte.
- Mukamel et al. (2004) untersuchten u. a. Risikofaktoren für den Qualitätsindikator „funktionaler Status“, der als Veränderung der Beeinträchtigung bei **Alltagsverrichtungen** (ADL) innerhalb einer Zeitspanne von drei Monaten definiert war. Als Risikofaktoren für einen Verlust an funktionalem Status wurden neben der kognitiven Beeinträchtigung zum Beispiel noch ein hohes Alter oder die Lebenssituation (alleine zu leben und nicht mit Partner oder Verwandtem) beschrieben. Daneben spielten auch der oftmals im Kontext von Demenzerkrankungen zu beobachtende gesteigerte Bewegungsdrang, Blasen- und Stuhlinkontinenz sowie Krebserkrankungen eine Rolle.
- Als Risikofaktoren für die **Dekubitusentstehung** identifizieren Berlowitz et al. (2005) ähnlich wie auch Bergquist-Beringer und Gajewski (2011) unter anderem das Alter, terminale Erkrankungen, Harnwegsinfektionen, Wundgeschwüre, Fremdabhängigkeit bei Mobilität, Toilettengängen und Lagerung sowie dem Vorliegen eines Dekubitus der Stufe 1. Mukamel et al. (2008) benennen in einer anderen Studie für bestimmte

Bewohnergruppen altersübergreifend noch weibliches Geschlecht, Bettlägerigkeit und Gewichtsverlust als Dekubitus-Risiken. Auf die Bedeutung der Faktoren Mobilitätseinschränkungen und Funktionsverlust verweisen Phillips, Shen, Chen und Sherman (2007) sowie Bostick, Rantz, Flesner und Riggs (2006). Wenngleich es bei der Risikoadjustierung in erster Linie um personenbezogene Merkmale geht, sei darauf hingewiesen, dass auf Struktur- bzw. Prozessebene ferner auch Variablen wie der Anteil an ausgebildeten Pflegefachkräften, Personalschlüssel oder das Vorliegen eines professionellen „Pflegeassessments“ einen nachweislichen Einfluss auf die Dekubitusentstehung haben (Popp, Profener, Stappenbeck, Reintjes, & Weber, 2006).

- Im Hinblick auf den Indikator **Stürze** bestätigen Nazir et al. (2012) zunächst die zentrale Rolle des Einflusses von kognitive Beeinträchtigung. Interessanterweise war in dieser Studie die Wahrscheinlichkeit für neue Stürze bei moderater kognitiver Beeinträchtigung am höchsten, während sie für geringe und hohe Beeinträchtigungsgrade vergleichsweise niedriger ausfiel. Letzteres steht mit eingeschränkter Mobilität bei hoher Beeinträchtigung in Zusammenhang. Neben der Tatsache, dass hier eher eine dreifach gestufte Schichtung der Risikogruppen nahegelegt wird, werden auf Basis einer logistischen Regression noch weitere assoziierte Risikofaktoren für eine höhere Sturzwahrscheinlichkeit beschrieben, so z. B. die Verwendung von Psychopharmaka, Vorhandensein von Verhaltensauffälligkeiten, gesteigertem Bewegungsdrang („wandering“) und größere Selbständigkeit bei Alltagsverrichtungen (gemessen als ADL). Geringer war das Risiko bei weiblichem Geschlecht und Personen mit eingeschränkter Mobilität bzw. Hemiplegie. Für Depressivität oder Gurtfixierungen war kein Einfluss feststellbar.
- Viele Autoren schlagen den Body-Mass-Index (BMI) zur Risikokontrolle im Rahmen des Qualitätsindikatorbereichs **Mangelernährung bzw. unbeabsichtigter Gewichtsverlust** vor (Bartholomeyczik et al., 2010; Olschewski, 2009). Der Indikator Mangelernährung erweist sich in der Studie von Bartholomeyczik et al. (2010) eher weniger als altersabhängig. Erst bei Bewohnern und Bewohnerinnen über 80 Jahren konnte ein Zusammenhang ermittelt werden. Das Risiko war hier darüber hinaus noch höher, wenn die BewohnerInnen Unterstützung beim Essen und Trinken benötigen, ebenfalls bei Menschen mit Demenzerkrankungen (Bartholomeyczik et al., 2010).
- Hjaltadottir, Ekwall, Nyberg und Hallberg (2012) fanden in einer Studie in Island zu mehreren Qualitätsindikatoren auf Basis eines Minimum Data Sets (MDS), dass die mit der CHESS-Skala gemessene Gesundheitsinstabilität ein signifikanter Prädiktor für 16 Qualitätsindikatoren war: Gesundheitsinstabilität stand beispielsweise mit dem Risiko für **Gurtfixierungen** in Zusammenhang. Bei sieben Qualitätsindikatoren stieg das Risiko mit geringerem sozialem Aktivitätsniveau der BewohnerInnen.

Im Bereich Gurtfixierungen werden von Mukamel et al. (2008) neben dem Alter noch die Faktoren eingeschränkte Verständigungs- und Rezeptionsfähigkeit, Verhaltensauffälligkeiten, eingeschränkte Mobilität im Bett sowie das Vorliegen von Stürzen im letzten halben Jahr als signifikante Risiken beschrieben.

- Bzgl. eines Indikators im Bereich **Verhaltensauffälligkeiten** findet die Studie von Berlowitz et al. (2005) die Risikofaktoren Alter, das Vorliegen dementieller Erkrankungen mit und ohne Psychose, Schizophrenie, psychotische Episoden sowie Unterstützungsbedarf beim Essen und bei der Mobilität.

- Bezüglich des Indikatorbereichs **Schmerz** werden von Mukamel et al. (2008) neben einem Zusammenhang mit dem Alter bei über 75jährigen Bewohnern und Bewohnerinnen u. a. noch geschlechtsspezifische Unterschiede sowie mangelnde Verständigungsfähigkeit, Depressivität, Stürze im letzten halben Jahr, Wundliegeneschwüre aller Stufen sowie kognitive Fähigkeiten als Risiken gefunden.
- Vahakangas, Noro, Finne-Soveri und Bjorkgren (2008) finden in ihrer Studie in Finnland, dass kognitive Beeinträchtigung einen signifikanten Einfluss auf verschiedene Qualitätsindikatoren hat, u. a. Dekubitusentstehung, Alltagsverrichtungen und Aktivitätsniveau. Ebenso haben aber beispielsweise auch Merkmale wie Depressivität oder physische Leistungsfähigkeit eine Rolle gespielt.
- Mor (2005) regt an, dass allgemein auch die Adäquatheit bzw. das Ausmaß der informellen Unterstützung der BewohnerInnen mit in die Risikoadjustierungsmodelle aufgenommen werden sollte, schließlich stehe nicht allen Bewohnern bzw. Bewohnerinnen eine ähnliche familiäre und soziale Unterstützung zur Verfügung.

Wie aus der Übersicht deutlich wird, liegen auch für einige Indikatorbereiche, für die Wingenfeld et al. (2011) auf eine Risikoadjustierung verzichten, Hinweise auf ggf. relevante Risikofaktoren vor. Dies trifft auf die Indikatorenbereiche Schmerz und Einsatz von Gurtfixierungen zu, wobei in letzterem Bereich wie einleitend beschrieben auch inhaltlich-fachliche Überlegungen den Verzicht auf eine Risikoadjustierung rechtfertigen könnten (Mor et al., 2011). Die vorliegenden empirischen Befunde können allerdings auch nicht einfach ignoriert werden. In Bezug auf den Bereich Schmerz muss an dieser Stelle darauf hingewiesen werden, dass Wingenfeld et al. (2011) den zugehörigen Indikator spezifisch als „Schmerzmanagement“ definiert haben, was in der vorliegenden Literatur allenfalls eine vage Entsprechung findet (s. Kapitel 4.1). Für die Wingenfeld-Indikatoren „Gestaltung des Alltagslebens und sozialer Kontakte“ sowie „Integrationsgespräch nach dem Heimeinzug“, für die ebenfalls keine Adjustierung vorgenommen wird, können auch in der Literatur keine gegenläufigen Hinweise gefunden werden. Allerdings handelt es sich auch hier um eher „exotische“ Indikatoroperationalisierungen, für die sich in anderen Studien nur schwer vergleichbare Definitionen ausmachen lassen.

In den Indikatorbereichen, in denen Wingenfeld et al. (2011) eine Risikoadjustierung vornehmen, fokussieren sie sich mit einer Ausnahme nahezu ausschließlich auf den Risikofaktor „kognitive Einbußen“ (s. Tabelle 8). Im Sinne einer pragmatischen Umsetzung des Stratifizierungsansatzes erscheint dies grundsätzlich nachvollziehbar. Wie gesagt, ist bei dieser Herangehensweise eine Beschränkung auf wesentliche Risikofaktoren unumgänglich. Für die Verwendung der „kognitiven Einbußen“ spricht, dass diese zumindest einer ganz zentralen Anforderungen an potentielle Risikofaktoren entspricht: Essentiell ist, dass jedes Risikomodell zwingend ein unvollständiges Prognosemodell sein muss. Es dürfen im Idealfall nur Faktoren berücksichtigt werden, die nicht direkt von den Anbietern beeinflusst werden können (vgl. Mukamel et al., 2004). Dies betrifft wie bereits eingangs beschrieben natürlich einerseits alle Prozessvariablen, kann aber auch Merkmale betreffen, die man auf den ersten Blick eher weniger „im Verdacht“ haben könnte. So weisen beispielsweise Hjaltadottir et al. (2012) in ihrer isländischen Studie darauf hin, dass bei der sehr gebräuchlichen Praxis der Risikoadjustierung über den Grad der Fähigkeit Alltagsverrichtungen (ADL-Scores) zu meistern, die mit einem steigenden Risiko von Dekubitusinzidenz und weiterem Verlust von Grundfähigkei-

ten im täglichen Leben einhergehe, häufig nicht bedacht werde, dass es durchaus Hinweise dafür gebe, dass in schlechten Pflegeeinrichtungen die Wahrscheinlichkeit für einen ADL-Verlust größer ist. Der Risikofaktor „kognitive Beeinträchtigung“ erscheint vor diesem Hintergrund, wenn unwahrscheinliche Szenarien wie ein extremer geistiger Abbau durch völlige Verwahrlosung im Rahmen eines reizarmen deprivierenden Settings einmal außer Acht gelassen werden, deutlich weniger anfällig.

Verbleiben noch die Wingenfeld-Indikatoren „Dekubitusentstehung“ und „Verhaltensauffälligkeiten bei Bewohnern bzw. Bewohnerinnen mit kognitiven Einbußen“, die, was die Risikoadjustierung anbetrifft, ein wenig eine Sonderstellung einnehmen. Bei der Dekubitusentstehung erfolgt die Stratifizierung anhand der Fähigkeit der BewohnerInnen Positionswechsel im Bett vorzunehmen. Faktoren, welche die Mobilität der BewohnerInnen im Bett kennzeichnen, werden auch in der Literatur als Risiken im Zusammenhang mit Dekubitusentstehung gesehen. Von daher erscheint es grundsätzlich sinnvoll, diesen Risikofaktor zur Adjustierung heranzuziehen. Das Problem sehen wir hier eher im Zusammenhang mit der Messmethode zur Erfassung des Risikos. Dies wird weiter unten im Abschnitt „Diskussion und Empfehlungen zu den von Wingenfeld vorgeschlagenen Verfahren zur Bestimmung der Risikofaktoren“ noch näher erörtert.

Für einen Indikator „Verhaltensauffälligkeiten bei Bewohnern mit kognitiven Einbußen“ ist ein Ausschluss von Bewohnern und Bewohnerinnen mit geringen oder keinen kognitiven Einbußen unmittelbar evident. Wie bereits erwähnt stellt sich hier eher die Frage, ob die Operationalisierung eines solchen „Ausschluss-Indikators“ grundsätzlich sinnvoll ist oder nicht (s. Kapitel 4.1). Nichtsdestotrotz ist auch hier das für die Ausschlüsse maßgebliche Messverfahren kritisch zu hinterfragen (siehe unten).

Fazit und Empfehlungen zur Weiterentwicklung des Adjustierungsverfahrens

Ein wesentlicher Vorteil der von Wingenfeld et al. (2011) bei der Risikoadjustierung präferierten stratifizierten Analyse ist deren „Schlichtheit“, Nachvollziehbarkeit und Transparenz, wobei die Systematik sich auch dem Laien schnell erschließt. Die Stratifizierung größtenteils anhand der kognitiven Einbußen vorzunehmen erscheint vor diesem Hintergrund praktikabel. Eine Bildung von Risikogruppen anhand von Pflegestufen wie es z. B. Weibler-Villalobos und Röhrig (2010) vorschlagen, wäre dahingegen nicht zu empfehlen gewesen, zumal hier einerseits die Messgenauigkeit und Vergleichbarkeit der Klassifikation fraglich erscheint sowie andererseits wie beschrieben auch der Konfundierungsgrad mit den Leistungen der stationären Pflegeeinrichtungen deutlich höher erscheint.

Auch wenn der bei Wingenfeld et al. (2011) vorliegende Adjustierungsansatz in sich schlüssig konzipiert erscheint und Beweggründe sowie Systematik durchaus nachvollziehbar sind, können wir ihn an dieser Stelle nicht empfehlen. Der entscheidende Grund ist, dass im Rahmen der Stratifizierung ganz einfach nicht alle praktisch relevanten Risikofaktoren einbezogen werden können. Bereits im vorangegangenen Abschnitt wurde deutlich, dass die Forschungslage in den sich mit Wingenfeld überschneidenden Indikatorbereichen eine Vielzahl weiterer signifikanter Risikofaktoren benennt. Auch wenn nicht alle der Studien explizit auch die kognitive Leistungsfähigkeit miteinbezogen haben, sind zumindest begründete Zweifel angebracht, ob diese wirklich in allen Qualitätsbereichen, bei denen Wingenfeld et al. (2011) über den Grad an kognitiven Einbußen adjustieren, den entscheidenden Anteil der jeweiligen

Outcome-Varianzen aufklärt oder sogar ob teilweise überhaupt einen signifikanten Beitrag geliefert wird. Die empirischen Belege hierzu werden von den Autoren nur unvollständig erbracht, was allerdings in Anbetracht der unterschiedlichen Indikatoroperationalisierungen der teils schwierigen Vergleichbarkeit von Forschungsbefunden geschuldet ist. Ein weiteres statistisches Problem des Stratifizierungsansatzes ist, dass die Messfehleranfälligkeit steigt, wenn in bestimmten Einrichtungen nur eine kleine Zahl von Klienten den Kriterien der einzelnen Schichten entspricht (Arling et al., 2007; Mor, 2006).

In einer spannenden Studie von Mukamel et al. (2008), bei der einfache (einfaktorielle) mit komplexeren (multifaktoriellen) Risikoadjustierungsmethoden verglichen wurden, konnte gezeigt werden, dass sich die Rankings von Einrichtungen bei einer vergleichenden Qualitätsberichterstattung teils deutlich verändern (vgl. auch Nazir et al., 2012), wenn zusätzliche Risiken mit in die Modellbildung aufgenommen werden. Ähnliche Befunde liegen auch bei Konetzka et al. (2013) vor. Eine umfangreichere Risikoadjustierung mit mehr Einflussfaktoren resultierte hier ferner auch noch in höheren Übereinstimmungen verschiedener Outcomemessungen innerhalb von Einrichtungen. D.h. die Qualitätsrankings der einzelnen Indikatoren waren weniger heterogen und spiegelten somit die Gesamtqualität der Einrichtung wohl besser wider. Einfache Modelle mögen in der Umsetzung praktikabler sein, können letztendlich die Wirklichkeit aber eben auch nur deutlich eingeschränkter abbilden. Nach Abraham, MacDonald und Nadzam (2006) adjustieren gute Outcomemessungen immer anhand möglichst vielfältiger Erkrankungen und Faktoren. Mukamel et al. (2008) verweisen ferner darauf, dass durch unzureichende Risikoadjustierung ggf. auch die falschen Marktanreize gesetzt werden, indem beispielsweise Anreize geschaffen werden, Menschen mit dem größten Pflegebedarf nicht aufzunehmen (sog. „cream skimming“ bzw. „Rosinen-Picken“).

Ein weiterentwickeltes Risikoadjustierungsmodell sollte aus Sicht der Gutachter daher immer möglichst umfassend potenzielle Risiken miteinbeziehen. Im Statistikjargon könnte man dies so ausdrücken, dass es sich von der univariaten auf die multivariate Ebene fortentwickeln sollte. In der von uns gesichteten vorwiegend US-amerikanischen, aber auch internationalen Literatur, sind solche multivariaten Ansätze bei der Risikoadjustierung quasi eine Standardprozedur. Ein weitläufig erprobtes und in statistischer Hinsicht gut handhabbares Verfahren stellt aus unserer Sicht etwa die **multiple logistische Regression** dar, die es bei gleichzeitig empfehlenswerter Kontrolle möglicher Multi-Kollinearitäts-Effekte ermöglicht, nahezu beliebig viele Einflussfaktoren mit Richtung (Risikoerhöhung bzw. -verminderung) und gradueller Stärke des Zusammenhangs mit in die Modellbildung aufzunehmen. Dadurch werden multiple Risiken abgebildet (Arling et al., 2007). In einer Erweiterung des Ansatzes (Cox & Wermuth, 1994) könnten theoretisch auch Wechselwirkungen zwischen den Risikofaktoren in die Modellbildung aufgenommen werden. Ein weiterer Vorteil ist, dass hier auch Faktoren miteinbezogen werden können, die ansonsten nur über Ausschlüsse handhabbar gewesen wären (s. Kapitel 4.1.3).

Mit dem Einwand einiger Autoren und Autorinnen, dass mathematische Modellrechnungen auf der Basis multivariater Regression wegen der Kompliziertheit der Interpretation und der fehlenden Transparenz in der Praxis nicht vermittelbar seien (vgl. Weibler-Villalobos & Röhrig, 2010), können wir nicht übereinstimmen. Man sollte sich klarmachen, dass der Transparenzfaktor an dieser Stelle insbesondere für die Anbieter, nicht aber unbedingt die Verbraucherperspektive (s. Kapitel 4.5) von Relevanz ist. So weist etwa Mor (2006) darauf hin, dass bei der Stratifizierung die Anbieter leicht erkennen könnten, welche Personen sich

in welcher Risikogruppe befinden, um sich so etwa spezifisch auf Qualitätsprobleme bei Hoch-Risiko-Klientel fokussieren zu können. Auf der anderen Seite schafft eine zu hohe Transparenz auf Anbieterseite mit Blick auf die Qualitätsberichterstattung auch „Spielraum“ für Datenmanipulationen, wenn beispielsweise Grenzfälle zwischen zwei Risikogruppen so „eingeordnet“ werden, dass sich am Ende dann ein besseres Qualitätsurteil ergibt. Derartige Anreize bestehen insbesondere dann, wenn das Qualitätsbewertungssystem vergütungsba-siert gestaltet wird (O'Connor & Davitt, 2012). Letztendlich erscheint es uns für ein Verfahren zur Outcomemessung, das einen bundesweiten Vergleich von Einrichtungen beinhaltet soll, unumgänglich, dass die Risikoadjustierung von einem unabhängigen Institut erfolgen muss. Da dies nur im Rahmen einer EDV-gestützten Datenerfassung und eines EDV-gestützten Datenaustauschs praktikabel erscheint, wäre es einerseits naheliegend, dass in diese Platt-form auch eine Auswertungsfunktionalität implementiert werden würde, die es den stationä-ren Pflegeeinrichtungen ermöglicht, etwa auch Risikogruppenprofile abzurufen. In diesem Zusammenhang wäre die Komplexität des zugrundeliegenden Algorithmus völlig unerheblich und gleichsam auch das Problem der schwierigen Praxisimplementierung gelöst.

