

**Die Intra-Cluster-Correlation  
im Kontext der Fallzahlschätzung  
cluster-randomisierter Studien  
in der Pflegewissenschaft**

INAUGURAL-DISSERTATION  
ZUR ERLANGUNG DES  
DOKTORGRADES DER PFLEGEWISSENSCHAFT  
(DR. RER. CUR.)

AN DER PFLEGEWISSENSCHAFTLICHEN FAKULTÄT DER  
PHILOSOPHISCH-THEOLOGISCHEN HOCHSCHULE VALLENDAR

BAND I

vorgelegt von:  
Marion Kowe

Erstguterachter:  
Prof. Dr. Albert Brühl

Zweitgutachterin:  
Prof. Dr. Sandra Bensch

vorgelegt im Juli 2019



# Inhaltsverzeichnis

<b>Band I</b>	<b>3</b>
<b>Abbildungsverzeichnis</b>	<b>3</b>
<b>Tabellenverzeichnis</b>	<b>3</b>
<b>Abkürzungsverzeichnis</b>	<b>4</b>
<b>Logische Operatoren</b>	<b>4</b>
<b>1 Einleitung</b>	<b>5</b>
<b>2 Thematische Einordnung</b>	<b>6</b>
2.1 Hinführung zum Thema . . . . .	6
2.2 Problembeschreibung . . . . .	8
2.2.1 Einfluss der Intra-Cluster-Correlation (ICC) auf die Fallzahl . . . . .	8
2.2.2 Relevanz des Themas . . . . .	11
2.3 Ziel dieser Arbeit . . . . .	15
<b>3 Theoretische Grundlagen</b>	<b>15</b>
3.1 Wissenschaftstheoretischer Hintergrund . . . . .	15
3.1.1 Festlegung der Begrifflichkeiten . . . . .	16
3.1.2 Aristoteles . . . . .	16
3.1.3 Kritischer Rationalismus . . . . .	16
3.1.4 Logischer Empirismus . . . . .	18
3.1.5 Strukturalismus . . . . .	19
3.2 Testtheorien zum statistischen Testen von Hypothesen . . . . .	21
3.2.1 Theorie des Signifikanztests von R.A. Fisher (FST) . . . . .	21
3.2.2 Theorie des statistischen Hypothesentestens von J. Neyman und E.S. Pearson (NPT) . . . . .	22
3.2.3 Hybridisierung von FST und NPT . . . . .	23
3.3 Fallzahlschätzung und Teststärkenanalyse . . . . .	24
3.4 Cluster-Randomisierung und ICC . . . . .	26
3.4.1 Clusterrandomisierung . . . . .	26
3.4.2 ICC . . . . .	27
<b>4 Datenanalyse</b>	<b>28</b>
4.1 Datenerhebung . . . . .	28
4.1.1 Stichprobe . . . . .	28
4.1.2 Variablenauswahl . . . . .	29
4.2 Datenauswertung . . . . .	40
4.2.1 Das Common-Correlation-Modell . . . . .	40
4.2.2 Varianzanalyse-Schätzer . . . . .	40
4.2.3 Pearson-Schätzer . . . . .	41
4.2.4 Moment-Schätzer . . . . .	43

4.2.5	probabilistischer Schätzer . . . . .	43
<b>5</b>	<b>Ergebnis</b>	<b>45</b>
5.1	ICC der einzelnen Variablen . . . . .	46
5.2	Streuung der ICC-Werte . . . . .	46
5.3	Vergleich der Schätzer . . . . .	46
5.4	Einfluss der Erfolgsrate auf die ICC . . . . .	48
5.5	Simulationsstudie . . . . .	51
<b>6</b>	<b>Ergebnisse anderer Studien</b>	<b>52</b>
6.1	Ulmer Sturz-Studie . . . . .	52
6.2	Fixierung . . . . .	55
<b>7</b>	<b>Diskussion</b>	<b>61</b>
7.1	Studiendesign . . . . .	61
7.2	Hypothesentesten - Möglichkeiten und Grenzen . . . . .	61
7.3	ICC-Schätzer . . . . .	62
7.4	ICC-Werte . . . . .	63
7.5	Empfehlungen . . . . .	65
7.6	Ausblick . . . . .	67
<b>8</b>	<b>Zusammenfassung</b>	<b>67</b>
	<b>Literaturverzeichnis</b>	<b>70</b>