Aus Verbraucherperspektive erscheint die Frage nach der Transparenz der Risikoadjustie-rung aus unserer Sicht eher von untergeordneter Relevanz. Für die Verbrauchersicht sollte die Maßgabe einer möglichst übersichtlichen und komprimierten Informationengabe gelten, etwa analog den Qualitätsurteilen der Stiftung Warentest. Da bei einer Risikoadjustierung auf Basis multipler Regression, bei der individuelle Risikowahrscheinlichkeiten auf Fallebene vorliegen, die Notwendigkeit einer Risikogruppenbildung entfällt, die in ihrer Einteilung insbe-sondere an den Klassengrenzen immer einem gewissen Grad an Willkürlichkeit unterworfen ist, bestünde auch bei den Wingenfeld-Indikatoren an vielen Stellen keine Notwendigkeit zu separaten Indikatoroperationalisierungen. In der Wingenfeld-Systematik könnte man sich auf jeweils einen einzelnen Indikator auf Ebene der in Tabelle 8 aufgeführten Indikatorberei-che beschränken. Dadurch könnte die Anzahl der in dieser Expertise diskutierten Indikatoren um ein Drittel von 15 auf 10 Qualitätsindikatoren verschlankt werden. Natürlich könnte man argumentieren, dass auch für die VerbraucherInnen interessant sein könnte zu wissen, wel-che Einrichtung eine besonders hohe Qualität etwa für die Betreuung von hochgradig demenziellen Menschen aufweist. Dies müsste aber eher über eine hierarchische Modellie-rung erreicht werden (siehe unten) als über separate Qualitätsindikatoren. Die Verbraucher könnten dann beispielsweise schon in einer Vorauswahl auf der Webseite des Qualitätsver-gleichs angeben, dass sie eine auf Demenzkranke spezialisierte Einrichtung suchen. In der Auswertung würden die Ergebnisse dann entsprechend spezifiziert, aber eben keine zusätz-lichen Indikatoren ausgewiesen. Sind die Indikatoren selbst wie bei Wingenfeld et al. (2011) nach erheblichen vs. geringen/ keinen kognitiven Einbußen getrennt, bleibt es weitestgehend der Phantasie der Suchenden überlassen, wie sie dieses Konstrukt für sich bzw. ihren Ange-hörigen definieren. Es ist zumindest diskussionswürdig, dieses Assessment quasi dem „Gut-dünken“ der Laien zu überlassen, wenn man sich schon bei den professionellen Messmetho-den nicht immer ganz sicher sein kann, ob wirklich eine zuverlässige Trennung gelingt (vgl. auch Abschnitt „Diskussion und Empfehlungen zu den von Wingenfeld vorgeschlagenen Ver-fahren zur Bestimmung der Risikofaktoren“).

Wie hat man sich den Ablauf einer multivariaten Risikoadjustierung auf der Basis logistischer Regression (bei binären Indikatoren) oder linearer Regression (bei stetigen Indikatoren) kon-kret vorzustellen? In einem ersten Schritt müssen die Einflussfaktoren, also die potenziellen

Risikofaktoren, identifiziert und erfasst werden. Wie bereits erwähnt, kommen hier keine Prozessmerkmale in Frage, die unmittelbar von den Anbietern beeinflusst werden. Prozessbezogene Messungen kommen in der Regel übrigens sehr gut ohne Risikoadjustierung aus, wobei allerdings dann oft die konkrete Beziehung zu den Outcomes fraglich ist (Naylor, 2007) und hohe Prozessqualität bekanntermaßen nicht automatisch zu guter Ergebnisqualität führen muss (Donabedian, 1980). Bei der Adjustierung durch Regression wird die Wahrscheinlichkeit für das Eintreffen des Outcomes ausschließlich als Funktion persönlicher Merkmale der BewohnerInnen abgebildet. Neben basalen soziographischen Merkmalen wie Alter, Geschlecht, Familienstand, ethnischer Zugehörigkeit o. ä. sind hier u. a. der Schweregrad von Begleit- und Vorerkrankungen relevant (vgl. Stemmer, 2005). Vorsicht ist geboten bei Risikomaßen, die zumindest indirekt auch mit der Qualität der Einrichtungen in Zusammenhang stehen. So waren etwa Phillips et al. (2007) im Zusammenhang mit dem Indikator Dekubitus davor, diesen über den Grad der Bettlägerigkeit zu adjustieren, da davon auszugehen sei, dass sich schlechte Einrichtungen auch durch einen zu geringen Aufwand an Mobilisierungsmaßnahmen kennzeichnen könnten (vgl. Mor, 2005; Phillips, Shen, Chen, & Sherman, 2007). In diesem Fall wäre dann der Bewohnerzustand mit der Anbieterleistung konfundiert. Allerdings muss auch bedacht werden, dass die Liste potentieller Risikofaktoren schnell sehr kurz werden würde, wenn man alle Faktoren, die ggf. irgendwie auch mit von Anbieterseite beeinflusst werden könnten, grundsätzlich ausschließen würde (Arling et al., 2005). Phillips et al. (2007) präferieren in diesem Zusammenhang ähnlich wie Wingenfeld et al. (2011) trotzdem eher personenbezogene Maße wie die kognitive Leistungsfähigkeit. Eine Randnotiz wäre hier aber wiederum, dass ausgerechnet beim Dekubitus-Indikator Wingenfeld et al. (2011) die kognitiven Einbußen gar nicht verwenden. Auch dies unterstreicht wieder die Wichtigkeit einer multifaktoriellen Betrachtung. Aber auch die Qualität des mehrfaktoriellen Ansatzes steht und fällt letztendlich damit, in welcher Breite potenzielle Einflüsse erhoben werden und damit überhaupt in die Auswertung einfließen können. Ggf. müssten auch vor dem Hintergrund der Outcomemessung insgesamt doch mehr Daten erhoben werden als von Wingenfeld et al. (2011) ursprünglich vorgesehen. Prinzipiell sollte dies im Rahmen der bereits vorgeschlagenen EDV-technischen Datenerfassung möglich sein. Sofern hinreichende Validität und Messgenauigkeit bestätigt werden können, kämen hierfür ggf. durchaus auch noch weitere Skalen aus dem Neuen Begutachtungsinstrument (NBA) (Wingenfeld et al., 2008) in Betracht. Um spezifischere Hinweise zu relevanten Risikofaktoren und damit auch zum notwendigen Erhebungsumfang zu erhalten, wäre es insgesamt in dieser Hinsicht empfehlenswert, zu den einzelnen Indikatorbereichen nochmals ein systematisches Literaturreview zu erstellen.

Im nächsten Schritt der Risikoadjustierung im Rahmen des Regressionsansatzes wird für jede Einrichtung die durchschnittliche Rate für den jeweils untersuchten Qualitätsindikator berechnet („observed QM rate“). Aus den Ergebnissen der multiplen Regression kann dann für jeden Fall die Wahrscheinlichkeit bestimmt werden, dass das entsprechende Outcome eintritt oder nicht (bei der logistischen Regression). Daraus kann dann eine einrichtungsbezogene Eintrittswahrscheinlichkeit bestimmt werden. Die durchschnittliche Rate auf dem jeweiligen Qualitätsindikator wird dann durch die mittlere Eintrittswahrscheinlichkeit der Einrichtungen dividiert. Der Quotient kann dann wie folgt interpretiert werden: Ein Wert von „1“ heißt, dass die Einrichtung risikoadjustiert genau im Durchschnitt liegt. Bei „2“ würden doppelt so viele Outcome-Ereignisse wie z. B. Stürze vorliegen als es risikoadjustiert zu erwarten gewesen wäre, bei „0,5“ würden nur halb so viele Ereignisse auftreten (vgl. Arling et al., 2007).

Kleinere Werte als 1 kennzeichnen bei Outcomes wie im Beispiel Stürze also positive Qualität, größere Werte als 1 dahingegen eine negative Qualität. Es ist natürlich leicht möglich, diese Werte auch auf eine intuitivere Skalenform wie beispielsweise ein Fünf-Sterne-Ranking ähnlich einer Hotelbewertung oder einer Schulnotenskala ähnlich Stiftung Warentest zu transformieren.

Obwohl wir diesen Ansatz favorisieren, muss an dieser Stelle auch darauf hingewiesen werden, dass auch die Anwendung klassischer multipler Regressionsmodelle gewisse Nachteile aufweist: So können für Einrichtungen mit kleinen Fallzahlen ungenaue und instabile Schätzungen resultieren, die dann zu falschen Qualitätseinschätzungen führen können (AQUA 2010). In der Literatur wird daher in der Regel eine Mindeststichprobengröße empfohlen. In der Regel liegt diese bei einem n von 20 (vgl. Mor et al., 2011; Sangl et al., 2005). Andere Autoren sind etwas konservativer und setzen den Schwellenwert für die Mindeststichprobengröße bei 30 an (etwa Li, Cai, Mukamel, & Glance, 2010; Zinn, Spector, Hsieh, & Mukamel, 2005). Wird diese unterschritten, sollten die entsprechenden Qualitätsindikatoren nicht berechnet werden. Li et al. (2010) zeigen eindrucksvoll, dass mitunter eine „Volumen-Outcome-Beziehung“ bestehen kann. In ihrer Studie zeigte sich bei sonst vergleichbaren Einrichtungen mit größeren Fallzahlen ein geringerer Abbau des Funktionsniveaus der BewohnerInnen als bei kleineren Einrichtungen. Es ist allerdings nicht klar, inwieweit diese Befunde auf andere Indikatoren generalisiert werden können.

Ein weiteres Problem, insbesondere im Hinblick auf eine vergleichende Qualitätsberichterstattung, ist, dass zum einen einige Risikofaktoren gleichzeitig auch Outcomekriterien sein können (Mukamel et al., 2008), z. B. limitierte Fähigkeiten in Bezug auf alltägliche Verrichtungen (ADL) als Risiko für Dekubitusentstehung. Zum anderen können einzelne Outcomes auch untereinander korreliert sein. So gehen nach Castle und Ferguson (2010) etwa Dekubitalgeschwüre auch mit Mangelernährung einher. Wenn die Ergebnismaße separat betrachtet werden, kann es dann sozusagen zu einer Doppel- bzw. Mehrfachbestrafung von qualitativ schlechten Pflegeeinrichtungen kommen, während Einrichtungen mit guter Qualität überproportional belohnt werden (vgl. Mukamel et al., 2008).

Vor dem Hintergrund derartiger Fragestellungen, vornehmlich aber um neben dem Beitrag der personenbezogenen Merkmale auch den einrichtungsbezogenen Beitrag zur Varianzaufklärung innerhalb der Outcomes, oder anders ausgedrückt die sogenannte Mehr-Ebenen-Struktur der Daten (z. B. BewohnerInnen innerhalb von bestimmten Einrichtungen) abbilden zu können, werden aktuell verstärkt auch hierarchische Modelle zur Risikoadjustierung diskutiert. Diese sind in u. a. auch als Multi-Ebenen-Modelle (HLM) (Arling et al., 2007), Bayessche hierarchische Netzwerke (Gajewski, Mahnken, & Dunton, 2008) oder einfach nur Bayessche Netzwerke (Goodson, Jang, & Rantz, 2008) bekannt. Nach Gajewski et al. (2008) sind hierarchische Modelle besonders gut für Probleme geeignet, die aus kleinen Stichproben und der damit verbundenen Messungenauigkeit resultieren. Es handelt sich ebenfalls um eine multivariate Methode, die aber deutlich komplexere Beziehungen zwischen einzelnen Variablen modellieren kann (vgl. Wu, Miller, Lapane, Roy, & Mor, 2005a) als die herkömmlichen Regressionsmodelle es zulassen, die klassischerweise neben den abhängigen Variablen bzw. Outcomes nur eine Ebene von Einflussfaktoren kennen. Hierarchische Modelle berücksichtigen auch Verschachtelungen: So können analog zu verschiedenen Schulklassen, die beispielsweise aufgrund unterschiedlich strenger Lehrer ganz andere Notenschnitte in denselben Schulfächern haben können, auch in stationären Pflegeeinrichtungen relevant

sein, z. B. im Hinblick darauf welche Bewohnercharakteristika zu welchen Einrichtungen „passen“. Hierbei können mehrere Ebenen einbezogen werden, z. B. Einrichtung und BewohnerInnen (Zwei-Ebenen-Modell) oder Einrichtung, Wohneinheit (bzw. Station) und BewohnerInnen (Drei-Ebenen-Modell). Dabei kann auch geprüft werden, ob die einrichtungsbezogenen Faktoren, zu denen auch Prozessmerkmale wie die zeitlichen Ressourcen des Pflegepersonals zählen können, eine stärkere Bedeutung als die personenbezogenen Unterschiede haben. Auf diese Weise lässt sich das immer wieder angesprochene Problem der Trennung dieser beiden Ebenen im Zuge der Risikoadjustierung statistisch-analytisch erhehlen. Durch Mehrebenen-Analyse können ferner noch Vorliegen und Ausmaß von potenziellen Beurteilungsfehlern („ascertainment bias“) aufgedeckt werden, die auf einen unterschiedlichen Trainingsstand bzw. divergierende diagnostische Fähigkeiten der Fachkräfte in den Einrichtungen zurückgehen (Wu et al., 2005a). Wie die Modellstudie von Horn et al. (2005) zeigt, kann hier durchaus aber auch ein Ergebnis sein, dass für bestimmte Outcomebereiche eine Mehr-Ebenen-Betrachtung nicht von Relevanz ist. In diesen Fällen wäre dann die klassische Modellierung völlig ausreichend.

Obwohl hierarchische Modelle gemeinhin als der Goldstandard (Gajewski et al., 2008; Arling et al., 2007) bei der Risikoadjustierung angesehen werden, können wir diese im Kontext der Weiterentwicklung der Adjustierung der Wingenfeld-Indikatoren aktuell noch nicht empfehlen. Auch mit Rücksicht auf die anspruchsvollen methodischen Anforderungen wäre dies in Anbetracht des vorliegenden Standes aus Sicht der Gutachter ein Schritt des Guten zu viel. Da es sinnbildlich gesehen gerne zum Stolpern kommt, wenn zwei Schritte auf einmal genommen werden sollen, wäre der nächste logische Entwicklungsschritt aus unserer Sicht zunächst die Einbeziehung zusätzlicher Risikofaktoren im Rahmen einer multiplen Regression. Dazu zählt auch die Klärung, welche Faktoren einbezogen werden können und wie diese praxistauglich erhoben werden können. Die Mehr-Ebenen-Struktur könnte dann in zukünftigen Entwicklungsschritten immer noch ergänzt werden.

Diskussion und Empfehlungen zu den von Wingenfeld vorgeschlagenen Verfahren zur Bestimmung der Risikofaktoren

Wenn man zur Weiterentwicklung der Risikoadjustierung der Wingenfeld-Indikatoren nicht den von uns vorgeschlagenen Weg hin zu einem multivariaten Adjustierungsverfahren einschlagen möchte, sehen wir auch innerhalb des bisherigen Vorgehens noch Optimierungspotenziale. Diese betreffen die im Rahmen von Wingenfeld et al. (2011) zur Stratifizierung verwendeten Methoden, mit denen die Ausprägung der von den Autoren ausgewählten Risikofaktoren bestimmt werden. Dabei greifen Wingenfeld et al. (2011) ausschließlich auf zwei Messverfahren bzw. Skalen zurück, anhand derer jeweils Personen vergleichsweise niedriger und hoher Ausprägung unterschieden werden. Die mit Abstand am häufigsten eingesetzte Skala zielt auf den Grad der kognitiven Einbußen ab. Es werden Personen unterschieden, die keine bzw. nur eine geringe kognitive Beeinträchtigung aufweisen und solche mit mindestens erheblicher kognitiver Funktionseinschränkung (vgl. Tabelle 8). Die zweite verwendete Skala soll vor dem Hintergrund des Indikators Dekubitusentstehung zwischen Personen mit hohem und niedrigem Dekubitusrisiko unterscheiden. Dass wir diese Auswahl auf dem Hintergrund des Stratifizierungsansatzes in inhaltlicher Hinsicht nachvollziehen können, wurde bereits dargelegt. In beiden Fällen wird auf das Neue Begutachtungsassessment (NBA) (Wingenfeld et al., 2008) zurückgegriffen. Es ist unmittelbar naheliegend, dass das Gelingen

der Risikoadjustierung, neben der Validität der jeweils dem NBA entnommenen Skalen, auch von deren Trennschärfe und Messgenauigkeit abhängt. Diese Frage ist völlig losgelöst von der inhaltlichen Plausibilität der zur Adjustierung herangezogenen Risikofaktoren zu diskutieren. In den vorausgehenden Überlegungen haben wir z. B. schon darauf hingewiesen, dass wir eine Risikoadjustierung über Pflegestufen aufgrund ihrer zweifelhaften Messgenauigkeit für nicht zielführend halten.

Das zur Bestimmung des Risikofaktors „Kognitive Einbußen“ verwendete NBA-Modul „Kognitive und kommunikative Fähigkeiten“ (NBA-Modul 2) beinhaltet 11 Items, anhand derer unter anderem die zeitliche und örtliche Orientierung oder das Gedächtnis anhand eines vierstufigen Ratings von „unbeeinträchtigt“ über „größtenteils vorhanden“, „in geringen Maße vorhanden“ sowie „nicht vorhanden“ einzuschätzen sind. Die Risikogruppenbildung beim Indikator Dekubitusentstehung erfolgt lediglich anhand des Einzelitems „Positionswechsel im Bett“ aus dem Modul NBA Mobilität, das auf einer vierstufigen Skala mit den Ausprägungen „selbstständig“, „überwiegend selbstständig“, „überwiegend unselbstständig“ und „unselbstständig“ zu bewerten ist. Zu keiner der beiden Skalen (bzw. Items) finden sich bislang befriedigende Belege zu den zentralen Testgütekriterien. Zwar wird der NBA-Skala „Kognitive und kommunikative Fähigkeiten“ mit einem gewichteten Kappa von 0,71 durchaus eine veritable Interrater-Reliabilität bescheinigt (Windeler et al., 2011), deren Validität wird aber durchaus kritisch gesehen. So gibt es deutliche Hinweise auf eine mangelhafte Konstruktvalidität. Die Ergebnisse von Rasch-Analysen deuten darauf hin, dass innerhalb des NBA Moduls 2 die Bereiche Kognition und Kommunikation eher eigenständigen Charakter haben und dementsprechend nicht in eine Gesamtskala aufsummiert werden dürften (Bensch, 2013; Bensch, 2011). Für den Gesamtskalenwert des Moduls bleibt somit in der Endkonsequenz unklar, was eigentlich genau gemessen wird. Handelt es beispielsweise eher um kognitive oder eher um kommunikative Beeinträchtigungen? Im Grunde genommen ist also gar nicht wirklich klar, welcher Risikofaktor mit dieser Skala hier überhaupt adjustiert wird.

Kritisch zu bewerten ist ebenfalls, dass bei der Risikogruppenbildung im Bereich Dekubitus nur auf ein einzelnes Item zurückgegriffen wird. Übereinstimmend etwa mit analogen Empfehlungen von Bates-Jensen und MacLean (2007) ist hier inhaltlich durchaus nachvollziehbar, mobilitätsbezogene Items wie die Fähigkeit zu Positionswechseln im Bett zur Adjustierung des Dekubitusrisikos heranzuziehen. Einzelitems haben aber testtheoretisch in der Regel immer eine deutlich geringere Reliabilität als umfangreichere Skalen, die den relevanten Sachverhalt mit zusätzlichen Items nochmals aus weiteren Blickwinkeln beleuchten. Das Reliabilitätskonzept der internen Konsistenz basiert beispielsweise im Grunde allein auf diesem Prinzip. Da die Hauptgütekriterien voneinander abhängig sind, ist bei fehlender Reliabilität auch hier die Validität der Messung sehr fraglich (vgl. Lienert & Raatz, 1998). Auch Bates-Jensen und MacLean (2007) sprechen sich dafür aus, das Dekubitusrisiko mit Hilfe einer standardisierten Skala zu adjustieren.

Selbst wenn man annehmen würde, dass die beiden von Wingefeld et al. (2011) verwendeten Skalen eine hinreichende Validität besäßen, wäre noch offen, ob sie wirklich eine zuverlässige Zuweisung bzw. Aufteilung der BewohnerInnen in die jeweils von den Autoren definierten Hoch- und Niedrigrisikogruppen ermöglichen. Dies ist eine wichtige Frage, da die Risikogruppen auch für die Operationalisierung der Qualitätsindikatoren verwendet werden. Um dies beurteilen zu können, müsste auch die Trennschärfe der Skalen in den Blick genommen werden. Eine maximale Trennschärfe kann nur erreicht werden, wenn eine mittlere

Itemschwierigkeit und eine hinreichende Streuung vorliegen. Angaben zu den relevanten Größen sind aber nicht ausgewiesen. Leider ist darüber hinaus nicht einmal der genaue Zuweisungsmodus von Bewohnern bzw. Bewohnerinnen in die jeweiligen Risikogruppen offengelegt. Es kann nur implizit aus den vorgelegten Auswertungsdaten gefolgert werden, dass vermutlich die Personen mit Werten größer bzw. gleich der NBA-Kategorie „überwiegend bzw. größtenteils vorhanden“ in die Risikogruppe mit dem größeren Ausprägungsgrad fallen.

Den Gutachtern erscheint es zwar insgesamt durchaus nachvollziehbar, dass zur Maximierung der Erhebungsökonomie Skalen bzw. Module des NBA herangezogen werden, in Anbetracht der essentiellen Bedeutung der Risikoadjustierung für die Definition, Interpretation und Vergleichbarkeit der Wingenfeld-Indikatoren wäre es hier ggf. aber doch ratsamer, zumindest in diesem Zusammenhang auf bereits ausreichend erprobte und abgesicherte Verfahren zurückzugreifen. So verweisen etwa auch Mor, Intrator, Unruh und Cai (2011) auf die große Anzahl an standardisierten Instrumenten für das physische, kognitive, emotionale, soziale und verhaltensseitige Funktionsniveau, die für die Erhebung von Risikofaktoren in Frage kommen. Wir möchten an dieser Stelle daher Alternativen für die beiden von Wingenfeld et al. (2011) verwendeten Skalen diskutieren.

Für die Bestimmung des Grades an kognitiver Beeinträchtigung wird in der Literatur vornehmlich auf die zwei Verfahren verwiesen: Die „Cognitive Performance Scale“ (CPS) sowie den „Mini-Mental-Status-Test“ (MMST) (Folstein, Folstein, & McHugh, 1975). Für beide Verfahren ist Messgenauigkeit und Validität hinreichend belegt. Auch die im deutschen Sprachraum weniger gängige CPS-Skala weist einer Studie zu Folge neben veritabler Konstruktvalidität eine Sensitivität von 94 %, eine Spezifität von 96 % und eine diagnostische Genauigkeit von 96 % auf (Capezuti, Boltz, Renz, Hoffman, & Norman, 2006). In Bezug auf die CPS-Skala verweisen Phillips et al. (2007) des Weiteren darauf, dass diese ebenfalls gegenüber deutlich umfangreicheren Instrumenten validiert wurde. Ferner sind beide Verfahren auch für Pflegefachkräfte verständlich und leicht anwendbar. Eine valide Anwendung setzt aber in der Anwendung geschulte Fachkräfte voraus (vgl. Nazir, Arling, Perkins, & Boustani, 2011). Paquay et al. (2007) stellen in einer in diesem Zusammenhang hoch relevanten Studie die diagnostische Leistungsfähigkeit der beiden Verfahren gegenüber. Insgesamt zeigt die Studie, dass beide Verfahren annähernd gleich gut die kognitive Beeinträchtigung von Pflegeheimbewohnern und -bewohnerinnen ermitteln können. Der MMST zeigte lediglich eine leicht höhere Sensitivität, während die CPS-Skala eine leicht höhere Spezifität aufwies. Paquay et al. (2007) sprechen in ihrer Studie vor dem Hintergrund von Spezifität und Sensitivität ferner auch Empfehlungen für optimierte Cut-Off-Werte aus, die teilweise von früheren Vorgaben in der Literatur abweichen. Diese könnten auch im Rahmen der Risikoadjustierung von Belang sein. Welches Verfahren kann nun empfohlen werden? In den US-amerikanischen Studien wird die CPS-Skala nahezu standardmäßig eingesetzt, der MMST spielt hier eher eine untergeordnete Rolle. Dies bedeutet aber nicht, dass der MMST ein praxisuntaugliches Verfahren wäre, sondern dürfte vielmehr damit zu tun haben, dass die CPS-Skala ein integrierter Bestandteil von bestehenden Datenerhebungssystemen, etwa dem „Minimum Data Set“ (MDS) oder dem „Resident Assessment Instrument“ (RAI) ist, deren Daten für Forschungszwecke leicht verfügbar sind. Da die entsprechenden Daten in den USA standardmäßig erhoben werden, empfehlen Paquay et al. (2007) die CPS-Skala, da hier kein zusätzlicher Erhebungsaufwand mit einem weiteren Testverfahren notwendig ist. Im Hinblick auf den Einsatz im deutschsprachigen Raum gilt dieses Argument freilich nicht. Zwar steht

die CPS etwa auch über die deutsche Version des RAI zur Verfügung, hat aber natürlich nicht annähernd die gleiche Verbreitung. Von daher wäre hier eher der Mini-Mental-Status-Test zu empfehlen, der hier einen höheren Bekanntheitsgrad genießt. Grundsätzlich wäre natürlich anzustreben, dass das Risikofaktorenassessment möglichst „nahtlos“ in die Dokumentation der Qualitätsindikatoren integriert wird. Wie bereits erwähnt, sehen wir das nur über die Nutzung einheitlicher Softwarelösung als realistisch an. Inwieweit ein Minimum Data Set auf das deutsche Pflegesystem übertragbar ist, bedarf noch näherer Abwägung (s. etwa Faust, 2004).

Zur Risikoadjustierung bei Dekubitus wird in der vorliegenden Literatur häufig die Braden-Skala verwendet (Großschädl & Lohrmann, 2011). Diese besteht aus sechs Items, die sensorisches Empfindungsvermögen, Feuchtigkeit der Haut, Aktivität, Mobilität, Ernährung sowie Krafteinwirkung in einer i. d. R. vierstufigen Skalen erfassen. Anhand der Summenwerte lässt sich das Dekubitusrisiko von maximalem Dekubitusrisiko (6 Punkte) bis zu keinem Dekubitusrisiko (23 Punkte) einstufen (Kottner et al., 2011). Nach Tannen, Bours, Halfens und Dassen (2006) konnte die prädiktive Validität und Reliabilität der Braden-Skala in unterschiedlichen Settings unter Beweis gestellt werden. Nach Kottner et al. (2011) liegen in Pflegeheimsettings zumindest ab Dekubitusgeschwüren der Stufe 2 hinreichende Interrater-Reliabilitäten von ca. 0,7 vor. Lediglich im Bereich der ambulanten Pflege sei die Reliabilität der Skala nicht gesichert (Bergquist-Beringer & Gajewski, 2011). Tannen et al. (2006) haben zur Risikogruppenbildung anhand der Braden-Skala einen Cut-Off von ≤ 20 verwendet. Es muss aber auch erwähnt werden, dass die Braden-Skala in der Literatur teilweise auch kritisch diskutiert wird. So wird in einigen Quellen berichtet, dass die Braden-Skala dazu neige, dass Dekubitusrisiko zu überschätzen (DNQP, 2010a).

Ohne die Verfahren an dieser Stelle in allen Facetten diskutieren zu können, sollte aus den Ausführungen deutlich geworden sein, dass es auch innerhalb der von Wingenfeld et al. (2011) vorgeschlagenen Adjustierungssystematik noch Optimierungspotenziale bestehen. In diesem Zusammenhang muss abschließend noch darauf hingewiesen werden, dass bei den von Wingenfeld et al. (2011) verwendeten Adjustierungsskalen auch der genaue Zuweisungsmodus nicht explizit offengelegt ist. D.h. es nicht transparent, welcher Skalen- oder Testwert erreicht werden muss, damit ein Bewohner bzw. eine Bewohnerin in eine bestimmte Risikogruppe eingeordnet werden kann. Die vorliegenden Zahlen deuten aber darauf hin, dass dies wohl anhand der inhaltlichen Bedeutung der Skalenwerte erfolgt ist, wahrscheinlich in der Form, dass alle Personen mit Werten größer bzw. gleich der NBA-Kategorie „überwiegend bzw. größtenteils vorhanden“ automatisch in die Hochrisikogruppe fallen. Ohne Kenntnis des Zuweisungsmodus können Sensitivität und Spezifität der Zuweisungen nicht abschließend beurteilt werden.

Als Alternativen bei der Erfassung der Risiken kämen aus unserer Sicht u. a. im Bereich „kognitive Einbußen“ etwa der Mini-Mental-Status-Test sowie im Bereich Dekubitus-Risiko ggf. die Braden-Skala in Betracht. Es müssten letztendlich nicht zwingend diese Verfahren sein. Im Bereich der kognitiven Beeinträchtigungen werden in der Literatur auch noch andere Verfahren wie u. a. das „Brief Interview für Mental Status“ (BIMS) (Saliba et al., 2012; Chodosh et al., 2008) oder der „Short Portable Mental Status Questionnaire“ (SPMSQ) (s. Bahrmann et al., 2010) diskutiert. Wichtig wäre, dass der Blick nochmals eingehender auf die Objektivität, Reliabilität, Validität und Trennschärfe der zur Risikoadjustierung eingesetzten

Messverfahren gerichtet werden sollte. Wenn der Nachweis für die bereits vorliegende Systematik erbracht werden könnte, bestünde keine Veranlassung, nach Alternativen zu suchen.

Unabhängig davon empfehlen wir nachdrücklich, nicht nur die Messverfahren zu überprüfen und wenn nötig zu optimieren, sondern dringend das Hauptaugenmerk auf eine Weiterentwicklung des Risikoadjustierungsverfahrens als solchem zu richten. Hierbei sollte im Vordergrund stehen, zukünftig multiple Risiken zu dokumentieren und in die Risikoadjustierung einzubeziehen.

4.4.2 Fazit

Der von Wingefeld et al. (2011) vorgelegte Vorschlag zur Risikoadjustierung ist in mehrfacher Hinsicht noch optimierungsfähig. Dabei ist nur ein Kritikpunkt, dass die Validität, Messgenauigkeit und Trennschärfe der aus dem NBA übernommenen Skalen zur Identifikation der Risikogruppen, z. B. BewohnerInnen mit hohen und niedrigen kognitiven Einbußen, nicht hinreichend belegt ist und hier daher ggf. besser erprobte Verfahren zum Einsatz kommen sollten. Die von Wingefeld et al. (2011) präferierte Adjustierungsmethode auf der Grundlage von Stratifizierung, weist an sich schon die wesentliche Limitierung auf, dass nur eine sehr begrenzte Anzahl an potenziellen Risiken einbezogen werden kann. Die internationale Literatur zeigt aber, dass bei allen zentralen Indikatorbereichen der Wingefeld-Systematik in der Regel eine Vielzahl von weiteren Risikofaktoren eine Rolle spielen. Mitunter kann selbst, die von Wingefeld et al. (2011) unterstellte, zentrale Rolle der kognitiven Beeinträchtigung nicht durchgehend belegt werden. Auf dieser Grundlage ist eine adäquate Risikoadjustierung nicht garantiert. Da diese für einen fairen Anbietervergleich unerlässlich ist und bei Wingefeld et al. (2011) sogar in die Definition der Indikatoren einfließt, indem bspw. separate Indikatoren für BewohnerInnen mit hohem und niedrigen Risiken gebildet werden, empfehlen wir eine alternative Risikoadjustierungsmethode, auf der Basis von multiplen Regressionsanalysen, bei der eine Vielzahl von Risikofaktoren gleichzeitig betrachtet und adjustiert werden können. Da diese Technik unabhängig von der Indikatordefinition ist, kann auf diese Weise, bei vergleichbarer Reichweite, die Gesamtzahl der Qualitätsindikatoren deutlich gesenkt werden. Entsprechende Verfahren sind vor allem in der amerikanischen Pflegeforschung Standard und prinzipiell auch auf den deutschsprachigen Raum übertragbar. Dies bedeutet allerdings auch, dass die Datenerhebung um die entsprechenden Risiken erweitert werden müsste. Auf diesem Hintergrund wäre ebenfalls ein systematisches Literaturreview, der für die relevanten Indikatorbereiche im Einzelnen maßgeblichen Risikofaktoren, ratsam.

4.5 Verbraucherperspektive

(Martin Müller, Jens Arnold & Renate Stemmer)

Das vorliegende Kapitel geht unter anderem der Frage nach, welche Indikatoren aus Verbraucherperspektive relevant sind. Hierzu wird der Begriff „Verbraucher“ zuerst definiert, und allgemeine Informationen zu Qualitätsindikatoren und konkrete Qualitätsindikatoren, die für die NutzerInnen relevant sind, benannt. Abschließend werden vor dem Hintergrund der Verbraucherperspektive Bezüge und Empfehlungen im Hinblick auf die von Wingensfeld et al. (2011) beschriebenen Qualitätsindikatoren (s. Kapitel 4.5.2) und die von den Autoren für den Vergleich von Anbietern vorgeschlagene Bewertungssystematik (s. Kapitel 4.5.3) diskutiert.

Der Verbraucherperspektive kommt insgesamt eine wichtige Rolle zu. Eine Qualitätsberichterstattung, ausschließlich von und für Fachkräfte, wäre letztendlich kaum als ausgewogen zu bezeichnen, da ein solcher Ansatz die Bedürfnisse der LeistungsadressatInnen in der Regel nicht hinreichend berücksichtigt. Die „Auswüchse“ eines derartigen Vorgehens können sich u. a. in einer Beschränkung auf rein defizitbezogene Indikatoren manifestieren, die Effektivität des Pflegesystems nicht in seiner genuinen Zielgröße, nämlich einer personenzentrierten Förderung bzw. Aufrechterhaltung von Lebensqualität, sondern lediglich als Absenz von in der Regel ausschließlich auf Unterlassungen zurückzuführenden, unerwünschten Ereignissen wie Wundliegegeschwüre, Mangelernährungssituationen o. Ä. zu definieren (vgl. Naylor, 2007; Mor et al., 2003). Die Forschungslage zeigt, dass zwischen derartigen Outcomes und den Lebensqualitätseinschätzungen von BewohnerInnen nicht zwingend Korrelationen bestehen müssen (Netten et al., 2012). Mit den geeigneten Instrumenten ist es dabei übrigens durchaus möglich, auch bei Menschen mit hohen kognitiven Einschränkungen etwa in Folge demenzieller Erkrankungen zu brauchbaren Einschätzungen der Lebensqualität zu kommen (vgl. Gräske, Fischer, Kuhlmeier, & Wolf-Ostermann, 2012). So fordern auch Tolson et al. (2011) die pflegebedürftigen Menschen und ihre Angehörigen möglichst aktiv in die Qualitätsdiskussion einzubinden. Qualitätsindikatoren sollten den Autoren folgend so entwickelt werden, dass sie zum einen in Bezug auf klinische und pflegerische Bedürfnisse sensitiv sind, jedoch vor allem würdevoll und respektvoll gegenüber dem Recht der Menschen mit Pflegebedarf sind. Auf welche Art und Weise die Einbindung der Betroffenen konkret zu erfolgen hat, wird an dieser Stelle allerdings nicht beschrieben.

Definition Verbraucher

Seit Mitte der neunziger Jahre des letzten Jahrhunderts setzte ein terminologischer Wandel bisheriger Begrifflichkeiten im pflegerischen Setting ein. Durch den Einzug marktwirtschaftlicher Terminologien und Gesichtspunkte im Gesundheitswesen ging der Trend von den entmündigenden und passiven Eigenschaften der Begrifflichkeiten Patient/ Bewohner hin zu mehr Autonomie und Souveränität. Der Terminus „Verbraucher“ zielt auf den Erwerb eines Gutes und den Kauf einer Dienstleistung ab. Primär geht es um die Bedürfnisbefriedigung und die persönliche Nutzenmaximierung. In diesem Zusammenhang wird ein weiterer Begriff – „Nutzer“ – diskutiert. Hier steht die Inanspruchnahme von Angeboten und Leistungen im Mittelpunkt (Schaeffer, 2004).