## Band I

### Abbildungsverzeichnis

1	Verweildauer der Patienten . . . . .	9
2	Verweildauer der Patienten im Vergleich der Stationen . . . . .	10

### Tabellenverzeichnis

1	Einteilung der Evidenz-Stärke von Interventionsstudien . . . . .	7
2	Beispielberechnung für die Fallzahl $n$ bei verschiedenen Clustergrößen $m$ und ihre Auswirkung auf die Anzahl der Cluster $k$ . . . . .	11
3	Einfluss von ICC $\rho$ und Clustergröße $m$ auf den Design-Effekt . . . . .	12
4	Wahrscheinlichkeiten für richtige und falsche Entscheidungen über das Zutreffen einer empirischen Hypothese (EH) . . . . .	20
5	Wahrscheinlichkeiten für richtige und falsche Entscheidungen über das Beibehalten einer statistischen Hypothese $H_0$ . . . . .	22
6	Das Vorgehen unter den beiden konkurrierenden Testtheorien FST und NPT	23
7	Design-Effekt $DE$ für $\rho = 0,01$ für unterschiedliche Clustergrößen . . . . .	27
8	Verteilung der Größe der Wohnbereiche . . . . .	29
10	kleinste ICC-Schätzer nach Wohnbereichs-Größe . . . . .	47
11	größte ICC-Schätzer nach Wohnbereichs-Größe . . . . .	47
12	kleinste ICC-Schätzer nach Erfolgsrate . . . . .	49
13	größte ICC-Schätzer nach Erfolgsrate . . . . .	50
14	Anzahl der Kombinationen . . . . .	51
15	Beschreibung der Studienteilnehmer . . . . .	54
16	Inzidenz von Stürzen und Frakturen in der Studienpopulation . . . . .	54
17	Schätzungen für die ICC der Stürze . . . . .	55
18	Schätzungen für die ICC der Frakturen . . . . .	55
19	Merkmale der Altenheime zu Beginn der Querschnitt-Studie . . . . .	56
20	Häufigkeit mechanischer Fixierungen . . . . .	57
21	Psychopharmaka . . . . .	58
22	Merkmale der Altenheime zu Studienbeginn . . . . .	58
23	Prävalenz von Fixierungen . . . . .	60

## Abkürzungsverzeichnis

<b>ASE</b>	Atemstimulierende Einreibung
<b>AWMF</b>	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
<b>ÄZQ</b>	Ärztliche Zentralstelle Qualitätssicherung
<b>Bew.</b>	Bewohner/in
<b>DE</b>	Design-Effekt
<b>EH</b>	empirische Hypothese
<b>FEM</b>	freiheitsentziehende Maßnahmen
<b>FST</b>	Theorie des Signifikanztests von R.A. Fisher
<b>ICC</b>	Intra-Cluster-Correlation
<b>NA</b>	not available
<b>NPT</b>	Theorie des statistischen Hypothesentestens von J. Neyman und E.S. Pearson
<b>RCT</b>	randomisierte kontrollierte Studie
<b>SPG</b>	Saarländische Pflegegesellschaft
<b>TS</b>	Teststatistik
<b>WB</b>	Wohnbereich

## Logische Operatoren

$\wedge$	und
$\rightarrow$	impliziert
$\neg$	nicht
$\vdash$	logische Beweisbarkeit

# 1 Einleitung

Zum Nachweis der Wirksamkeit von Interventionen werden in der Regel kontrollierte Studien durchgeführt; idealerweise werden die Probanden hierbei zufällig der Interventions- oder der Kontrollgruppe zugewiesen.

Bei Studien, die pflegerische Interventionen zum Gegenstand haben, wird häufig nicht über die Individuen randomisiert, sondern über die jeweilige Pflegeeinheit, z.B. den Wohnbereich oder die Station. Dafür gibt es verschiedene Gründe (Donner, Klar, 2000; Kuß et al., 2009):

- Praktikabilität bzw. Wirtschaftlichkeit

Die Interventionen müssen von allen beteiligten Pflegekräften durchgeführt werden, d.h. es wäre ein unverhältnismäßig hoher Aufwand, für jeweils einen Probanden alle Pflegekräfte entsprechend zu schulen.

Besteht die Intervention in einer Schulung der Probanden, kann es aus ökonomischen Gründen sinnvoll sein, die Schulung in Gruppen durchzuführen.

Beruhet die Intervention auf einer die gesamte Organisationseinheit betreffenden Veränderung, z.B. der Einführung eines neuen Verpflegungskonzeptes, wäre es unwirtschaftlich, diese in mehreren Einrichtungen für jeweils einen Probanden durchzuführen.

- Gefahr der Kontamination

Wenn Kontrollgruppe und Interventionsgruppe derselben Untersuchungseinheit angehören, besteht die Gefahr der Kontamination, wenn beispielsweise Probanden oder Angehörige miteinander in Kontakt kommen und sich über die Intervention austauschen. So erfährt der Proband aus der Kontrollgruppe eventuell von der Maßnahme und setzt sie um.

Zudem bestünde die Gefahr, dass Pflegekräfte versehentlich die zu untersuchende Intervention bei Probanden der Kontrollgruppe durchführen.

- Ethische Gründe

Falls Kontrollgruppe und Interventionsgruppe derselben Untersuchungseinheit angehören, müssten Pflegekräfte eine Maßnahme, von der sie annehmen, dass sie positive Effekte hat, einem Teil der Patienten oder Bewohner vorenthalten.

Ein Einfluss der Clusterzugehörigkeit auf die zu beobachtenden Effekte kann nicht ausgeschlossen werden, so dass von einer Unabhängigkeit der Ergebnisse nicht mehr ausgegangen werden kann (Kuß et al., 2009).

Der Nachweis des Effekts einer Intervention erfolgt in der Regel über den Vergleich von Mittelwerten und inwieweit sie signifikant voneinander abweichen. Der Anteil der Clusterzugehörigkeit an dieser Differenz ist schon bei der Planung einer Studie zu berücksichtigen und in Form der Intra-Cluster-Correlation (ICC) in die Berechnung der erforderlichen Fallzahl einzubeziehen (Kuß et al., 2009).

In Kapitel 2.2.1 wird dargestellt, welchen Einfluss die aus der Cluster-Randomisierung resultierende ICC auf die erforderliche Stichprobengröße hat. Da die ICC vor Beginn der

Studie nicht vorliegt, muss auf Werte aus bereits durchgeführten vergleichbaren Studien zurück gegriffen werden. Der Versuch, diese im Rahmen einer Literatur-Recherche zu ermitteln, wird in Kapitel 2.2.2 beschrieben.