Für den Sektor der stationären Altenpflege werden zwei Nutzergruppen benannt. Die primäre Gruppe besteht aus den Menschen mit Pflegebedarf, da sie die HauptabnehmerInnen der erbrachten Leistungen darstellen. Die zweite Nutzergruppe sind die Angehörigen. Diese sind gesondert zu beachten, da sie eine Vielzahl von Funktionen erfüllen. Hierzu gehören beispielsweise in vielen Fällen die Heimauswahl und die Funktion als Bindeglied zwischen Pflegepersonal und Menschen mit Pflegebedarf (Josat, Schubert, Schnell, & Köck, 2006). Aus diesem Grund sollten nach Möglichkeit relevante Ergebnisqualitätsindikatoren für beide Nutzergruppen getrennt dargestellt werden.

4.5.1 Welche Indikatoren sind aus der Verbraucherperspektive relevant?

Zuerst soll auf die übergreifenden Kriterien Bezug genommen werden, die für die VerbraucherInnen bei der Auswahl einer Pflegeeinrichtung von Bedeutung sind. Seit 2009 gibt es für Deutschland verbindliche und flächendeckende Prüfungen von Pflegeeinrichtungen. Die hieraus abgeleiteten Pflegetransparenzberichte werden von den Verbrauchern und VerbraucherInnen als sinnvolle Ergänzung bei der Auswahl von Pflegeeinrichtungen angesehen (Hasseler & Wolf-Ostermann, 2010).

Inwiefern die Informationen dieser Berichte die Informationsbedürfnisse der NutzerInnen von Pflegeeinrichtungen tatsächlich erfüllen, beschreiben Geraedts et al. (2012) im Gesundheitsmonitor 2011. Zentrale Frage ist, welche der 36 ausgewählten Kriterien die VerbraucherInnen dazu veranlassen, eine Pflegeeinrichtung für ihre Angehörigen auszuwählen. Grundlage bildet eine Befragung von 1782 Personen. Dargestellt werden jeweils der ermittelte Rangplatz der Kriterien und der Grad der von den befragten Personen eingeschätzten Wichtigkeit.

Die Befragten räumen den Pflegeergebniskriterien erstaunlich wenig Relevanz ein. Die entsprechenden Kriterien finden sich auf den hinteren Rängen wieder. Um dies zu veranschaulichen, sind im Folgenden ausschließlich die Pflegeergebniskriterien nach der in der Studie ermittelten Wichtigkeit und dem zugehörigem Rangplatz, beginnend mit den aus Nutzersicht subjektiv relevantesten Aspekten, aufgelistet (in Klammern jeweils Rangplatz und Wichtigkeit in Prozent):

- geringer Anteil an BewohnerInnen mit Druckgeschwüren (Rangplatz 24 / 69,8%)
- geringer Anteil an BewohnerInnen mit mittelschweren oder schweren Schmerzen (Rangplatz 27 / 56,7%)
- geringer Anteil an BewohnerInnen mit einer Zunahme bei Depressionen oder Ängsten (Rangplatz 29 / 44,1%)
- geringer Anteil an BewohnerInnen die beim Aufenthalt in der Pflegeeinrichtung zu viel Gewicht verloren haben (Rangplatz 30 / 42,1%)
- hoher Anteil an BewohnerInnen mit Gripeschutzimpfung (Rangplatz 31 / 38,8%)
- geringer Anteil an bettlägerigen BewohnerInnen (Rangplatz 32 / 37,4%)
- geringer Anteil an BewohnerInnen, die Harndrang oder Stuhlgang nicht mehr kontrollieren können (Rangplatz 32 / 37,4%)
- geringer Anteil an BewohnerInnen mit Dauerkatheter (Rangplatz 34 / 34,4%)

- geringer Anteil an BewohnerInnen, die bei alltäglichen Verrichtungen stetig mehr Unterstützung brauchen (Rangplatz 35 / 30,9%)

Die größte Relevanz zeigte sich nach Geraedts et al. (2012) dahingegen bei Kriterien, die Bereiche des Pflegeprozesses, des Personals und anderer Strukturmerkmale umfassen. Die wichtigsten zehn Kriterien bei der Auswahl einer Pflegeeinrichtung waren dabei: (1) Respektvoller Umgang mit den pflegebedürftigen Menschen, (2) Freundlichkeit des Personals, (3) ausreichend Zeit für die Pflege, (4) Qualifikation der Pflegekräfte, (5) Anzahl der Pflegekräfte, (6) Qualität von Speisen und Getränken, (7) ausreichend Zeit für Zuwendung und Gespräche, (8) erster Eindruck der Einrichtung auf BesucherInnen, (9) Sauberkeit der Einrichtung und (10) regelmäßige ärztliche Betreuung. Unter den wichtigsten Kriterien finden sich somit größtenteils personalbezogene Kriterien, Kriterien der Lebensqualität und der medizinischen Versorgung. Zusammenfassend kommen die AutorInnen zu dem Schluss, dass die Qualität des Personals bei der Bewertung von Pflegeeinrichtungen einer der zentralen Aspekte für die Nutzerperspektive ist. Aus diesem Grund sollen künftige bundesweit einheitliche Qualitätsvergleiche von Pflegeeinrichtungen den Fokus auf verständliche und aus Sicht der VerbraucherInnen relevante Kriterien richten. Hierbei sei insbesondere auf die persönliche Zuwendung und die menschliche Betreuung der Menschen mit Pflegebedarf abzielen (Geraedts et al., 2012).

Was sind die Gründe, warum die von vielen PflegeexpertInnen so einhellig geforderten Outcomekriterien bei den VerbraucherInnen offensichtlich weniger hoch im Kurs stehen? Größere Relevanz haben Struktur- und Prozessmerkmale häufig deshalb, weil sie aus Verbraucherperspektive besser zu beobachten und einzuordnen sind (Pesis-Katz et al., 2013). Die Kriterien, denen Pflegeergebnisse zugrunde liegen, werden daher oft aus Gründen der Nachvollziehbarkeit und Verständlichkeit vernachlässigt. Es ist in diesem Zusammenhang absolut nachvollziehbar, dass VerbraucherInnen Qualität eher an der konkreten Dienstleistung orientiert bewerten. Zudem kann davon ausgegangen werden, dass VerbraucherInnen intuitiv unterstellen, dass hier ein unmittelbarer Ergebnisbezug besteht. Die Erkenntnis, dass eine hohe Fachlichkeit im Sinne von überdurchschnittlicher Struktur- und Prozessqualität nicht automatisch auch gute Ergebnisse bedingen muss (was letztendlich ein Hauptgrund ist, dass Ergebnisindikatoren eine essentielle Rolle im Qualitätsdialog einnehmen sollten), oder anders ausgedrückt, dass beispielsweise durch ein in jeder Hinsicht vorbildlich persönlich zugewandtes, respektvolles und gut qualifiziertes Pflegepersonal in einer modernen Wohnanlage mit höchsten Hygienestandards und hervorragendem Essen nicht zwingend verhindert werden kann, dass HeimbewohnerInnen gravierende Gesundheitsschädigungen erleiden, ist eine (wissenschaftliche) Erkenntnis, die ein hohes Abstraktionsniveau erfordert und keineswegs besonders trivial oder naheliegend ist. Nichtsdestotrotz muss ein praxistaugliches Qualitätsberichterstattungssystem natürlich auch den Informationsbedürfnissen der VerbraucherInnen gerecht werden (vgl. auch Abschnitt „Allgemeine Anforderungen an Qualitätsberichte aus Verbraucherperspektive“ im Kapitel 4.5.3 weiter unten).

Wie einleitend dargestellt, erscheint es lohnenswert, die Frage der aus Verbrauchersicht relevanten Qualitätsindikatoren den spezifischen Sichtweisen sowohl der pflegebedürftigen Menschen als auch deren Angehörigen entsprechend, nochmals einer differenzierteren Betrachtung zu unterziehen. Dies soll in den folgenden Abschnitten geschehen.

Spezifisch für Menschen mit Pflegebedarf relevante Qualitätsindikatoren

Für Menschen mit Pflegebedarf erscheinen Qualitätsindikatoren besonders relevant, die Informationen über Lebensqualität, Autonomie und Zufriedenheit enthalten (Mor, 2006). Die Bezugnahme auf Zufriedenheit ist in vielen Qualitätsberichten einzelner Staaten der USA als Qualitätsindikator vorzufinden. Als „Stimme der Verbraucher“ wird hierbei oft zwischen Bewohner- und Angehörigenzufriedenheit differenziert. Im Gegensatz zu den eher „medizinlastigen“ Qualitätsindikatoren stellt die Zufriedenheit auch nach Castle & Ferguson (2010) eine Möglichkeit dar, den Nutzerinteressen besser gerecht zu werden.

Josat et al. (2006) beschreiben auf der Grundlage eines Reviews Qualitätskriterien, die aus Sicht der pflegebedürftigen Menschen als besonders relevante Kriterien für die Qualität gewertet werden. Hierzu gehört zunächst die autonome Lebensführung. Zentrales Anliegen der Menschen mit Pflegebedarf ist das selbstbestimmte Leben. Die Qualität von Altenheimen wird demnach anhand folgender Aspekte bewertet: Wahlmöglichkeiten bei der Zimmerart, den Aktivitäten und dem Essensangebot. Ein weiteres Qualitätskriterium bezieht sich auf die sozialen Kontakte. Wichtig sind den AltenheimbewohnerInnen Kontakte zu MitbewohnerInnen, zu ihrer Familie und zu den Pflegekräften. Darüber hinaus wird auf den Bereich der Versorgung Bezug genommen. Hierzu gezählt wird die Essensversorgung, die Wäscheversorgung, Pflege, Unterstützung und das Angebot von Aktivitäten. Das Kriterium des Personals umfasst die Empathie, mit der das Personal dem Bewohner bzw. der Bewohnerin gegenüber treten sollte. Hinzu kommt, dass Zeit für Gespräche und genügend gut qualifiziertes und deutsch sprechendes Personal vorhanden sein soll. Letztlich geht es u. a. auch um Merkmale wie die Lage und um die Ausstattung des Pflegeheims. Wichtige Aspekte sind hierbei etwa zentrale Lage, Nähe zum ursprünglichen Wohnort sowie die Ausstattung und Atmosphäre des Pflegeheims.

Zusammenfassend können für Menschen mit Pflegebedarf sieben relevante Bereiche von Qualitätsindikatoren ausfindig gemacht werden:

1. Qualitätsindikatoren, die die Lebensqualität der pflegebedürftigen Menschen abbilden
2. Qualitätsindikatoren, die die Autonomie der pflegebedürftigen Menschen abbilden
3. Qualitätsindikatoren, die die Zufriedenheit der pflegebedürftigen Menschen abbilden
4. Qualitätsindikatoren, die soziale Kontakte der pflegebedürftigen Menschen abbilden
5. Versorgungsbezogene Qualitätsindikatoren
6. Personalbezogene Qualitätsindikatoren
7. Institutionsbezogene Qualitätsindikatoren

Bei näherer Betrachtung dieser Indikatoren wird wie bereits eingangs beschrieben recht schnell deutlich, dass nur wenige davon im engeren Sinne Ergebnisqualität abbilden. Dazu sind die Bereiche Lebensqualität, Autonomie und soziale Kontakte zu rechnen. Die anderen Bereiche referenzieren in der Mehrzahl auf die teilweise miteinander verschränkten Prozess- und Strukturmerkmale der Anbieter (Zufriedenheit, Versorgung, Personal und Institutionsmerkmale).

Spezifisch für Angehörige relevante Qualitätsindikatoren

Auf die Frage, welche Qualitätsindikatoren für die Gruppe der Angehörigen als relevant einzustufen sind, geben Geraedts et al. (2012) wichtige Hinweise. Die AutorInnen beschreiben, dass der Fokus der zu entwickelnden Qualitätsindikatoren, die aus Sicht der Angehörigen relevant sind, vorrangig Aspekte der Zuwendung des Personals gegenüber den Menschen mit Pflegebedarf und Aspekte der guten menschlichen Betreuung enthalten. Konkret beziehen sich die Aussagen auf die am wichtigsten eingestuften Kriterien: respektvoller Umgang mit den pflegebedürftigen Menschen, Freundlichkeit des Personals und ausreichend Zeit für die Pflege.

Auf der Grundlage einer qualitativen Studie beschreibt Josat (2005) weitere Kategorien, die für Angehörige wichtig sind. Hierzu die Institution (bauliche Gegebenheiten, Sauberkeit, Ordnung und Geruch), die Versorgung (familiäre Art und Weise der Versorgung, pflegerische Grundversorgung und Kümmern) und das Wohlbefinden der BewohnerInnen betreffen. Die Angehörigen legten dabei Wert auf eine individuell, an den BewohnerInnen angepasstes Angebot. Die Angehörigen hätten den AutorInnen zufolge ferner konkrete Vorstellungen in Bezug auf das Personal, einerseits was Kompetenz und Verantwortungsübernahme, andererseits aber auch eine adäquate Anzahl an Fachkräften anbetrifft. Der Faktor „Lage und Ausstattung“ werde zudem auch vor dem Hintergrund des Preises der Einrichtung bewertet. Das Wohlfühlen der Menschen mit Pflegebedarf werde von den Angehörigen schließlich als am Wichtigsten eingestuft (Josat et al., 2006).

Zusammenfassend können den Quellen zufolge fünf Indikatorenbereiche identifiziert werden, die für die Gruppe der Angehörigen relevant sind:

1. Personalbezogene Qualitätsindikatoren
2. Betreuungsbezogene Qualitätsindikatoren
3. Versorgungsbezogene Qualitätsindikatoren
4. Institutionsbezogene Qualitätsindikatoren
5. Auf das Wohlbefinden der Menschen mit Pflegebedarf bezogene Qualitätsindikatoren

Im Vergleich zu den spezifisch für pflegebedürftige Menschen relevanten Qualitätsindikatoren erscheinen die für die Angehörigen bedeutsamen Qualitätsbereiche insgesamt noch weniger ergebnisbezogen. Lediglich der Bereich „Wohlbefinden der pflegebedürftigen Menschen“ tangiert diese Ebene, wobei auch hier Fragen nach dem Pflegesetting und einer guten Betreuung im Vordergrund stehen dürften.

4.5.2 Sind Indikatoren bzw. die Empfehlungen von Wingenfeld et al. für eine verständliche, übersichtliche und vergleichbare Information der Verbraucher geeignet? Ist die verfügbare Information in ihrer Auswahl hinreichend, in ihrer Bewertung und Darstellung verständlich und erlaubt sie einen Vergleich der Einrichtungen/Dienste insbesondere bezogen auf Ergebnisqualität? Wenn nein, welche Voraussetzungen, Änderungen, Ergänzungen wären hierfür notwendig?

Bevor der konkrete Bezug zum Ansatz von Wingenfeld et al. (2011) hergestellt werden kann, ist es zunächst erforderlich grundlegende Anforderungen an die Qualitätsberichterstattung zu diskutieren, die unmittelbar mit der Verbraucherperspektive verknüpft sind.

Allgemeine Anforderungen an Qualitätsberichte aus Verbraucherperspektive

Durch eine Verfügbarkeit von qualitativ hochwertigen Informationen erhöht sich die Präzision der Nachfrage von Seiten der NutzerInnen in Bezug auf Qualität (Konetzka et al., 2013). So empfiehlt es sich, ein umfassendes Verbraucherinformationssystem zu entwickeln, das Informationen sowohl zu Strukturen, Prozessen als auch Outcomes enthält. Diese Prämisse wird von den internationalen Erfahrungen unterstützt (Hasseler & Wolf-Ostermann, 2010). Es ist unumgänglich, dass solche Qualitätsberichte bestrebt sein müssen, die häufig vorhandenen Informationsasymmetrien in Bezug auf die Pflegequalität zwischen Pflegeheimen und VerbraucherInnen zu reduzieren (Troyer & Thompson, 2004)

NutzerInnen wünschen sich das „gute“ Pflegeheim. Dies stellt die Qualitätsberichterstattung aus Nutzerperspektive vor große Herausforderungen. Notwendig wären Qualitätsindikatoren, die transparente und öffentliche Kriterien beinhalten (Mor et al., 2003; Mor, 2006). Hasseler und Fünfstück (2012) definieren vor diesem Hintergrund spezifische Anforderungen an die Qualitätsberichterstattung aus Nutzerperspektive, die in Tabelle 9 musterhaft dargestellt sind.

Entscheidungen über die Auswahl einer Pflegeeinrichtung werden häufig sehr kurzfristig getroffen, aus diesem Grund müssen jederzeit aktuelle Informationen zur Verfügung stehen. Hierbei meint Angemessenheit, dass nicht nur die Ergebnisse präsentiert werden. Es sollen auch Auskünfte darüber gegeben werden, wie die Ergebnisse generiert wurden. Relevanz bedeutet in diesem Zusammenhang, dass die aus der Nutzerperspektive wirklich wichtigen Kriterien eingeschlossen sind. Die Nützlichkeit ist darin definiert, dass die enthaltenen Informationen zu einer fundierten Entscheidung für oder gegen eine Einrichtung führen können. In Bezug auf die Verständlichkeit der Informationen verweisen die AutorInnen auf die Anforderungen sowohl an die Technik der Webseiten als auch an Entscheidungshilfen und die Einfachheit der Sprache. Informationspräferenzen werden angesprochen, da für die Nutzerperspektive nicht immer fachliche Kriterien ausschlaggebend sind. So können beispielsweise Faktoren wie Wohnortnähe, Kosten, Versorgungsformen, Ausstattung, persönliche Zuwendung und Kommunikation ausschlaggebend sein. Für die Zielgruppe der VerbraucherInnen ist es notwendig, dass konsequent auf die Verständlichkeit der Formulierungen geachtet wird, um sicherzustellen, dass die Informationen auch von Laien problemlos nachvollzogen werden kann. Die Veröffentlichung von Qualitätsberichten sollte eine analytische und eine

intuitive Denkmethode berücksichtigen, damit LeserInnen in ihren individuellen Denkschemata angesprochen werden (Hasseler & Fünfstück, 2012).

Tabelle 9: Kritische Punkte und Anforderungen an eine systematische Qualitätsberichterstattung (Hasseler & Fünfstück, 2012)

Kritischer Punkt	Anforderung an Qualitätsbericht
<p>Zeitpunkt der Entscheidung (Entscheidung unter Zeitdruck, fehlende health literacy, Informationen zu verstehen etc.)</p>	<p>Qualitätsberichte müssen zeitnah aktuelle Informationen vermitteln und zu jeder Zeit abrufbar sein. Die Informationen müssen adäquat, relevant, nützlich und flächendeckend verfügbar sein.</p>
<p>Verständlichkeit der Informationen (Menge der Informationen, Sprache etc.)</p>	<p>Die dargestellten Informationen sollen verständlich bzw. erklärt sein. Die Menge der Informationen kann von Verbraucher selbst gesteuert werden. Idealerweise ist ein Entscheidungsstool integriert.</p>
<p>Informationspräferenzen</p>	<p>Qualitätsberichte sollen Informationen enthalten, die für Verbraucher in Situationen der Entscheidungsfindung relevant sind. Die dargestellten Informationen sind von Bedeutung für Verbraucher und an Informationsbedürfnisse angepasst</p>
<p>Layout und Design der Webseite</p>	<p>Qualitätsberichte sollten strukturiert sein und Verbrauchern ermöglichen, nach eigenen Suchkriterien und Gewichtungen auszuwählen. Das Layout soll ein intuitives Bedienen der Webseite ermöglichen. Das visuelle Design und die Navigation müssen der Zielgruppe angepasst sein.</p>
<p>Art und Weise der Darstellung</p>	<p>Unterschiedliche Zielgruppen verarbeiten Informationen auf verschiedene Weise. Die Präsentation der Informationen sollen Verbraucher darin unterstützen, Informationen zu verarbeiten. Dies kann durch Symbole, Wörter oder Sterne geschehen. Ältere Menschen verarbeiten Informationen anders als jüngere, so dass emotionalisierende Informationen angemessener erscheinen.</p>
<p>Menge der Information</p>	<p>Es kann nur eine bestimmte Anzahl von Informationen verarbeitet werden. Zu viele Informationen verkürzen den Such- und Entscheidungsprozess. Die Informationsmenge muss der Zielgruppe angepasst sein oder es sollte ermöglicht sein, dass die Zielgruppe die Menge der Informationen selbständig steuert.</p>
<p>Limitation von Qualitätsberichten</p>	<p>Qualitätsberichterstattungen unterliegen Limitationen, diese sollten angemessen und transparent dargestellt werden.</p>

Im Folgenden werden, soweit wissenschaftliche Belege vorliegen, Aussagen dazu getroffen, auf welche Art und Weise die teils abstrakten Anforderungen aus Tabelle 9 umgesetzt werden können.

Um VerbraucherInnen nicht zu verwirren, sollten verständliche Informationen gegeben und der Umfang der Informationen vom Nutzer bzw. von der Nutzerin selbst bestimmt werden. Wichtige Mittel, dies zu erreichen, sind das Layout und das Design der Webseite, die Art und Weise der Darstellung sowie die Menge der Informationen (Hasseler & Fünfstück, 2012). Von Hasseler und Wolf-Ostermann (2010) ausgewertete Interviews zeigten eine Präferenz für eine sachorientierte Gestaltung. Die Suchoberfläche sollte intuitiv zu bedienen und Barrierefreiheit sichergestellt sein. Auch andere Quellen belegen, dass internetbasierte Qualitätsberichte für VerbraucherInnen ein im Vergleich zu anderen Publikationsmedien nützliches Mittel zur Auswahl einer Einrichtung zu sein scheinen (vgl. Berlowitz et al., 2005; Carlisle, 2007).

Die Verständlichkeit der Qualitätsberichte kann auch im Hinblick auf die Verbraucherperspektive durch die Verwendung eines einheitlichen Maßsystems erreicht werden (Mor, 2005; Abraham, MacDonald, & Nadzam, 2006). In diesem Zusammenhang sollte ebenfalls die Forderung nach mindestens vier Bewertungsstufen erwähnt werden (Weibler-Villalobos & Röhrig, 2010). Eine differenziertere Bewertungssystematik kann auch den Nutzer bzw. die Nutzerin in seiner Auswahl unterstützen (s. Kapitel 4.3.1 bzw. siehe unten). Die Informationen der Qualitätsberichte können für ältere Nutzergruppen zudem in erzählerischer Form abgefasst werden, um auf diese Weise ein emotionales Interesse zu wecken. Nach Hasseler und Fünfstück (2012) kann so erreicht werden, dass die Informationen besser verstanden werden.

An dieser Stelle sollen aber vor dem Hintergrund der Qualitätsberichterstattung im Zusammenhang mit der Verbraucherperspektive auch kritische Aspekte und Limitierungen nicht verschwiegen werden. Häufig sind gar nicht die VerbraucherInnen selbst diejenigen, die den Pflegeheimplatz auswählen (wollen). Die Auswahl des Pflegeheims kann auch durch Krankenhäuser und deren Sozialdienste, Pflegeberatungen, patientenbegleitende Dienste der Krankenkassen (Löcherbach, Hermsen, Macsenaere, Arnold, & Klein, 2010) oder andere Akteure des Gesundheitswesens mitbeeinflusst werden. So können öffentliche Qualitätsvergleichssysteme mitunter nur einen geringen oder nur mittelbaren Einfluss auf die Pflegeheimauswahl der VerbraucherInnen haben (Mor, 2005). Nach Mor (2005) wird neben einrichtungsbezogenen Charakteristika oft einfach auf Grundlage der geographischen Nähe entschieden. Auch in anderen Fällen werden Qualitätsberichte von vielen Verbrauchern und Verbraucherinnen nicht in dem gewünschten Ausmaß genutzt. Eine US-Studie zeigte beispielsweise, dass weniger als ein Drittel der Angehörigen bei ihrer Pflegeheimwahl explizit auf solche Quellen zurückgegriffen haben. Dies wird u. a. auch auf die teils schwierige Verständlichkeit der Berichte zurückgeführt (Pesis-Katz et al., 2013). Verständlichkeit rekuriert im Wesentlichen auf die Nutzung einer Sprache, die der Zielgruppe angemessen ist. Konkret bedeutet dies auf Abkürzungen zu verzichten, Fachbegriffe weitgehend zu vermeiden, Inhalte im Zusammenhang darzustellen und so wenig wie möglich mit Verweisen zu arbeiten.

Einordnung der Wingenfeld-Indikatoren

Vor dem Hintergrund der bisherigen Ausführungen muss die Frage, ob die von Wingenfeld et al. (2011) vorgelegten Qualitätsindikatoren eine aus Verbrauchersicht hinreichende Auswahl darstellen, verneint werden. Das grundlegende „Handicap“ ist hier aber, dass VerbraucherInnen sich bei der Auswahl von Pflegeheimen zunächst eher an strukturellen und prozessualen Kriterien orientieren wollen. Dies kann einem Instrument, das die spezifische Zielsetzung verfolgt, gesundheitsbezogene Ergebnisqualität in der stationären Altenhilfe zu untersuchen, natürlich nicht zur Last gelegt werden. Trotzdem gehen auch die beschriebenen Indikatoren nicht völlig an den Informationsbedürfnissen der VerbraucherInnen vorbei. Insbesondere die Indikatoren des Qualitätsbereichs 1 wie „Alltagsverrichtungen“ und „Gestaltung des Alltagslebens und sozialer Kontakte“ umfassen Bereiche, die wie beschrieben speziell für die pflegebedürftigen Menschen eine hohe Relevanz haben.

Auch die Wingenfeld-Indikatoren des Qualitätsbereichs 2, im Rahmen derer es um die Häufigkeit von gesundheitsschädigenden Einflüssen und Belastungen wie Dekubitusentstehung, Stürze und Gewichtsverlust geht, sind aus Sicht der GutachterInnen der vorliegenden Expertise für die Verbraucherinformation nicht grundsätzlich ungeeignet. Dabei sollte sich deren Relevanz und Bedeutung auch dem Laien bzw. der Laiin schnell erschließen. Auch wenn die VerbraucherInnen bei der Suche nach Pflegeeinrichtungen natürlich nicht in erster Linie potenziell negative gesundheitliche Folgeerscheinungen im Sinn haben, sind derartige Informationen vor dem Hintergrund einer umfassenden Aufklärung trotzdem wichtig und nützlich. Hierbei muss auch bedacht werden, dass die Informationsbedürfnisse der VerbraucherInnen sich natürlich ebenfalls nach dem verfügbaren Informationsangebot ausrichten. Insofern kann eine aussagekräftige Qualitätsberichterstattung nicht ausschließlich die „Kundenwünsche“ in den Mittelpunkt stellen. Ergebnisindikatoren sollten hierbei ein zentraler Themenbereich sein. Die Informationsbedürfnisse der LeistungsadressatInnen dürfen im Hinblick auf die Akzeptanz aber auch nicht völlig ausgeblendet werden. Es ist daher dringend zu empfehlen, die Qualitätsberichterstattung nicht alleine auf Outcomes zu beschränken. Eine zentrale Dimension mit sehr hoher Nutzerrelevanz, die auf jeden Fall Berücksichtigung finden sollte, ist der Bereich der Lebensqualität. Dies lässt sich ebenfalls mit den Maßstäben und Grundsätzen (MuG) nach § 113 SGB XI für die vollstationäre Pflege in Übereinklang bringen (MuG stat., 2011). Neben den vorliegenden Vorschlägen der Arbeitsgruppe des Instituts für Sozialforschung und Gesellschaftspolitik ISG für die gesundheitsbezogenen Wingenfeld-Qualitätsbereiche ergänzende Qualitätsindikatoren zur Erfassung von Lebensqualitätsaspekten (vgl. Wingenfeld et al., 2011), die allerdings nicht Gegenstand der vorliegenden Expertise sind, beschreibt die Literatur in diesem Zusammenhang eine Reihe von geeigneten Verfahren (vgl. Haywood, Garratt, & Fitzpatrick, 2005). Ein vielversprechendes Instrument zur Messung der Lebensqualität bei kognitiv beeinträchtigten Menschen ist z. B. das QUALIDEM „Quality of life in dementia“ (Ettema, Droes, de Lange J., Mellenbergh, & Ribbe, 2007a; Ettema, Droes, de Lange J., Mellenbergh, & Ribbe, 2007b; Bouman et al., 2011). Da es sich in diesem Beispiel um eine Fremdeinschätzung durch Pflegefachpersonen handelt, sei an dieser Stelle darauf hingewiesen, dass im Sinne einer ausgewogenen Evaluation, soweit möglich, allerdings auch die persönlichen Einschätzungen der Betroffenen durch unmittelbare Befragung der BewohnerInnen oder Angehörigen eingeholt werden sollten.

Neben der outcomebezogenen Lebensqualität sollte vor dem Hintergrund der Verbraucherperspektive auch in Erwägung gezogen werden, die Qualitätsberichte um zentrale Aspekte

der Struktur- und Prozessqualität zu ergänzen. Eine relevante Dimension könnten hier beispielsweise die Angabe von Personalkapazitäten und Betreuungsschlüsseln sein (s. Kapitel 4.1.2). Eine Erweiterung der ergebniszentrierten Perspektive zugunsten weiterer Qualitätsdimensionen wäre auch insofern interessant, als dass nur so im Rahmen der Qualitätsberichterstattungen auch Aussagen zu den für die Ergebnisse verantwortlichen Wirkfaktoren ermöglicht werden würden. Beschränkt sich die Dokumentation ausschließlich auf die Ergebnisqualität, so liegt letztendlich eine sog. „Black Box“-Evaluation vor, die Erfolge und Misserfolge zwar belegen, bei negativem Outcome aber keine unmittelbar steuerungsrelevanten Optimierungspotenziale aufzeigen kann. Dabei wäre es auch aus Forschersicht wünschenswert, wenn die Zusammenhänge zwischen den Ergebnisdimensionen noch genauer beleuchtet werden könnten.

Die von Wingenfeld et al. (2011) entwickelten Indikatoren haben im Qualitätsbereich 3 („spezifische Bedarfslagen“) eher den Charakter von Output- und weniger von klassischen Outcomeindikatoren. Während es definitorisch betrachtet bei den Outcomes um Ergebnisqualität im engeren Sinne geht, umschreiben Outputs die Versorgungsleistung. Mit diesem Fokus müssten sie eigentlich den dargestellten Informationsbedürfnissen der VerbraucherInnen eher entgegenkommen. Dies ist nach Einschätzung der GutachterInnen bei den vorliegenden Indikatoren aber nicht der Fall, da die verbraucherrelevanten Bereiche hier nicht abgedeckt sind (vgl. Kapitel 4.5.1). Das heißt natürlich nicht, dass sie damit aus Fachkräftesicht irrelevant wären. In dieser Hinsicht sind auf jeden Fall die Bereiche Schmerz, Verhaltensauffälligkeiten und Gurtfixierungen von Bedeutung.

Übersichtlichkeit und mitunter auch Verständlichkeit der Indikatoren hängen neben der Art ihrer Operationalisierung unmittelbar mit der Bewertungssystematik (Kapitel 4.3) und im vorliegenden Fall ebenfalls mit der von Wingenfeld et al. (2011) gewählten Risikoadjustierungstechnik (Kapitel 4.4) zusammen. Die Bezüge zur Verbraucherperspektive wurden in den betreffenden Kapiteln bereits ausführlich dargestellt und werden an dieser Stelle daher nur noch bilanzierend zusammengefasst. Wie in Kapitel 4.3.1 beschrieben, sollte mit Blick auf die Verbraucherperspektive bedacht werden, dass die Anzahl der betrachteten Indikatoren in einem überschaubaren Rahmen bleibt. Für die Verbraucherseite wäre es daher notwendig, anders als bei der Aufbereitung der Daten für die professionellen Akteure und Akteurinnen, eine Auswahl zentraler bzw. besonders relevanter Qualitätsindikatoren zu treffen. Unseren Empfehlungen folgend wäre die Veröffentlichung von mehr als etwa drei bis sechs Indikatoren zur Information der VerbraucherInnen hier wohl kontraindiziert. Diese Herausforderung stellt sich insbesondere vor dem Hintergrund, dass nach aktuellem Sachstand von einer Berechnung komprimierter Gesamtwerte, die sich aus mehreren Einzelindikatoren zusammensetzen, abgesehen werden sollte (vgl. Kapitel 4.3.1). Es wäre allerdings wie in Kapitel 4.3.1 beschrieben durchaus denkbar, Auswahl und Gewichtung der relevanten Bereiche ein Stück weit den Verbrauchern und Verbraucherinnen selbst zu überlassen. Auch Mor et al. (2003) machen darauf aufmerksam, dass unterschiedliche Zielgruppen für Outcomemessungen ggf. mit einer individuellen Indikatorenauswahl „bedient“ werden müssen. Auf das Beispiel des kanadischen „Quality Compass“ (Health Quality Ontario, 2014c) wurde in diesem Zusammenhang bereits verwiesen. Auf ein weiteres Mustermodell wird weiter unten im Abschnitt „Umsetzungsbeispiel“ eingegangen.

Jenseits der auch für eine vergleichende Information elementaren Differenzierungsfähigkeit, auf die in Kapitel 4.5.3 eingegangen wird, sollte sich eine für VerbraucherInnen verständliche

Bewertungssystematik möglichst intuitiv erschließen. Aufgrund seiner Ähnlichkeit mit den bei Hotels üblichen Sternevergaben würden wir daher ein fünfstufiges Bewertungsraster wie es beispielsweise in den USA mit der Fünf-Sterne-Bewertung quasi Standard ist, auch für den deutschsprachigen Raum präferieren (vgl. Kapitel 4.3.1).

Hinsichtlich der Verständlichkeit wurde ferner bereits in Kapitel 4.4.1 die von Wingefeld et al. (2011) vorgenommene Risikoadjustierung kritisiert, die als elementarer Bestandteil der Indikatordefinition zwischen Hoch- und Niedrigrisikogruppen differenziert, wobei zumeist zwischen BewohnerInnen mit vergleichsweise hohen und niedrigen kognitiven Einschränkungen unterschieden wird. Es darf bezweifelt werden, dass VerbraucherInnen die notwendigen diagnostischen Einschätzungen zugemutet werden können, ohne deren Klärung nicht eindeutig ist, welcher Qualitätsindikator im konkreten Fall tatsächlich maßgeblich ist. Durch die von uns vorgeschlagene alternative multifaktorielle Risikoadjustierungsmethode könnte nicht nur darauf verzichtet, sondern auch der Gesamtumfang der Indikatoren um ein Drittel (von 15 auf 10 Qualitätsindikatoren) reduziert werden. Dies würde gleichsam auch zu einer besseren Übersichtlichkeit beitragen. Sofern dieses Problem gelöst werden könnte, erscheinen uns ansonsten wie bereits oben erwähnt die Ergebnisindikatoren der Qualitätsbereiche 1 und 2 der Wingefeld-Systematik durchaus auch verbraucherseitig nachvollziehbar und verständlich, wobei den Indikatoren des Qualitätsbereichs 1 aus Verbrauchersicht gemäß den vorliegenden Befunden zumindest die subjektiv größere Relevanz zukommen dürfte. Für eine bessere Verständlichkeit und optimierte Darstellung sollten schließlich noch die in Tabelle 9 dargestellten Leitsätze verbraucherfreundlicher Qualitätsberichte berücksichtigt werden.

Umsetzungsbeispiel zur Orientierung

Ein modellhaftes Beispiel, die bereits genannten kritischen Punkte und Anforderungen an die systematische Qualitätsberichterstattung umzusetzen und die relevanten Qualitätsindikatoren für pflegebedürftige Menschen und deren Angehörigen abzubilden, stellt das „Nursing Home Compare“ dar (Center for Medicare and Medicaid Services, 2012). Qualitätsberichte, die auf diese Art und Weise dargestellt werden, decken die meisten beschriebenen Anforderungen aus der Verbraucherperspektive ab:

Das Center for Medicare and Medicaid Services (2012) beschreibt in der Bedienungsanleitung für dieses System eine Onlineplattform, in der verschiedene Qualitätskriterien von Pflegeheimen in einer für VerbraucherInnen verständlichen Art und Weise dargestellt sind. Die Bewertungssystematik erfolgt in einem Fünf-Sterne-System, das für die Nutzergruppe wertvolle und verständliche Vergleichsparameter liefert. Durch die Funktion der Webseite alle Qualitätsbereiche einzeln recherchieren zu können, wird dem Nutzer bzw. der Nutzerin ermöglicht, die Menge der Informationen individuell zu steuern. Im Weiteren stellt das „Nursing Home Compare“ die Aktualität der Daten durch ein monatliches Update sicher. Auf diese Weise wird erreicht, dass zu jedem Zeitpunkt aktuelle Qualitätsdaten vorliegen. Die beschriebenen Aspekte erfüllen zentrale Anforderungen an die Qualitätsberichterstattung, hierzu zählen Zeitpunkt, Verständlichkeit der Informationen, Art und Weise der Darstellung und Menge der Informationen.