Für ein vertieftes Verständnis der Methode des Hypothesentestens und der zugrunde zu legenden Werte für das Signifikanz-Niveau  $\alpha$  und den Fehler 2. Art  $\beta$  bzw. die Power  $1 - \beta$  einer Studie werden die wissenschaftstheoretische Entwicklung sowie die daraus hervorgegangenen Testtheorien in Kapitel 3 ausführlich dargelegt. Auch der Zusammenhang zwischen Fallzahlschätzung, Teststärkenanalyse und der Notwendigkeit der Berücksichtigung der ICC wird dort erläutert.

Um Forschenden für zukünftige Studien die für die Berechnung der erforderlichen Stichprobengröße benötigte ICC zur Verfügung zu stellen, wurde eine Sekundär-Analyse der im Rahmen der am Lehrstuhl für Statistik und standardisierte Methoden für Pflegeforschung der Philosophisch-Theologischen Hochschule Vallendar im Auftrag der Saarländischen Pflegegesellschaft e.V. durchgeführten PiSaar-Studie (Brühl, Planer; 2013) erhobenen Daten durchgeführt. Aus den Daten wurde mit verschiedenen Schätzern die ICC für 62 Wohnbereiche und 53 bewohnerbezogene Merkmale bestimmt und verglichen. Das zugrunde liegende Modell sowie die verwendeten Schätzer werden in Kapitel 4 beschrieben.

Die Auswertung erfolgte sowohl über alle Wohnbereiche als auch getrennt nach Wohnbereichsgröße und Erfolgsquote. Desweiteren wurde eine Simulationsstudie durchgeführt, indem die ICC für jede mögliche Kombination aus  $k = 2$  sowie  $k = 3$  Clustern berechnet wurde. Die Ergebnisse sind in Kapitel 5 dargestellt. Die aus der Literatur ermittelten Werte für die ICC finden sich in Kapitel 6.

In Kapitel 7 werden Methode und Ergebnisse dieser Arbeit diskutiert und Empfehlungen für den Umgang mit den hier ermittelten Werten sowie für die eigene Berechnung der ICC gegeben.

## 2 Thematische Einordnung

### 2.1 Hinführung zum Thema

Es gibt zahlreiche wissenschaftliche Methoden zum Nachweis der Wirksamkeit pflegerischer Interventionen. Die Wahl des Paradigmas - qualitativ oder quantitativ - sowie der Methode zur Datenerhebung und -auswertung sind dabei unter anderem abhängig vom gewählten Endpunkt, der Zielgruppe und nicht zuletzt von den Rahmenbedingungen, unter denen die Untersuchung stattfinden soll.

Inwieweit es sinnvoll ist, die Evidenzklassifikation der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) und der Ärztlichen Zentralstelle Qualitätssicherung (ÄZQ) (siehe Tabelle 1) für die Beurteilung quantitativer Studien im Rahmen der Pflegeforschung zugrunde zu legen, soll hier nicht weiter diskutiert werden. Letztlich werden die Forschungsarbeiten nach diesen Kriterien bewertet – zumindest dann, wenn sie in Fachzeitschriften veröffentlicht werden sollen, in denen die Beiträge einem Peer-Review unterzogen werden.



Grad	ÄZQ
Ia	Evidenz aufgrund von Metaanalysen randomisierter kontrollierter Studien
Ib	Evidenz aufgrund mindestens einer randomisierten kontrollierten Studie
IIa	Evidenz aufgrund mindestens einer gut angelegten kontrollierten Studie ohne Randomisierung
IIb	Evidenz aufgrund mindestens einer gut angelegten quasi-experimentellen Studie
III	Evidenz aufgrund gut angelegter, nicht experimenteller deskriptiver Studien (z.B. Vergleichsstudien, Korrelationsstudien, Fall-Kontrollstudien)
IV	Evidenz aufgrund von Berichten/Meinungen von Expertenkreisen, Konsensus-Konferenzen und/oder Erfahrung anerkannter Autoritäten

Tabelle 1: Einteilung der Evidenz-Stärke von Interventionsstudien - AWMF und ÄZQ, 2001 (aus Kottner, Tannen, 2015 S. 5)

Entscheidend ist, dass das Ergebnis einer Studie unabhängig von deren Qualität eine Pflegefachkraft nicht davon entbindet, situative Kriterien bei der Entscheidungsfindung für oder gegen eine Intervention in einer spezifischen Situation zu berücksichtigen. Daher ist es unerlässlich, Studienergebnisse nicht nur zu kennen, sondern diese auch entsprechend bewerten zu können (Sackett et al., 1996)

Um von allen Pflegenden genutzt werden zu können, müssen Studienergebnisse in entsprechenden Datenbanken zur Verfügung stehen und im Rahmen systematischer Literaturrecherchen gefunden werden können. Schlömer (1999) hat im Rahmen einer Literaturstudie untersucht, inwieweit von Pflegenden durchgeführte deutschsprachige Studien in den Datenbanken Medline und Cinahl zu finden sind. Sie zeigt, dass deutschsprachige RCTs mit pflegerischem Bezug mittels systematischer Literaturrecherche nicht zu ermitteln sind. Dies liege zum einen daran, dass Studien in den entsprechenden Datenbanken meist inhaltlich und weniger methodisch indiziert werden. Zum anderen werden in Deutschland viele Studien im Rahmen von Qualifizierungsarbeiten durchgeführt, die in diese Datenbanken nicht aufgenommen werden.