Inhaltlich werden Daten aus drei Qualitätsbereichen zur Verfügung gestellt: (1) Qualitätsprüfungen im Gesundheitswesen, (2) Personalbesetzung und (3) Pflegeergebniskriterien. Der erste Bereich umfasst jährliche Überprüfungen der Pflegeheime im Hinblick auf Umgebung,

Küchen- und Essensversorgung, Betroffenenrechte und Lebensqualität. Die Personalbesetzung wird unter Berücksichtigung zweier Aspekte, der Pflegezeit durch examiniertes Pflegepersonal und der Pflegezeit durch das gesamte Pflegepersonal (einschließlich Pflegehilfpersonal) bewertet. Der Bereich der Pflegeergebniskriterien umfasst neun der Qualitätsindikatoren des Minimum Data Sets. Im Besonderen wird die für die pflegebedürftigen Menschen relevante Selbständigkeit abgebildet (Center for Medicare and Medicaid Services, 2012). Die Qualitätsbereiche decken wesentliche Qualitätsindikatoren für die beiden Verbrauchergruppen ab. Hierzu gehören für die Menschen mit Pflegebedarf Autonomie, Versorgung, Personal, Institution und Lebensqualität und für die Angehörigen die Bereiche Personal, Betreuung, Institution, Versorgung und Wohlbefinden. Zudem ist das Minimum Data Set speziell für die Nutzung durch VerbraucherInnen entwickelt worden (Straker & Bailer, 2008), so liegt es nahe, dass das beschriebene Bewertungssystem für die Verbraucherperspektive gut geeignet ist.

4.5.3 Sind die Empfehlungen von Wingenfeld et al. zur Bewertungssystematik auf Ebene der Indikatoren und der Indikatorenbereiche aus Verbraucherperspektive geeignet, eine hinreichende Differenzierung der Einrichtungen zu ermöglichen? Sind hierfür voraussichtlich fortlaufende Nachjustierungen am Bewertungssystem notwendig?

Die Beurteilung der Eignung der von Wingenfeld et al. (2011) vorgelegten Bewertungssystematik hinsichtlich ihrer Fähigkeit zwischen Einrichtungen zu differenzieren, kann nicht spezifisch aus Sicht der VerbraucherInnen diskutiert werden. Da auch empirische Aspekte wie die Diskriminationsfähigkeit der Indikatoren eine Rolle spielen, sollten letztendlich für alle Stakeholder bzw. Adressatengruppen der Qualitätsberichtserstattung einheitliche Bewertungsmaßstäbe zur Verfügung gestellt werden. Entsprechend wurde die Verbraucherperspektive im Rahmen unserer in Kapitel 4.3 dargestellten Empfehlungen bereits hinlänglich mitbedacht. Wir können an dieser Stelle daher nur auf die dortigen Ausführungen verweisen.

Kurz rekapitulierend wird das von Wingenfeld et al. (2011) vorgeschlagene dreistufige Indikatorenbewertungsraster für eine angemessene Differenzierung von Einrichtungen als zu grob empfunden. Entsprechend des Hinweises am Ende von Kapitel 4.5.2, stellt ein fünfstufiges Raster, das in weit überdurchschnittliche, überdurchschnittliche, durchschnittliche, unterdurchschnittliche und weit unterdurchschnittliche Anbieter unterteilt bei entsprechender visueller Aufbereitung wie etwa in Form des in den USA verbreiteten „Fünf-Sterne-Systems“ einen guten Kompromiss dar, der sowohl in statistischer Hinsicht eine hinreichende Differenzierung erlaubt als auch für die NutzerInnen verständlich und intuitiv nachvollziehbar ist.

Bei allen Qualitätsbewertungssystemen, unabhängig davon, ob diese norm- oder kriterienbezogen konstruiert wurden, sind aufgrund der mit der Systematik beabsichtigten fortschreitenden Qualitätsentwicklungsprozesse und sich ändernder Rahmenbedingungen fortlaufende Nachjustierungen am Bewertungssystem unabdingbar. Für die von uns präferierte kriterienbezogene Methode wäre dies auf diesem Hintergrund sogar unabdingbar, um

langfristig die Differenzierungsfähigkeit aufrechtzuerhalten (vgl. auch die detaillierten Ausführungen dazu in Kapitel 4.3.2).

4.5.4 Fazit

VerbraucherInnen haben andere Informationsbedürfnisse als Fachkräfte. Outcomeindikatoren nehmen hier grundsätzlich nur einen untergeordneten Stellenwert ein. Die VerbraucherInnen legen, im Rahmen der Entscheidung für oder gegen ein Pflegeheim, insgesamt mehr Wert auf Prozessqualität wie etwa die Freundlichkeit und Zuwendung des Personals oder strukturelle Merkmale der Einrichtung wie z. B. die geographische Nähe zum ursprünglichen Wohnort. Da der Verbraucherperspektive im Rahmen der Qualitätsberichterstattung insgesamt eine zentrale Rolle zukommt, können die Informationsbedürfnisse der VerbraucherInnen nicht ignoriert werden. Dies heißt allerdings im Umkehrschluss nicht, dass Ergebnisindikatoren hier komplett außer Acht zu lassen wären. Neben Kriterien wie der Verständlichkeit der Sprache, ist insbesondere aber auf eine überschaubare Informationsmenge zu achten. Die Veröffentlichung von insgesamt mehr als drei bis sechs Indikatoren wäre aus Sicht der GutachterInnen eher nicht zielführend. Die Auswahl der zur Verfügung gestellten Indikatoren sollte dabei auf die jeweiligen Interessensgruppen abgestimmt werden. Mit Hilfe einer geeigneten Online-Plattform könnte die Auswahl und Gewichtung allerdings teilweise auch durch die NutzerInnen selbst geschehen. Sofern man unseren Empfehlungen zur Bewertungssystematik und zur Risikoadjustierung, bei denen die Nutzersicht bereits mitgedacht wurde, sowie den allgemeinen Gestaltungsleitlinien für „kundenfreundliche“ Qualitätsberichterstattung folgt, wäre auch auf Grundlage, vieler von Wingenfeld et al. (2011) beschriebener Indikatorbereiche, ein angemessene Einrichtungsvergleich möglich, der sowohl im Hinblick auf Übersichtlichkeit als auch Verständlichkeit der Verbraucherperspektive gerecht werden kann. Dies gilt insbesondere in Bezug auf die von Wingenfeld et al. (2011) entwickelten Indikatoren der Qualitätsbereiche 1 und 2, deutlich weniger aber für die Indikatoren des Qualitätsbereichs 3.

5 Roadmap

(Renate Stemmer)

Mit dem von Wingenfeld et al. (2011) vorgelegten Projektergebnis liegt erstmals ein Vorschlag für die Einführung von Outcomeindikatoren in der stationären Altenhilfe vor. Hier wurden erste Überlegungen angestellt, die nun weiter zu entwickeln sind.

Der Entwicklungsbedarf betrifft

1. die Outcomeindikatoren selbst
2. das Verfahren der Risikoadjustierung
3. die Bewertungssystematik
4. das Verfahren der Datenerhebung in den Einrichtungen sowie
5. den Aufbau / die Auswahl einer Auswertungsstelle.

Jenseits der Empfehlung von Wingenfeld et al. (2011) hat die angestrebte gemeinsame Nutzung der Qualitätsdaten für interne und externe Qualitätsprozesse auch Konsequenzen für das interne Qualitätsmanagement. Dies betrifft die Datenerhebung ebenso wie die Datennutzung.

Weiterentwicklung der Outcomeindikatoren

Die Outcomeindikatoren sind hinsichtlich Auswahl und Ausgestaltung entsprechend den Ausführungen im Kapitel 4.1 weiter zu entwickeln. Dazu gehört einerseits die Frage der Konkretisierung wie am Beispiel des Schmerzmanagements oder der Operationalisierung wie am Beispiel der Mobilität. Es liegen weit akzeptierte Verfahrensweisen zur Indikatorenentwicklung vor (AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH, 2013). Diese sind bei der Weiterentwicklung zu beachten. Eine Expertengruppe ist mit der Weiterentwicklung der Indikatoren zu beauftragen. Dieser Expertengruppe sollten auch BetroffenenvertreterInnen angehören.

Verfahren der Risikoadjustierung

Es wird die Entwicklung einer Risikoadjustierungsmethode auf der Grundlage multipler Regression wie im Text beschrieben empfohlen. Ausgehend von den ausgewählten Indikatoren

- ist ein Literaturreview zu den Risikofaktoren für diese Outcome-Indikatoren zu erstellen,
- sind auf der Grundlage des Reviews repräsentative Aussagen darüber zu machen, welche Risikofaktoren jeweils eine Rolle spielen, um zu wissen welche Daten dann noch zusätzlich erhoben müssen,
- müssen ergänzend auch Messmethoden / Testverfahren recherchiert werden (dies wäre auch ggf. notwendig, wenn man nur bessere Testverfahren für die Risikogruppentrennung suchen würde),

- ist eine repräsentative Datengrundlage mit den identifizierten Risikofaktoren zu schaffen. Dies kann theoretisch auch nach einer bundesweiten Einführung geschehen, da man die Risikoadjustierung auch später noch auswerten kann und nicht von Beginn an mitdenken muss (sie ist nicht definitorischer Bestandteil der Indikatoren wie bei Wingefeld et al. (2011),
- sind die risikoadjustierten Maße auszuwerten / zu visualisieren bspw. zur Veröffentlichung auf der Webseite,

Als weiteren Schritt des Ausbaus der Risikoadjustierung könnte ggf. nachfolgend ein Hierarchisches Risikoadjustierungsmodell auf der Grundlage der bereits erhobenen Daten eingesetzt werden.

Bewertungssystematik

- Mit dem empfohlenen fünfstufigen System kann eigentlich schon auf Grundlage der vorhandenen Daten gestartet werden (aktuell aber nur normorientiert)
- Es wird die Entwicklung von Kriteriumsstandards mit folgenden Vorgehen empfohlen:
 - Bildung eines Expertengremiums aus ForscherInnen, Leitungen und Fachkräften
 - Übermittlung und Darstellung der empirischen Grundlagen (Verteilungen und ranggeordnete Ergebnisse der einzelnen Qualitätsindikatoren) als Grundinformation für die ExpertInnen. (Dies setzt Datenerhebung voraus).
 - Festlegung von oberen und unteren Schwellenwerten für die Fragen, wann Pflegeergebnisse nicht mehr akzeptabel sind und wann exzellente Qualität beginnt. Dies findet auf der Basis von Delphi-Prozessen (Ammon, 2009) statt. In einer ersten Befragungsrunde schätzen die ExpertInnen die Fragen unabhängig von der Gruppe auf Grundlage der Daten, aktueller Forschungstrends und eigener Erfahrungen ein. Die Ergebnisse werden dann in der Regel anonym an einen „Mediator“ übermittelt, der diese zusammenfassend auswertet (bspw. die mittleren Experten-Schwellenwerte für jeden Indikator).
 - In weiteren Runden (im vorliegenden Beispiel drei) wird die zusammengefasste „Gruppenmeinung“ an die ExpertInnen zurückgespiegelt. Die einzelnen Schwellenwerte werden dann so lange adjustiert, bis die Streuung der Einschätzung in der Gruppe soweit minimiert ist, dass ein akzeptabler Grad an Übereinstimmung erreicht ist.
- Kriteriumsstandards sollten alle 5-10 Jahre überarbeitet werden

Verfahren der Datenerhebung in den Einrichtungen

Als Grundlage für die Datenerhebung und Datenübermittlung wird die Herstellung der Bedingungen für eine edv-gestützte Datenerhebung und –übermittlung empfohlen.

Erforderlich sind dazu

- die verpflichtende Einführung einer edv-gestützten Pflegedokumentation,

- die Sicherstellung der Kompatibilität der Schnittstelle zur Übermittlungssoftware,
- die Herstellung eines Konsenses bzgl. eines einheitlichen Dateiformates (z.B. Health Level 7), das den Austausch der erbetenen Daten ermöglicht,
- die Integration der erforderlichen Items für die Outcomemessung in die Dokumentationssoftware.

Bereits heute gibt es verschiedene Aktivitäten von Softwarefirmen für die Erstellung elektronischer Pflegedokumentationen. Wenn der Konsens bzgl. des Dateiformates hergestellt ist, könnte voraussichtlich die technische Umsetzung relativ rasch erfolgen.

Aufbau / Auswahl einer Auswertungsstelle

Das Datenmanagement für die externe Outcomemessung, zu dem auch die Überwachung der Datenqualität gehört, sollte von einem unabhängigen Institut übernommen werden. Der Auftrag für dieses Institut ist zu spezifizieren. Beispielsweise ist zu klären, welche Rechte und Pflichten dieses Institut gegenüber den Einrichtungen hat, ob die Veröffentlichung der Outcomedaten ebenfalls in sein Aufgabengebiet fällt, wie das Verhältnis zum MDK gestaltet wird usw. Ein geeignetes Institut ist auszuwählen und zu beauftragen.

Literatur

- Abel, R. L., Warren, K., Bean, G., Gabbard, B., Lyder, C. H., Bing, M. et al. (2005). Quality improvement in nursing homes in Texas: results from a pressure ulcer prevention project. *J Am Med Dir Assoc*, 6, 181-188.
- Abraham, I. L., MacDonald, K. M., & Nadzam, D. M. (2006). Measuring the quality of nursing care to Alzheimer's patients. *Nurs Clin North Am*, 41, 95-104.
- ActiZ, V&N, LOC, NVVA, Sting, IGZ et al. (2007). Quality framework responsible care. <http://www.biomedcentral.com/content/supplementary/1472-6963-10-95-s1.pdf> (Abruf: 03.03.2014) [Electronic version].
- AGE Platform Europe (2010). Europäische Charta der Rechte und Pflichten älterer hilfe- und pflegebedürftiger Menschen. http://www.ede-eu.org/pdf/projects/DE_European_Charter_Daphne.pdf (Abruf: 03.05.2014) [Electronic version].
- Ammon, U. (2009). Delphi-Befragung. In S.Kühl, P. Strodtholz, & A. Taffertshofer (Eds.), *Handbuch Methoden der Organisationsforschung* (pp. 458-476). Wiesbaden: VS Verlag für Sozialwissenschaften.
- AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH (2010). Allgemeine Methoden im Rahmen der sektorenübergreifenden Qualitätssicherung im Gesundheitswesen nach § 137a SGB V. http://www.aqua-institut.de/aqua/upload/CONTENT/Projekte/137a/Methodenpapier/AQUA_Allgemeine_Methoden_Version_2-0.pdf (Abruf: 02.05.2014) [Electronic version].
- AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH (2013). Qualitätsreport 2012. <https://www.sgg.de/sgg/upload/CONTENT/Qualitaetsberichte/2012/AQUA-Qualitaetsreport-2012.pdf> (Abruf: 10.05.2014) [Electronic version].
- Arling, G., Kane, R. L., Lewis, T., & Mueller, C. (2005). Future development of nursing home quality indicators. *Gerontologist*, 45, 147-156.
- Arling, G., Lewis, T., Kane, R. L., Mueller, C., & Flood, S. (2007). Improving quality assessment through multilevel modeling: the case of nursing home compare. *Health Serv Res*, 42, 1177-1199.
- Australian Institute of Health and Welfare (2009). Towards national indicators of safety and quality in health care. <http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442457172> (Abruf: 12.03.2014) [Electronic version].
- Bahrman, A., Abel, A., Specht-Leible, N., Abel, A., Worz, E., Holscher, E. et al. (2010). Behandlungsqualität bei geriatrischen Patienten mit Diabetes mellitus in verschiedenen häuslichen Versorgungsstrukturen. *Z Gerontol Geriatr*, 43, 386-392.
- Balzer, K., Junghans, A., Behncke, A., & Lühmann, D. (2013). Literaturanalyse Expertenstandard Sturzprophylaxe in der Pflege. http://www.wiso.hs-osnabrueck.de/fileadmin/groups/607/Literaturstudie_Akt_Sturzprophylaxe_Langfassung.pdf (Abruf: 10.05.2014) [Electronic version].

- Bartholomeyczik, S., Reuther, S., Luft, L., van Nie, N., Meijers, J., Schols, J. et al. (2010). Prävalenz von Mangelernährung, Maßnahmen und Qualitätsindikatoren in deutschen Altenpflegeheimen - erste Ergebnisse einer landesweiten Pilotstudie. *Gesundheitswesen*, 72, 868-874.
- Bates-Jensen, B. M., Alessi, C. A., Cadogan, M., Levy-Storms, L., Jorge, J., Yoshii, J. et al. (2004). The minimum data set bedfast quality indicator: differences among nursing homes. *Nurs Res*, 53, 260-272.
- Bates-Jensen, B. M. & MacLean, C. H. (2007). Quality indicators for the care of pressure ulcers in vulnerable elders. *J Am Geriatr Soc*, 55, 409-416.
- Bensch, S. (2011). Konstruktvalidität der Module 'Mobilität' und Kognitive und kommunikative Fähigkeiten' des Neuen Begutachtungsassessments zur Feststellung von Pflegebedürftigkeit. http://www.pthv.de/fileadmin/user_upload/PDF_Pflege/Abschlussarbeiten/PTHV-PW_Dissertation_Bensch.pdf (Abruf: 10.06.2012) [Electronic version].
- Bensch, S. (2013). Das neue Begutachtungs-Assessment (NBA): Messen die Module "Mobilität" und "Kognitive und kommunikative Fähigkeiten" Pflegebedürftigkeit? *Pflegewissenschaft*, 15, 608-618.
- Bergquist-Beringer, S. & Gajewski, B. J. (2011). Outcome and assessment information set data that predict pressure ulcer development in older adult home health patients. *Adv Skin Wound Care*, 24, 404-414.
- Berlowitz, D. R., Rosen, A. K., Wang, F., Tsilimingras, D., Tariot, P. N., Engelhardt, J. et al. (2005). Purchasing or providing nursing home care: can quality of care data provide guidance. *J Am Geriatr Soc*, 53, 603-608.
- Bostick, J. E., Rantz, M. J., Flesner, M. K., & Riggs, C. J. (2006). Systematic review of studies of staffing and quality in nursing homes. *J Am Med Dir Assoc*, 7, 366-376.
- Bouman, A. I. E., Ettema, T. P., Wetzels, R. B., van Beek, A. P., de Lange J., & Dröes, R. M. (2011). Evaluation of Qualidem: a dementia-specific quality of life instrument for persons with dementia in residential settings; scalability and reliability of subscales in four Dutch field surveys. *Int J Geriatr Psychiatry*, 26, 711-722.
- Bundesministerium für Familie, S. F. u. J. & Bundesministerium für Gesundheit (2010). Charta der Rechte hilfe- und pflegebedürftiger Menschen. <http://www.pflege-charta.de/> (Abruf: 17.03.2014) [Electronic version].
- Bundesministerium für Gesundheit (2006). Rahmenempfehlungen zum Umgang mit herausforderndem Verhalten bei Menschen mit Demenz in der stationären Altenhilfe. https://www.bundesgesundheitsministerium.de/fileadmin/dateien/Publikationen/Pflege/Berichte/Bericht_Rahmenempfehlungen_zum_Umgang_mit_herausforderndem_Verhalten_bei_Menschen_mit_Demenz_in_der_stationaeren_Altenhilfe.pdf (Abruf: 03.04.2014) [Electronic version].
- Burfield, A. H., Wan, T. T., Sole, M. L., & Cooper, J. W. (2012). Behavioral cues to expand a pain model of the cognitively impaired elderly in long-term care. *Clin Interv Aging*, 7, 207-223.
- Canadian Institute for Health Information (2013). When a nursing home is home: how do Canadian nursing homes measure up on quality?