Um in diese Datenbanken aufgenommen zu werden, müssen Studien bestimmte Kriterien erfüllen. Für therapeutische und präventive Interventionen gelten randomisierte kontrollierte Studien als Goldstandard.

Im Verlauf einer Recherche über randomisierte kontrollierte Studien im Rahmen der Pflegeforschung kristallisierte sich folgendes Spezifikum für pflegerische Interventionsstudien heraus: Die zu untersuchenden Personen werden nicht zufällig und unabhängig voneinander ausgewählt, sondern sind in der Regel Teil einer bestehenden Gruppe, sogenannte Cluster, z.B. ein Wohnbereich in einem Altenheim oder eine Station in einem Krankenhaus. Sie erhalten somit nicht nur die gleiche Intervention sondern unterliegen auch denselben Rahmenbedingungen, die ihrerseits Einfluss auf den Endpunkt haben können. Diese Ähnlichkeit zwischen den Individuen einzelner Cluster wird als Intra-Cluster-Correlation (ICC) bezeichnet und ist nicht nur bei der Datenauswertung zu berücksichtigen, sondern ist auch ein wichtiger Parameter in der vor Beginn der Studie durchzuführenden Fallzahlschätzung (Kron, 2003)

Die A-priori-Fallzahlschätzung ist erforderlich um sicher zu stellen, dass der erwartete Effekt auch tatsächlich nachgewiesen werden kann. Dieser Zusammenhang wird in Kapitel 3.3 umfassend dargelegt. Von der erforderlichen Fallzahl hängt wiederum ab, ob eine Studie mit den vorhandenen Ressourcen überhaupt zielführend durchführbar ist (Erdfelder et al., 2004).

## 2.2 Problembeschreibung

### 2.2.1 Einfluss der ICC auf die Fallzahl

Die Notwendigkeit der Berücksichtigung der ICC bei der Fallzahlschätzung soll an einem fiktiven Beispiel dargestellt werden. Die Erläuterung des zugrunde liegenden Modells sowie der Fallzahlberechnung erfolgt in den Kapiteln 4.2.1 und 3.3.

Beispiel: In einer orthopädischen Klinik beträgt die durchschnittliche Verweildauer nach einem bestimmten Eingriff 10 Tage. Mit einem neuen Mobilisationskonzept soll die Verweildauer reduziert werden. Da die Anwendung des Konzeptes mit mehr Zeitaufwand für die Pflegekräfte verbunden ist, muss sich die Verweildauer um mindestens 1 Tag reduzieren, um wirtschaftlich vertretbar zu sein.

Das Konzept wird nun probeweise auf einigen Stationen eingeführt. Anschließend wird verglichen, ob sich die Verweildauer der Patienten dieser Stationen von denen der Vergleichsstationen unterscheidet. Dabei werden zunächst die allgemeinen Konventionen mit einem Signifikanzniveau von  $\alpha = 0,05$  und einer Power von 80% zugrunde gelegt. Nun stellt sich die Frage, wie viele Patienten in die Untersuchung einbezogen werden müssen, um den Erfolg des Konzeptes belegen zu können.

Für die Berechnung der Stichprobengröße wird folgende Formel für einen einseitigen Test herangezogen (siehe z.B. Donner, Klar, 2000, S. 57)

$$n = \frac{(z_{\alpha} + z_{\beta})^2 2(\sigma^2)}{(\mu_1 - \mu_2)^2}$$

Die Differenz zwischen den Stichprobenmittelwerten entspricht dem zu erwartenden Effekt von mindestens einem Tag Verweildauer. Für die Berechnung der Varianz wurde eine Zufallsstichprobe aus den Akten zweier Stationen gezogen mit einem Umfang von  $n=30$  Patienten pro Station. Das Ergebnis ist in der mit der Software SPSS<sup>®</sup> 23 erstellten Abbildung 1 dargestellt.

Somit ergibt sich eine Fallzahl von

$$\begin{aligned} n &= \frac{(1,65 + 0,84)^2 \cdot 2 \cdot 1,25}{1^2} \\ &= 15,5 \end{aligned}$$

Patienten pro Studienarm.

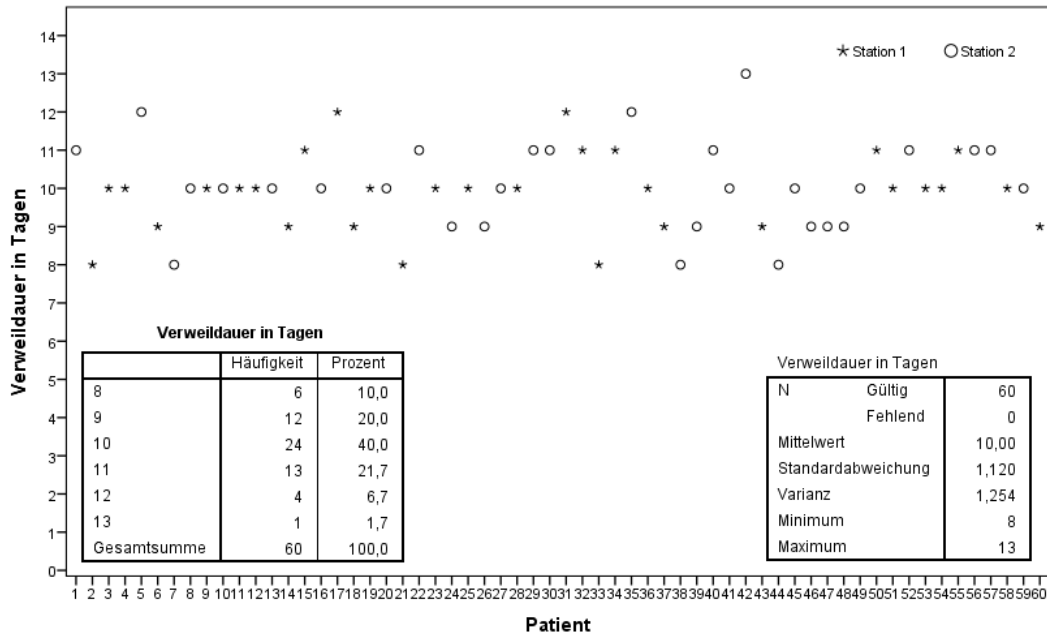


Abbildung 1: Verweildauer der Patienten

Zum Vergleich: Würde man davon ausgehen, dass eine Reduzierung der Verweildauer von einem halben Tag realistischer wäre, erhöht sich die Fallzahl auf

$$\begin{aligned}
 n &= \frac{(1,65 + 0,84)^2 \cdot 2 \cdot 1,25}{0,5^2} \\
 &= 62
 \end{aligned}$$

Patienten pro Studienarm.

Bereits hier wird die Notwendigkeit einer A-priori-Fallzahlschätzung deutlich.

Da nicht ausgeschlossen werden kann, dass die Verweildauer durch weitere stationsspezifische Faktoren mit beeinflusst wurde, ist Gruppenzugehörigkeit in der Fallzahlschätzung mit zu berücksichtigen.

Für die oben aufgeführten Gruppen ergibt sich der in Abbildung 2 dargestellte Unterschied: Der Anteil der Gruppenzugehörigkeit an der Gesamtvarianz wird mit der ICC angegeben. Sie wird als  $\rho$  bezeichnet und wie folgt berechnet (Donner, Klar, 2000):

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

wobei  $\sigma_b^2$  für die Varianz zwischen (**between**) und  $\sigma_w^2$  für die Varianz innerhalb (**within**) der Gruppen steht.

Hier ergibt sich eine ICC von

Verweildauer in Tagen			Station 1		
Station 1	Häufigkeit	Prozent	Verweildauer in Tagen		
8	3	10,0	N	Gültig	30
9	6	20,0		Fehlend	0
10	14	46,7	Mittelwert		9,90
11	5	16,7	Standardabweichung		1,029
12	2	6,7	Varianz		1,059
Gesamtsumme	30	100,0	Minimum		8
			Maximum		12

Verweildauer in Tagen			Station 2		
Station 2	Häufigkeit	Prozent	Verweildauer in Tagen		
8	3	10,0	N	Gültig	30
9	6	20,0		Fehlend	0
10	10	33,3	Mittelwert		10,10
11	8	26,7	Standardabweichung		1,213
12	2	6,7	Varianz		1,472
13	1	3,3	Minimum		8
Gesamtsumme	30	100,0	Maximum		13

Abbildung 2: Verweildauer der Patienten im Vergleich der Stationen

$$\begin{aligned}\rho &= \frac{0,3}{0,3 + 1,27} \\ &= 0,19\end{aligned}$$

Naturgemäß hat der Einfluss der Gruppenzugehörigkeit bei wenigen großen Clustern mehr Gewicht als bei vielen kleinen Clustern. Daher ist die Gruppengröße  $m$  ein Parameter, der im Vorfeld festgesetzt oder im Rahmen der Berechnung iterativ ermittelt werden muss.

Der Einfluss von Clustergröße und ICC wird als Design-Effekt (DE) oder Varianz-Inflations-Faktor bezeichnet (Donner, Klar, 2000):

$$DE = (1 + (m - 1)\rho)$$

und bei der Fallzahlschätzung wie folgt berücksichtigt (Donner, Klar, 2000) :

$$n = \frac{(z_\alpha + z_\beta)^2 2(\sigma^2)(1 + (m - 1)\rho)}{(\mu_1 - \mu_2)^2}$$

Wird  $\sigma^2$  durch mehr als eine Variable beeinflusst (hier: Intervention und Gruppenzugehörigkeit), wird

$$\sigma^2 = \sigma_w^2 + \sigma_b^2$$

gesetzt (Donner, Klar, 2000) .

Angenommen, im vorgesehenen Zeitraum können pro Station 20 Patienten in die Untersuchung aufgenommen werden. Dann benötigt man für den Nachweis einer Reduzierung der Verweildauer um einen Tag

$$\begin{aligned}
 n &= (1,65 + 0,84)^2 \cdot 2(0,3 + 1,27)(1 + (20 - 1)0,19) \\
 &= 89,76
 \end{aligned}$$

Patienten pro Studienarm. Bei einer Gruppengröße von 20 Patienten bedeutet dies, dass für die Interventionsgruppe sowie für die Kontrollgruppe jeweils mindestens  $k=5$  Stationen (Cluster) einbezogen werden müssen.

In Tabelle 2 sind zum Vergleich die Fallzahlen sowie die daraus resultierenden Clusterzahlen  $k$  pro Studienarm für unterschiedliche Clustergrößen und zwei verschiedene Effekte dargestellt. Der hier ermittelten ICC von 0,19 wurde zusätzlich die halbierte ICC gegenübergestellt, um zu zeigen, welchen Einfluss die ICC auf die erforderliche Stichprobengröße hat.