- https://secure.cihi.ca/free_products/CCRS_QualityinLongTermCare_EN.pdf (Abruf: 12.03.2014) [Electronic version].
- Canadian Nurses Association (2008). Mapping Canadian clinical outcomes in ICNP. http://c.hobic.cna-aiic.ca/documents/pdf/ICNP_Mapping_2008_e.pdf (Abruf: 02.04.2014) [Electronic version].
- Capezuti, E., Boltz, M., Renz, S., Hoffman, D., & Norman, R. G. (2006). Nursing home involuntary relocation: clinical outcomes and perceptions of residents and families. *J Am Med Dir Assoc*, 7, 486-492.
- Carlisle, R. T. (2007). Internet report cards on quality: what exists and the evidence on impact. *W V Med J*, 103, 17-21.
- Castle, N. G. & Ferguson, J. C. (2010). What is nursing home quality and how it is measured? *Gerontologist*, 50, 426-442.
- Center for Medicare and Medicaid Services (2011). OASIS-C guidance manual. <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/HHQIOASISUserManual.html> (Abruf: 12.06.2012) [Electronic version].
- Center for Medicare and Medicaid Services (2012). Design for nursing home compare five-star quality rating system: technical users' guide. <http://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/CertificationandCompliance/downloads/usersguide.pdf> (Abruf: 28.05.2014) [Electronic version].
- Centers for Medicare & Medicaid Services (2013). MDS quality measures. User's manual. <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Downloads/MDS-30-QM-User%E2%80%99s-Manual-V80.pdf> (Abruf: 02.05.2012) [Electronic version].
- Centers for Medicare & Medicaid Services (2014). Quality measures. <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/NHQIQualityMeasures.html> (Abruf: 03.03.2014) [Electronic version].
- Chodosh, J., Edelen, M. O., Buchanan, J. L., Yosef, J. A., Ouslander, J. G., Berlowitz, D. R. et al. (2008). Nursing home assessment of cognitive impairment: development and testing of a brief instrument of mental status. *J Am Geriatr Soc*, 56, 2069-2075.
- Compas, C., Hopkins, K. A., & Townsley, E. (2008). Best practices in implementing and sustaining quality of care. A review of the quality improvement literature. *Res Gerontol Nurs*, 1, 209-216.
- Cox, D. R. & Wermuth, N. (1994). Tests of linearity, multivariate normality and the adequacy of linear scores. *Statistics*, 43, 347-355.
- Dellefield, M. E. (2007). Implementation of the resident assessment instrument/minimum data set in the nursing home as organization: implications for quality improvement in RN clinical assessment. *Geriatr Nurs*, 28, 377-386.
- Department of Health (2010). Essence of care 2010. Benchmarks for the fundamental aspects of care.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216691/dh_119978.pdf (Abruf: 15.03.2014) [Electronic version].

Department of Health (2012). The NHS outcomes framework 2013-14. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213055/121109-NHS-Outcomes-Framework-2013-14.pdf (Abruf: 03.03.2014) [Electronic version].

DNQP (2009). *Expertenstandard Pflege von Menschen mit chronischen Wunden*. Osnabrück: Fachhochschule Osnabrück.

DNQP (2010a). *Expertenstandard Dekubitusprophylaxe in der Pflege*. (1. Aktualisierung ed.) Osnabrück.

DNQP (2010b). Expertenstandard Ernährungsmanagement zur Sicherstellung und Förderung der oralen Ernährung in der Pflege. Entwicklung - Konsentierung - Implementierung. <http://www.wiso.hs-osnabrueck.de/38093.html> (Abruf: 20.04.2014) [Electronic version].

DNQP (2011a). *Expertenstandard Schmerzmanagement in der Pflege bei akuten Schmerzen. 1. Aktualisierung*. Osnabrück: Hochschule Osnabrück.

DNQP (2011b). Methodisches Vorgehen zur Entwicklung, Einführung und Aktualisierung von Expertenstandards in der Pflege. http://www.wiso.hs-osnabrueck.de/fileadmin/groups/607/DNQP_Methodenpapier.pdf (Abruf: 02.05.2014) [Electronic version].

DNQP (2013). *Expertenstandard Sturzprophylaxe in der Pflege*. (1. Aktualisierung ed.) Osnabrück.

DNQP (2014a). Aktuelle Veröffentlichungen des DNQP. <http://www.dnqp.de> (Abruf: 12.02.2014) [Electronic version].

DNQP (2014b). Expertenstandard Förderung der Harnkontinenz in der Pflege. <http://www.wiso.hs-osnabrueck.de/fileadmin/groups/607/AktKo-Auszuglnet.pdf> (Abruf: 01.06.2014) [Electronic version].

DNQP (2014c). *Expertenstandard Schmerzmanagement in der Pflege bei chronischen Schmerzen*. Osnabrück: Hochschule Osnabrück.

Donabedian, A. (1980). *The definition of quality and approaches to its assessment. Explorations in quality assessment and monitoring*. (vols. 1) Ann Arbor, Mich: Health Administration Press.

Donald, F., Martin-Misener, R., Carter, N., Donald, E. E., Kaasalainen, S., Wickson-Griffiths, A. et al. (2013). A systematic review of the effectiveness of advanced practice nurses in long-term care. *J Adv Nurs*, 69, 2148-2161.

Doran, D., Mildon, B., & Clarke, S. (2011). Towards a national report card in nursing: a knowledge synthesis. *Nurs Leadersh (Tor Ont)*, 24, 38-57.

Doran, D. I., O'Brien-Pallas, L. L., Sidani, S., Hall, L. M., Petryshen, P., Hawkins, J. et al. (2003). An evaluation of nurse sensitive outcomes for quality care. *Int Nurs Perspect*, 3, 109-125.

- Dunckley, M., Aspinall, F., Addington-Hall, J. M., Hughes, R., & Higginson, I. J. (2005). A research study to identify facilitators and barriers to outcome measure implementation. *Int J Palliat Nurs*, 11, 218-225.
- Ettema, T. P., Drees, R. M., de Lange J., Mellenbergh, G. J., & Ribbe, M. W. (2007a). QUALIDEM: development and evaluation of a dementia specific quality of life instrument-validation. *Int J Geriatr Psychiatry*, 22, 424-430.
- Ettema, T. P., Drees, R. M., de Lange J., Mellenbergh, G. J., & Ribbe, M. W. (2007b). QUALIDEM: development and evaluation of a dementia specific quality of life instrument. Scalability, reliability and internal structure. *Int J Geriatr Psychiatry*, 22, 549-556.
- Europäisches Zentrum für Wohlfahrtspolitik und Sozialforschung et al. (2010). Verbesserung messen. Ergebnisorientierte Qualitätsindikatoren für Alten- und Pflegeheime. http://www.euro.centre.org/data/progress/PROGRESS_GERMAN.pdf (Abruf: 03.05.2014) [Electronic version].
- Expertenkreis Qualitätsindikatoren für Nationale VersorgungsLeitlinien (2009). Qualitätsindikatoren. Manual für Autoren. <http://www.aezg.de/mdb/edocs/pdf/schriftenreihe/schriftenreihe36.pdf> (Abruf: 02.05.2014) [Electronic version].
- Faust, R. (2004). Indikatoren zur Beurteilung pflegesensitiver Ergebnisqualität in Pflegeheimvergleichen. Chancen RAI MDS basierte Qualitätsindikatoren für eine pflegerische Qualitätsmessung in Deutschland. *PR-Internet für die Pflege*, 6, 222-230.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*, 12, 189-198.
- Gajewski, B. J., Mahnken, J. D., & Dunton, N. (2008). Improving quality indicator report cards through Bayesian modeling. *BMC Med Res Methodol*, 8:77.
- Gebert, A. & Kneubühler, H.-U. (2003). *Qualitätsbeurteilung und Evaluation der Qualitätssicherung in Pflegeheimen*. (2 ed.) Bern: Huber.
- Geraedts, M., Brechtel, T., Zöll, R., & Hermeling, P. (2012). Beurteilungskriterien für die Auswahl einer Pflegeeinrichtung. In J. Böcken, B. Braun, & U. Repschläger (Eds.), *Gesundheitsmonitor 2011. Bürgerorientierung im Gesundheitswesen. Kooperationsprojekt der Bertelsmann Stiftung und der BARMER GEK* (pp. 155-172). Gütersloh: Verlag Bertelsmann Stiftung.
- Gnass, I., Bartoszek, G., Thiesemann, R., & Meyer, G. (2010). Erworbene Kontraktoren der Gelenke im höheren Lebensalter. Eine systematische Literaturanalyse. *Z Gerontol Geriatr*, 43, 147-157.
- Goodson, J., Jang, W., & Rantz, M. (2008). Nursing home care quality: insights from a Bayesian network approach. *Gerontologist*, 48, 338-348.
- Grando, V. T., Rantz, M. J., & Maas, M. (2007). Nursing home staff's views on quality improvement interventions: a follow-up study. *J Gerontol Nurs*, 33, 40-47.
- Gräske, J., Fischer, T., Kuhlmei, A., & Wolf-Ostermann, K. (2012). Quality of life in dementia care - differences in quality of life measurements performed by residents with dementia and by nursing staff. *Aging Ment Health*, 16, 819-827.

- Großschädl, F. & Lohrmann, C. (2011). Druck in der Pflege. Ergebnisse einer Prävalenzerhebung zu Dekubitus in österreichischen Krankenhäusern und Pflegeheimen. *Pflege Z*, 64, 356-361.
- Hasseler, M. (2011). Ergebnisqualität und Lebensqualität in stationären Einrichtungen: Welche Indikatoren und Instrumente brauchen wir? Kommentar aus Sicht der Wissenschaft. <http://www.mds-ev.de/media/pdf/6 - Statement Hasseler.pdf> (Abruf: 05.05.2014) [Electronic version].
- Hasseler, M. & Fünfstück, M. (2012). Die Erstellung und Erprobung von Qualitätsberichten nach § 12 des Landesgesetzes über Wohnformen und Teilhabe (LWTG) in Rheinland-Pfalz. Bericht zur wissenschaftlichen Begleitung. (Langversion). http://www.sozialportal.rlp.de/uploads/media/Abschlussbericht_Qualitaetsberichte_Langversion.pdf (Abruf: 14.05.2014) [Electronic version].
- Hasseler, M. & Wolf-Ostermann, K. (2010). *Wissenschaftliche Evaluation zur Beurteilung der Pflege-Transparenzvereinbarungen für den ambulanten (PTVA) und stationären (PTVS) Bereich*.
- Haywood, K. L., Garratt, A. M., & Fitzpatrick, R. (2005). Quality of life in older people: a structured review of generic self-assessed health instruments. *Qual Life Res*, 14, 1651-1668.
- Health Information and Quality Authority (2009). National quality standards for residential care settings for older people in Ireland. http://www.higa.ie/system/files/HIQA_Residential_Care_Standards_2008.pdf (Abruf: 12.03.2014) [Electronic version].
- Health Information and Quality Authority (2013). Guidance on developing key performance indicators and minimum data sets to monitor healthcare quality. www.higa.ie/system/files/Hi_KPI_Guidelines.pdf (Abruf: 03.03.2014) [Electronic version].
- Health Quality & Safety Commission (2012). Describing the quality of New Zealand's health and disability services: developing our health quality and safety indicators. www.hqsc.govt.nz/assets/Health-Quality-Evaluation/PR/HQSI-indicators-summary-report-Dec-2012.pdf (Abruf: 03.04.2014) [Electronic version].
- Health Quality Ontario (2013). Indicator technical specifications. Quality improvement plan 2014/15. http://health.gov.on.ca/en/pro/programs/ecfa/legislation/qualityimprove/qip_tech.pdf (Abruf: 03.03.2014) [Electronic version].
- Health Quality Ontario (2014a). Information about long-term care quality indicators. <http://www.hqontario.ca/public-reporting/long-term-care/information-about-quality-indicators> (Abruf: 10.05.2014).
- Health Quality Ontario (2014b). Quality compass. Navigate from knowledge to action. <http://qualitycompass.hqontario.ca/language/en-us/home/subtopics/consistency-of-assignment/overviewd071.html> (Abruf: 15.05.2014) [Electronic version].
- Health Quality Ontario (2014c). Quality compass. Navigate from knowledge to action. <http://qualitycompass.hqontario.ca/> (Abruf: 04.06.2014) [Electronic version].

- Health Service Executive (2010). DML integrated minimum data set. <http://hse.ie/eng/services/publications/olderpeople/dmlmds.pdf> (Abruf: 12.03.2014) [Electronic version].
- Helsedirektoratet (2013). Najsonalt kvalitetsindikatorsystem. Statusrapport fra arbeidsgruppen for pleie og omsorg. <http://helsedirektoratet.no/publikasjoner/nasjonalt-kvalitetsindikatorsystem-statusrapport-fra-arbeidsgruppen-for-pleie-og-omsorg/Publikasjoner/statusrapport%20fra%20arbeidsgruppen%202012%20endelig%20versjon.pdf> (Abruf: 03.03.2014) [Electronic version].
- Helsedirektoratet (2014a). About IPLOS - a register for statistics on municipal health and care services. <http://helsedirektoratet.no/kvalitet-planlegging/iplos-registeret/about-iplos/Sider/default.aspx> (Abruf: 13.03.2014) [Electronic version].
- Helsedirektoratet (2014b). Najsonalt kvalitetsindikatorsystem. Statusrapport fra arbeidsgruppen for pleie og omsorg desember 2013. <http://helsedirektoratet.no/publikasjoner/nasjonalt-kvalitetsindikatorsystem/Publikasjoner/Nasjonalt%20kvalitetsindikatorsystem.pdf> (Abruf: 03.03.2014) [Electronic version].
- Hjaltadottir, I., Ekwall, A. K., Nyberg, P., & Hallberg, I. R. (2012). Quality of care in Icelandic nursing homes measured with minimum data set quality indicators: retrospective analysis of nursing home data over 7 years. *Int J Nurs Stud*, *49*, 1342-1353.
- Horn, S. D., Buerhaus, P., Bergstrom, N., & Smout, R. J. (2005). RN staffing time and outcomes of long-stay nursing home residents: pressure ulcers and other adverse outcomes are less likely as RNs spend more time on direct patient care. *Am J Nurs*, *105*, 58-71.
- Josat, S. (2005). Welche Qualitätskriterien sind Angehörigen in der stationären Altenpflege wichtig? *Pflege*, *18*, 169-175.
- Josat, S., Schubert, H.-J., Schnell, M. W., & Köck, C. (2006). Qualitätskriterien, die Altenpflegeheimbewohnern und Angehörigen wichtig sind. Eine Literaturanalyse. *Pflege*, *19*, 79-87.
- Kelley, E. & Hurst, J. (2006). Health care quality indicators project. Conceptual framework paper. <http://www.oecd.org/els/health-systems/36262363.pdf> (Abruf: 05.05.2014) [Electronic version].
- Kirkevold, O. & Engedal, K. (2008). Quality of care in Norwegian nursing homes - deficiencies and their correlates. *Scand J Caring Sci*, *22*, 560-567.
- Konetzka, R. T., Polsky, D., & Werner, R. M. (2013). Shipping out instead of shaping up: rehospitalization from nursing homes as an unintended effect of public reporting. *J Health Econ*, *32*, 341-352.
- Köpke, S., Gerlach, A., Möhler, R., Haut, A., & Meyer G (2012). Leitlinie FEM. <http://www.leitlinie-fem.de/download/LeitlinieFEM.pdf> (Abruf: 12.05.2014) [Electronic version].
- Kottner, J., Dassen, T., & Lahmann, N. A. (2011). Dekubitus in deutschen Pflegeheimen. Häufigkeiten, Grade und Entstehungsorte. *Z Gerontol Geriatr*, *44*, 318-322.

- Kunz, R., Kalid, K., Kleinen, J., & Antes, G. (2009). *Systematische Übersichtsarbeiten und Meta-Analysen*. (2 ed.) Bern: Huber.
- Lahmann, N., Kottner, J., Lützenkendorf, D., Raeder, K., Rahn, Y., Schmitz, G. et al. (2012). Pflegeprobleme in Deutschland. Ergebnisse von 12 Jahren Forschung in Pflegeheimen und Kliniken 2001-2012. http://igpw.charite.de/pflegewissenschaft/forschungsschwerpunkte/pflegeprobleme_in_deutschland/ (Abruf: 10.05.2014) [Electronic version].
- Li, Y., Cai, X., Mukamel, D. B., & Glance, L. G. (2010). The volume-outcome relationship in nursing home care: an examination of functional decline among long-term care residents. *Med Care*, *48*, 52-57.
- Li, Y., Schnelle, J., Spector, W. D., Glance, L. G., & Mukamel, D. B. (2010). The "nursing home compare" measure of urinary/fecal incontinence: cross-sectional variation, stability over time, and the impact of case mix. *Health Serv Res*, *45*, 79-97.
- Lienert, G. A. & Ratz, U. (1998). *Testaufbau und Testanalyse*. (6 ed.) Weinheim: Beltz PVU.
- Löcherbach, P., Hermsen, T., Macsenaere, M., Arnold, J., & Klein, J. (2010). Evaluationsstudie zur Patientenbegleitung der Bosch BKK. http://www.kh-mz.de/ifib/pdf/Evaluationsstudie_BKK.pdf (Abruf 02.06.2014) [Electronic version].
- MDS & GKV (2014). Qualitätsprüfungs-Richtlinien. Transparenzvereinbarung. <http://www.mds-ev.de/media/pdf/> (Abruf: 05.05.2014) [Electronic version].
- Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen (MDS) (2012). 3. Bericht des MDS nach § 114a Abs. 6 SGB XI. http://www.mds-ev.de/media/pdf/MDS_Dritter_Pflege_Qualitaetsbericht_Endfassung.pdf (Abruf: 10.04.2014) [Electronic version].
- Meißner, A. & Althammer, T. (2012). *Pflegedokumentation mit EDV*. Hannover: Vincentz.
- Mitchell, P. H., Ferketich, S., & Jennings, B. M. (1998). Quality health outcomes model. American academy of nursing expert panel on quality health care. *Image J.Nurs.Sch*, *30*, 43-46.
- Moorhead, S., Johnson, M., & Mass, M. (2004). *Nursing Outcomes Classification (NOC)*. (3 ed.) St. Louis: Mosby.
- Mor, V. (2005). Improving the quality of long-term care with better information. *Milbank Q*, *83*, 333-364.
- Mor, V. (2006). Defining and measuring quality outcomes in long-term care. *J Am Med Dir Assoc*, *7*, 532-540.
- Mor, V., Angelelli, J., Gifford, D., Morris, J., & Moore, T. (2003). Benchmarking and quality in residential and nursing homes: lessons from the US. *Int J Geriatr Psychiatry*, *18*, 258-266.
- Mor, V., Gruneir, A., Feng, Z., Grabowski, D. C., Intrator, O., & Zinn, J. (2011). The effect of state policies on nursing home resident outcomes. *J Am Geriatr Soc*, *59*, 3-9.
- Mor, V., Intrator, O., Unruh, M. A., & Cai, S. (2011). Temporal and geographic variation in the validity and internal consistency of the nursing home resident assessment Minimum Data Set 2.0. *BMC Health Serv Res*, *11*:78.