	ICC = 0,19				ICC = 0,095			
$\mu_1 - \mu_2 = 1$	$m$	10	20	30	$m$	10	20	30
	$n$	52,76	89,76	126,75	$n$	36,12	54,61	73,11
	$k$	6	5	5	$k$	4	3	3
$\mu_1 - \mu_2 = 0,5$	$m$	10	20	30	$m$	10	20	30
	$n$	211,05	359,03	507,00	$n$	144,47	218,45	292,44
	$k$	22	18	17	$k$	15	11	10

Tabelle 2: Beispielberechnung für die Fallzahl  $n$  bei verschiedenen Clustergrößen  $m$  und ihre Auswirkung auf die Anzahl der Cluster  $k$

Tabelle 3 zeigt, welchen Einfluss schon sehr kleine ICC insbesondere bei großen Clustern (z.B. ein Krankenhaus oder ein Altenheim) auf den Design-Effekt haben. Der Design-Effekt ist die Größe, um die die für eine unabhängige Stichprobe ermittelte Fallzahl erhöht werden muss.

Hier wird deutlich, dass bei einer Clustergröße von  $m = 30$  bereits eine kleine ICC von  $\rho = 0,005$  eine Erhöhung der Fallzahl um 14,5 % erfordert. Dies kann zur Folge haben, dass ein oder mehrere zusätzliche Cluster in die Untersuchung einbezogen werden müssen.

### 2.2.2 Relevanz des Themas

Es gibt verschiedene Methoden, die ICC aus den erhobenen Daten zu ermitteln(siehe Kapitel 4.2). Für die Fallzahlschätzung wird der Wert jedoch vor der Untersuchung benötigt,

		$\rho$				
		0,001	0,003	0,005	0,008	0,01
$m$	10	1,009	1,027	1,045	1,072	1,09
	15	1,014	1,042	1,07	1,112	1,14
	30	1,029	1,087	1,145	1,232	1,29
	50	1,049	1,147	1,245	1,392	1,49
	100	1,099	1,297	1,495	1,792	1,99

Tabelle 3: Einfluss von ICC  $\rho$  und Clustergröße  $m$  auf den Design-Effekt

d.h. es muss auf die Ergebnisse bereits durchgeführter ähnlicher Studien zurück gegriffen werden.

Diese Untersuchung ist darauf ausgelegt, insbesondere jene Pflegeforschenden bei der Fallzahlschätzung zu unterstützen, die im Rahmen kleinerer Forschungsvorhaben wie beispielsweise Qualifizierungsarbeiten nicht die Ressourcen haben, einen Biometriker mit dieser Berechnung zu beauftragen. Da die ICC hauptsächlich durch die Rahmenbedingungen beeinflusst wird, sollte sie unter ähnlichen Bedingungen ermittelt worden sein, um realistisch zu sein. Daher beschränkt sich die Literaturrecherche auf Studien, die im deutschsprachigen Raum durchgeführt wurden.

Zunächst sollte eine Literaturrecherche durchgeführt werden mit dem Ziel, Studien zu identifizieren, in deren Veröffentlichung die ICC berichtet wird, um diese im Rahmen einer Metaanalyse Pflegeforschenden zur Verfügung zu stellen.

Eine eigene systematische Literaturrecherche erbrachte dermaßen indifferente Resultate, dass die Ergebnisse nicht verwendet und hier auch nicht weiter aufgeführt werden sollen; es entspricht im Wesentlichen den Ergebnissen von Schlömer (1999). Die Recherche lieferte jedoch einen ersten Anhaltspunkt dafür, dass nach wie vor wenig pflegerische Interventionsstudien durchgeführt bzw. veröffentlicht werden und vor allem Fallzahlschätzung und ICC im Rahmen der Studien kaum thematisiert werden.

Stattdessen wurde eine Handsuche in den gängigen deutschsprachigen Pflegezeitschriften von 2003 bis 2013 mit folgenden Zielen durchgeführt:

1. Identifikation von Studien, in denen die ICC veröffentlicht wird
2. Identifikation von Studien, in denen Daten über Cluster erhoben wurden, so dass im Rahmen einer Sekundäranalyse die ICC ermittelt werden kann. (Dieses Ziel wurde ergänzt, nachdem sich sehr schnell gezeigt hatte, dass die ICC nicht berichtet wird.)

Hierfür wurden die Inhaltsverzeichnisse der Zeitschriften

- Pflege
- Plexus
- Pflegemagazin 2003 – 2005, danach aufgegangen in Pflege & Gesellschaft
- Pflege& Gesellschaft

- Nightingale 2003-2005, danach aufgegangen in Altenpflege
- Psych. Pflege Heute
- Pflegewissenschaft
- Procare
- Heilberufe science ab 2/2010
- Pflege aktuell 2003 bis 6/2006, danach aufgegangen in Die Schwester/Der Pfleger
- Die Schwester/Der Pfleger
- Zeitschrift für Pflegewissenschaft und psychische Gesundheit 2007-2010
- Reihe Pflegewissenschaft
- Pflegezeitschrift

online oder händisch in der Zentralen Fachbibliothek für Medizin und Gesundheitswesen in Deutschland (ZBMed) in Köln auf die Begriffe Effekt, Auswirkung, Intervention, Maßnahme, Studie, RCT, erhöhen, senken, verbessern, verschlechtern und ähnliche untersucht. In die Analyse wurden diejenigen Artikel aufgenommen, in denen eine Datenerhebung über Cluster erfolgt war. Hierbei zeigte sich, dass im Bereich der Pflegewissenschaft zunehmend auch Studien durchgeführt werden, die die Arbeitsbedingungen der Pflegekräfte und deren Folgen zum Gegenstand haben. Auch diese wurden in die Literaturrecherche einbezogen.

Insgesamt wurden 117 Artikel identifiziert, die den obigen Einschluss-Kriterien entsprechen. Thematisch unterteilen Sie sich wie folgt:

- Patient/Bewohner (67)
  - Schmerz (8)
  - Dekubitus/Wunde (18)
  - Ernährung (6)
  - Sturz (13)
  - (Im)Mobilität (2)
  - FEM (5)
  - Kinsäthetik (1)
  - ASE (3)
  - Inkontinenz (4)
  - Pflegebedürftigkeit/-aufwand (7)
  - Wissen (1)
  - Basale Stimulation (2)
  - Demenz (4)
  - Sonstige (15)
- Angehörige (7)
  - Belastung (2)
  - Wissen (1)

- Bedürfnisse/Zufriedenheit (3)
- Beteiligung (1)
- Pflegepersonal/Organisation (45)
  - Belastung/Gesundheit (30)
  - Assessment (4)
  - Wissen (2)
  - Selbstdarstellung/-wahrnehmung (2)
  - Umsetzung(2)
  - Qualität (1)
  - Zufriedenheit (2)
  - Sonstige (5)
- Sonstige/Allgemein (1)

Es wurden 53 Artikel aus der weiteren Analyse ausgeschlossen. Gründe dafür waren:

- Es handelt sich um einen Verweis auf oder einen Bericht über eine Studie, die dann an anderer Stelle berücksichtigt wurde.
- Die Studie wurde wiederholt; es liegen aktuellere Daten vor.
- Die untersuchte Intervention ermöglicht keine Berechnung der ICC, z.B. weil sich die Maßnahmen in den Clustern unterscheiden oder weil nicht über Pflegeeinheiten randomisiert wurde sondern über Erzähl-Cafés oder Krankenkassen-Geschäftsstellen.
- Es handelt sich um den Auszug einer an anderer Stelle berücksichtigten Studie.
- Es handelt sich um eine qualitative Studie.
- Der Artikel stammte zwar aus dem Untersuchungszeitraum, die Studie liegt jedoch länger zurück.
- Die Studie konnte nicht abgeschlossen werden.
- Forschungsgegenstand war ein Instrument.
- Der selbe Artikel wurde in einer anderen Zeitschrift veröffentlicht und dort bereits berücksichtigt.

Zusätzlich aufgenommen wurde der Artikel "Effect of a Guideline-Based Multicomponent Intervention on Use of Physical Restraints in Nursing Homes" (Köpke et. al., 2012), auf den der Text "Pflege ohne Freiheitsentzug ist machbar" (in: Die Schwester/Der Pfleger 51 (8)) verweist. In der hier dargestellten Studie wurden die ICC sowohl bei der Fallzahlschätzung berücksichtigt als auch aus der Studie bestimmt und berichtet. Die für die Fallzahlschätzung verwendeten Werte stammten aus der Studie "Restraint use among nursing home residents" (Meyer et al.,2009), die ebenfalls in diese Arbeit aufgenommen wurde.



Damit verblieben 65 Studien oder Artikel (siehe Anhang ??), die einer weiteren Analyse unterzogen wurden. Keine Studie veröffentlichte die ICC. In nur wenigen Studien wurde die Fallzahlberechnung beschrieben. Lediglich die Studie von Köpke et. al. (2012) bezieht die ICC aus einer vorangegangenen Studie in die Fallzahlberechnung ein und bestimmt sie auch für die eigenen Ergebnisse (siehe Kapitel 6).

Es ist nicht auszuschließen, dass bei dieser Art der Recherche einzelne Studien übersehen werden. Es ist jedoch zu erwarten, dass diese von Pflegeforschenden im Rahmen ihrer themenbezogenen Literaturrecherche berücksichtigt werden.

Die Literaturrecherche zeigt somit, dass die ICC höchst selten veröffentlicht wird. Pflegewissenschaftler/innen haben somit keine Referenzgröße für ihre Fallzahlschätzung.

### **2.3 Ziel dieser Arbeit**

Mit dieser Arbeit sollen ICC zu häufig beforschten pflegerelevanten Themen wie Demenz, Dekubitus, Sturz, Inkontinenz, Mobilität, Ernährung oder freiheitsentziehende Maßnahmen ermittelt und Pflegeforschenden als Grundlage für ihre Fallzahlschätzung zur Verfügung gestellt werden.

Die ursprüngliche Intention der Arbeit war die Analyse bereits veröffentlichter ICC nach dem Vorbild von Adams et al. (2004). Nachdem die Literaturrecherche gezeigt hat, dass ICC in der Pflegeforschung nicht in entsprechendem Ausmaß veröffentlicht werden, sollen die ICC im Rahmen einer Sekundäranalyse der in der Literaturrecherche identifizierten Studien ermittelt werden.

Hierfür sollen verschiedene in der Literatur beschriebene Schätzer vorgestellt und verglichen werden.