- Mueller, C. & Karon, S. L. (2004). ANA nurse sensitive quality indicators for long-term care facilities. *J Nurs Care Qual*, 19, 39-47.
- MuG stat. (2011). Bekanntmachung der "Maßstäbe und Grundsätze für die Qualität und die Qualitätssicherung sowie für die Entwicklung eines einrichtungsinternen Qualitätsmanagements nach § 113 des Elften Buches Sozialgesetzbuch (SGB XI) in der vollstationären Pflege vom 27.Mai 2011" sowie der "Anlage nach Ziffer 5 (ambulant) bzw. 7 (stationär) der Maßstäbe und Grundsätze für die Qualität und die Qualitätssicherung sowie für die Entwicklung eines einrichtungsinternen Qualitätsmanagements nach § 113 SGB XI in der ambulanten und stationären Pflege in Bezug auf die Anforderungen an die Zuverlässigkeit, Unabhängigkeit und Qualifikation von Prüfinstitutionen und unabhängigen Sachverständigen nach § 114 Absatz 4 SGB XI sowie die methodische Verlässlichkeit von Zertifizierungs- und Prüfverfahren". *Bundesanzeiger*, 2573-2580.
- Mukamel, D. B., Glance, L. G., Li, Y., Weimer, D. L., Spector, W. D., Zinn, J. S. et al. (2008). Does risk adjustment of the CMS quality measures for nursing homes matter? *Med Care*, 46, 532-541.
- Mukamel, D. B., Peterson, D. R., Bajorska, A., Temkin-Greener, H., Kunitz, S., Gross, D. et al. (2004). Variations in risk-adjusted outcomes in a managed acute/long-term care program for frail elderly individuals. *Int J Qual Health Care*, 16, 293-301.
- National Aged Care Alliance (2014). Quality indicators for residential care advisory paper. <http://www.naca.asn.au/Publications/NACA%20Advisory%20Paper%20-%20Quality%20Indicators%20in%20Residential%20Care%20-%20Jan%202014.pdf> (Abruf: 29.05.2014) [Electronic version].
- National Institute for Health and Care Excellence (2014). Indicators process guide. http://www.nice.org.uk/media/03E/31/Indicators_process_guide.pdf (Abruf: 15.05.2014) [Electronic version].
- National Quality Forum (2011). National voluntary consensus standards for nursing homes. A consensus report. http://www.qualityforum.org/Publications/2011/07/National_Voluntary_Consensus_Standards_for_Nursing_Homes.aspx (Abruf: 15.05.2014) [Electronic version].
- Naylor, M. D. (2007). Advancing the science in the measurement of health care quality influenced by nurses. *Med Care Res Rev*, 64, 144S-169S.
- Nazir, A., Arling, G., Perkins, A. J., & Boustani, M. (2011). Monitoring quality of care for nursing home residents with behavioral and psychological symptoms related to dementia. *J Am Med Dir Assoc*, 12, 660-667.
- Nazir, A., Mueller, C., Perkins, A., & Arling, G. (2012). Falls and nursing home residents with cognitive impairment: new insights into quality measures and interventions. *J Am Med Dir Assoc*, 13, 819.e1-819.e6.
- Netten, A., Trukeschitz, B., Beadle-Brown, J., Forder, J., Towers, A. M., & Welch, E. (2012). Quality of life outcomes for residents and quality ratings of care homes: is there a relationship? *Age Ageing*, 41, 512-517.
- Nygaard, H. A. (2008). Measuring body mass index (BMI) in nursing home residents: the usefulness of measurement of arm span. *Scand J Prim Health Care*, 26, 46-49.

- O'Connor, M. & Davitt, J. K. (2012). The Outcome and Assessment Information Set (OASIS): a review of validity and reliability. *Home Health Care Serv Q*, 31, 267-301.
- O'Reilly, M., Courtney, M., Edwards, H., & Hassall, S. (2011). Clinical outcomes in residential care: setting benchmarks for quality. *Australas J Ageing*, 30, 63-69.
- Ollenschläger, G. (2007). Institutionalisation der Qualitätsentwicklung in der Pflege. Gutachten für die Bundeskonferenz zur Qualitätssicherung im Gesundheits- und Pflegewesen e.V. (BUKO-QS). http://buko-qs.de/cms/upload/pdf/GAGO_BUKOQS_0704final.pdf (Abruf: 04.06.2014) [Electronic version].
- Olschewski, U. (2009). Ernährung und Flüssigkeitsversorgung von Bewohnern in stationären Pflegeeinrichtungen aus der Sicht von Angehörigen und Betreuern. Eine Follow-Up-Studie. *Pflegewissenschaft*, 11, 1-6.
- Paquay, L., De Lepeleire, J., Schoenmakers, B., Ylief, M., Fontaine, O., & Buntinx, F. (2007). Comparison of the diagnostic accuracy of the cognitive performance scale (minimum data set) and the mini-mental state exam for the detection of cognitive impairment in nursing home residents. *Int J Geriatr Psychiatry*, 22, 286-293.
- Pesis-Katz, I., Phelps, C. E., Temkin-Greener, H., Spector, W. D., Veazie, P., & Mukamel, D. B. (2013). Making difficult decisions: the role of quality of care in choosing a nursing home. *Am J Public Health*, 103, e31-e37.
- Pfaff, H. & Schrappe, M. (2011). Einführung in die Versorgungsforschung. In H. Pfaff, E. Neugebauer, G. Glaeske, & M. Schrappe (Eds.), *Lehrbuch Versorgungsforschung. Systematik, Methodik, Anwendung* (pp. 1-39). Stuttgart: Schattauer.
- Phillips, C. D., Chen, M., & Sherman, M. (2008). To what degree does provider performance affect a quality indicator? The case of nursing homes and ADL change. *Gerontologist*, 48, 330-337.
- Phillips, C. D., Shen, R., Chen, M., & Sherman, M. (2007). Evaluating nursing home performance indicators: an illustration exploring the impact of facilities on ADL change. *Gerontologist*, 47, 683-689.
- Popp, J., Pröfener, F., Stappenbeck, J., Reintjes, R., & Weber, P. (2006). Der Einfluss des Fachkräfteanteils auf die Dekubitusinzidenz in Pflegeheimen. *Pflege*, 19, 303-307.
- Pringle, D. & Doran, D. (2003). Patient outcomes as an accountability. In *Nursing sensitive outcomes - state of the science* (pp. 1-26). Sudbury, Massachusetts: Jones and Bartlett.
- Rantz, M. J., Cheshire, D., Flesner, M., Petroski, G. F., Hicks, L., Alexander, G. et al. (2009). Helping nursing homes "at risk" for quality problems: a statewide evaluation. *Geriatr Nurs*, 30, 238-249.
- Reiter, A., Fischer, B., Kötting, J., Geraedts, M., Jäckel, W. H., Barlag, H. et al. (2014). QUALIFY: Ein Instrument zur Bewertung von Qualitätsindikatoren. <http://online.abo.bqs-online.com/download/qualify/qualify-down.pdf> (Abruf: 10.06.2013) [Electronic version]. Available: <http://online.abo.bqs-online.com/download/qualify/qualify-down.pdf>

- Rondas, A. A., Schols, J. M., Stobberingh, E. E., & Halfens, R. J. (2013). Prevalence of chronic wounds and structural quality indicators of chronic wound care in Dutch nursing homes. *Int.Wound.J.*
- RTI International (2012). MDS 3.0 quality measures. User's manual. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/downloads/MDS30QM-Manual.pdf> (Abruf: 05.05.2014) [Electronic version].
- Saliba, D., Buchanan, J., Edelen, M. O., Streim, J., Ouslander, J., Berlowitz, D. et al. (2012). MDS 3.0: brief interview for mental status. *J Am Med Dir Assoc*, 13, 611-617.
- Sangl, J., Saliba, D., Gifford, D. R., & Hittle, D. F. (2005). Challenges in measuring nursing home and home health quality: lessons from the first national healthcare quality report. *Med Care*, 43, I-24-I-32.
- Schaeffer, D. (2004). *Der Patient als Nutzer. Krankheitsbewältigung und Versorgungsnutzung im Verlauf chronischer Krankheit*. Bern: Verlag Hans Huber.
- Schneekloth, U. & von Törne, I. (2007). Entwicklungstrends in der stationären Versorgung - Ergebnisse der Infratest-Repräsentativerhebung. In U.Schneekloth & H. W. Wahl (Eds.), (München: Bundesministerium für Familie, Senioren, Frauen und Jugend.
- Schnelle, J. F., Bates-Jensen, B. M., Levy-Storms, L., Grbic, V., Yoshii, J., Cadogan, M. et al. (2004). The minimum data set prevalence of restraint quality indicator: does it reflect differences in care? *Gerontologist*, 44, 245-255.
- Sektion for Klinisk Information & Data (2014). Landsdaekkende database for geriatri (Landesweite Datenbank für Geriatrie). http://www.regionh.dk/kliniskedatabaser/menu/Nationale+databaser/Landsdaekkende_database_for_Geriatri.htm (Abruf: 03.03.2014) [Electronic version].
- Stemmer, R. (2005). Zur aktuellen Lage der pflegesensitiven Outcomemessung - eine Übersicht. *PR-Internet für die Pflege*, 7, 42-47.
- Stemmer, R. & Arnold, J. (2012). *Kurzexpertise zur "Bewertung von Qualitätsindikatoren in der ambulanten und stationären Langzeitpflege" mit besonderer Berücksichtigung der vom Institut für Pflegewissenschaft der Universität Bielefeld entwickelten und erprobten Instrumente zur Beurteilung der Ergebnisqualität in der stationären Altenhilfe*". o.V.
- Stengel, D., Bhandari, M., & Hanson, B. (2011). *Statistik und Aufbereitung klinischer Daten*. Stuttgart: Thieme.
- Straker, J. K. & Bailer, A. J. (2008). A review and characterization of the MDS process in nursing homes. *J Gerontol Nurs*, 34, 36-44.
- Tannen, A., Bours, G., Halfens, R., & Dassen, T. (2006). A comparison of pressure ulcer prevalence rates in nursing homes in the Netherlands and Germany, adjusted for population characteristics. *Res Nurs Health*, 29, 588-596.
- The Scottish Government (2008). Leading better care. Report of the senior charge nurse review and clinical quality indicators project. <http://www.scotland.gov.uk/Resource/Doc/225218/0060938.pdf> (Abruf: 03.03.2014) [Electronic version].

- Tolson, D., Rolland, Y., Andrieu, S., Aquino, J. P., Beard, J., Benetos, A. et al. (2011). International association of gerontology and geriatrics: a global agenda for clinical research and quality of care in nursing homes. *J Am Med Dir Assoc*, 12, 184-189.
- Troyer, J. L. & Thompson, H. G. Jr. (2004). The impact of litigation on nursing home quality. *J Health Polit Policy Law*, 29, 11-42.
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Publishing Company.
- Vahakangas, P., Noro, A., Finne-Soveri, H., & Bjorkgren, M. (2008). Association between rehabilitation care practices and care quality in long-term care facilities. *J Nurs Care Qual*, 23, 155-161.
- Wagner, L. M., McDonald, S. M., & Castle, N. G. (2012). Joint commission accreditation and quality measures in U.S. nursing homes. *Policy Polit Nurs Pract*, 13, 8-16.
- Weibler-Villalobos, U. & Röhrig, B. (2010). Methodische Anforderungen an einrichtungsbezogene Qualitätsberichte in der Pflege. *Gesundheitswesen*, 72, 780-789.
- Weidner, F., Laag, U., & Brühl, A. (2011). Evaluation der Umsetzung der Pflege-Transparenzvereinbarung ambulant (PTVA) durch den MDK in Rheinland-Pfalz. http://www.dip.de/fileadmin/data/pdf/material/110530_dip-MDK-Abschlussbericht.pdf (Abruf: 07.06.2014) [Electronic version].
- Wilson, A. & Twiss, A. (1999). Using outcomes and OASIS as strategic tools. *Home Healthc Nurse Manag*, 3, 19-25.
- Windeler, J., Görres, S., Thomas, S., Kimmel, A., Langner, I., Reif, K. et al. (2011). Maßnahmen zur Schaffung eines neuen Pflegebedürftigkeitsbegriffs und eines Begutachtungsinstrumentes zur Feststellung der Pflegebedürftigkeit. http://www.gkv-spitzenverband.de/upload/GKV-Schriftenreihe_Pflege_Band_3_18963.pdf (Abruf: 01.06.2012) [Electronic version].
- Wingenfeld, K., Büscher, A., & Gansweid, B. (2008). *Das neue Begutachtungsassessment zur Feststellung von Pflegebedürftigkeit*. Bielefeld; Münster.
- Wingenfeld, K., Engels, D., Engel, H., Franz, S., Kleina, T., & Mehlan, S. (2011). *Entwicklung und Erprobung von Instrumenten zur Beurteilung der Ergebnisqualität in der stationären Altenhilfe*. Bielefeld; Köln.
- Wu, N., Miller, S. C., Lapane, K., Roy, J., & Mor, V. (2005a). Impact of cognitive function on assessments of nursing home residents' pain. *Med Care*, 43, 934-939.
- Wu, N., Miller, S. C., Lapane, K., Roy, J., & Mor, V. (2005b). The quality of the quality indicator of pain derived from the minimum data set. *Health Serv Res*, 40, 1197-1216.
- Zinn, J., Spector, W., Hsieh, L., & Mukamel, D. B. (2005). Do trends in the reporting of quality measures on the nursing home compare web site differ by nursing home characteristics? *Gerontologist*, 45, 720-730.

Anlage: Übersicht über die Fragestellungen der Expertise

Themenbereich 1: Reichweite der Indikatoren

- Bildet das Indikatorenset von Wingenfeld die für eine vergleichende Qualitätsberichterstattung relevanten Pflegebereiche in der erforderlichen Themenbreite und –tiefe hinreichend ab? Auf welche von Wingenfeld et al. vorgeschlagenen Themenbereiche/ Indikatoren kann verzichtet werden, welche sollten ergänzend berücksichtigt werden? Wenn relevante Themenbereiche nicht abgedeckt sind, wie könnten diese einbezogen werden, wenn eine Indikatorenbildung methodisch nicht valide ist?
- Bestimmte Personen werden aus der Bewertung ausgeschlossen. Sind die jeweils definierten Ausschlusskriterien (z. B. Schlaganfall oder akute Verschlechterung einer Gesundheitsstörung), die zum Ausschluss aus der Bewertung führen, akzeptabel oder führen sie zu einer systematischen Verzerrung der Qualitätsberichterstattung?

Themenbereich 2: Datengrundlage/ Externe Qualitätsprüfung

- Sind für eine Bewertung der Indikatoren neben der Pflegedokumentation weitere Datenquellen notwendig?
- Wie kann erreicht werden, dass die Daten des internen Qualitätsmanagements flächendeckend eine zuverlässige und valide Datengrundlage liefern?
- Können mit den von Wingenfeld et al. vorgeschlagenen Plausibilitätsprüfungen im Rahmen externer Qualitätsprüfungen die Daten über die Qualität der Pflege verifiziert werden?

Themenbereich 3: Bewertungssystematik

- Stellen die von Wingenfeld et al. gegebenen Empfehlungen zur Bewertungssystematik eine Form der Bewertung von Indikatoren/ Indikatorenbereichen dar, die eine Differenzierung von Einrichtungen ermöglicht und sind die vorgenommenen Setzungen ausreichend fachlich begründet?
- Gibt es Alternativen, z. B. die ergänzende Berücksichtigung von zu erfüllenden Mindestniveaus, und wie werden diese bewertet?

Themenbereich 4: Risikoadjustierung

- Sind die der Risikoadjustierung zugrunde gelegten Kriterien (Items aus dem NBA) angemessen, um eine valide Risikoadjustierung zu ermöglichen? Gibt es Hinweise oder Vorschläge zu Alternativen oder zur Weiterentwicklung?

Themenbereich 5: Verbraucherperspektive

- Welche Indikatoren sind aus der Verbraucherperspektive relevant?
- Sind Indikatoren bzw. Empfehlungen von Wingenfeld et al. für eine verständliche, übersichtliche und vergleichbare Information der Verbraucher geeignet? Ist die ver-

fügbare Information in ihrer Auswahl hinreichend, in ihrer Bewertung und Darstellung verständlich und erlaubt sie einen Vergleich der Einrichtungen/ Dienste, insbesondere bezogen auf Ergebnisqualität? Wenn nein, welche Voraussetzungen, Änderungen, Ergänzungen wären hierfür notwendig?

- Sind die Empfehlungen von Wingenfeld et al. zur Bewertungssystematik auf Ebene der Indikatoren und der Indikatorenbereiche aus Verbraucherperspektive geeignet, eine hinreichende Differenzierung der Einrichtungen zu ermöglichen? Sind hierfür voraussichtlich fortlaufende Nachjustierungen am Bewertungssystem notwendig?

Abbildungsverzeichnis

Abbildung 1: Vorgehen Datenbankrecherche.....18

Tabellenverzeichnis

Tabelle 1: Thematische Hinweise für die grundsätzliche inhaltliche Passung18

Tabelle 2: Codiersystematik der Abstracts20

Tabelle 3: BQS-QUALIFY und Beurteilungsraster Expertise21

Tabelle 4: Qualitätsmessung nach Ländern26

Tabelle 5: Übersicht Ausschlusskriterien bei den Wingenfeld-Indikatoren.....58

Tabelle 6: Übersicht Bewertungssystematik bei den Wingenfeld-Indikatoren.....73

Tabelle 7: Übergeordnete Bewertungssystematik der Wingenfeld-Qualitätsbereiche.....76

Tabelle 8: Übersicht Risikoadjustierungskriterien bei den Wingenfeld-Indikatoren89

Tabelle 9: Kritische Punkte und Anforderungen an eine systematische
Qualitätsberichterstattung (Hasseler & Fünfstück, 2012).....110