Pflegeforscher/innen, die kleine Studien mit geringen Ressourcen durchführen möchten, soll so ein Instrument an die Hand gegeben werden, mit dem sie mit möglichst geringem Aufwand eine Fallzahlschätzung durchführen und somit einschätzen können, ob der erwartete Effekt mit den verfügbaren Mitteln überhaupt nachgewiesen werden kann.

Die Arbeit will somit einen Beitrag dazu leisten, dass kleine Effekte in cluster-randomisierten Studien auch tatsächlich nachgewiesen werden können.

## **3 Theoretische Grundlagen**

### **3.1 Wissenschaftstheoretischer Hintergrund**

Die statistische Prüfung von Hypothesen ist eine vor allem in den Disziplinen der Psychologie und der Medizin häufig angewandte Methode zum Nachweis der Wirksamkeit von Interventionen. Dabei werden die Ergebnisse solcher Untersuchungen oft fehlinterpretiert als "statistische Beweise" (Hager, 2004). Für eine angemessene Anwendung des Verfahrens und die korrekte Interpretation der Ergebnisse ist es unerlässlich, den wissenschaftstheoretischen Hintergrund und die Entwicklung des Hypothesentestens zu betrachten. In diesem

Kapitel werden die historische Entwicklung des Verfahrens sowie seine Möglichkeiten und Grenzen beschrieben.

Hierbei geht es nicht um die Bewertung bestimmter wissenschaftstheoretischer Ansätze, sondern lediglich um ihre Darstellung zum Hintergrundverständnis für eine daraus entstandene Methode.

### **3.1.1 Festlegung der Begrifflichkeiten**

Carnap und Popper differenzieren in ihren Ausführungen nicht klar zwischen Hypothesen und Theorie. Im weiteren Verlauf wird so weit als möglich die Unterscheidung zugrunde gelegt zwischen Theorie "als systematisches Gefüge von Ideen und Annahmen über einen definierten Gegenstandsbereich" und Hypothesen als Annahmen oder Behauptungen, die aus Theorien abgeleitet werden und empirisch überprüft werden können. (Sedlmeier, Renkewitz, 2011)

### **3.1.2 Aristoteles**

Aristoteles hat ein Modell der wissenschaftlichen Erkenntnis entwickelt, wonach wahre Aussagen schrittweise aus gewissen ersten Prinzipien, sogenannten Axiomen, deduziert werden. Die Axiome ihrerseits sind nicht beweisbar, sie entstehen durch Erfahrung und Induktion, d.h. durch Erfahrung und Induktion wird von der Betrachtung vieler Einzelfälle auf die Allgemeingültigkeit von Regeln geschlossen. (Lauth, Sareiter, 2005)

Daraus haben sich zwei unterschiedliche erkenntnistheoretische Grundpositionen entwickelt – die empiristische Sicht, wonach Theorien induktiv aus Beobachtungen entwickelt werden, und die rationalistische Sicht, wonach Theorien den Untersuchungen vorangehen, d.h. aus Theorien Vorhersagen abgeleitet und diese anhand von Untersuchungsergebnissen überprüft werden. (Westermann, 2000)

### **3.1.3 Kritischer Rationalismus**

Das Schließen von besonderen Sätzen, die Beobachtungen beschreiben, auf allgemeine Sätze in Form von Theorien wird als Induktionsschluss bezeichnet. Das sich daraus ergebende Induktionsproblem lässt sich anhand zweier Fragen darstellen (Popper, Keuth, 2005):

1. Wann ist der Induktionsschluss berechtigt, d.h. wie viele Beobachtungen sind erforderlich, um auf einen allgemeinen Satz schließen zu können?
2. Was rechtfertigt den Induktionsschluss an sich? Seine Anwendung beruht auf der wiederholten Beobachtung, dass dieses Verfahren bisher erfolgreich angewendet wurde, ist also selbst ein Induktionsschluss.

Für Popper kann daher die Tätigkeit des wissenschaftlichen Forschers nicht mit dem Induktionsschluss enden; an das Aufstellen einer Theorie muss sich die Überprüfung anschließen,

inwieweit die Theorie nachprüfbar ist und inwieweit sie von anderen Theorien logisch abhängt oder mit ihnen in Widerspruch steht.

Für diese Überprüfung werden aus der Theorie auf logisch-deduktivem Weg Folgerungen abgeleitet, die sowohl untereinander als auch mit anderen Theorien auf logische Beziehungen wie Äquivalenz, Ableitbarkeit, Vereinbarkeit oder Widerspruch untersucht werden. Insgesamt nennt Popper vier Schritte zur Durchführung der Prüfung – nach der logischen Überprüfung der Theorie auf ihre innere Widerspruchslosigkeit zum einen und auf ihren Charakter einer wissenschaftlich-empirischen Theorie zum anderen folgen die Einschätzung, inwieweit die Theorie als wissenschaftlicher Fortschritt zu bewerten ist und zuletzt die Prüfung der abgeleiteten Forderungen durch empirische Anwendung.

Die Vorgehensweise im letztgenannten Prüfungsverfahren ist wieder deduktiv: Aus der Theorie werden Folgerungen abgeleitet, die auf bereits anerkannten Sätzen beruhen und möglichst leicht nachprüfbar oder anwendbar sind. Über diejenigen Folgerungen, die nicht aus bekannten Theorien ableitbar sind oder mit ihnen in Widerspruch stehen, wird in Experimenten oder in der praktischen Anwendung entschieden:

Eine positive Entscheidung bedeutet, dass die Folgerungen anerkannt, also verifiziert, werden. Die Theorie hat die Prüfung somit vorläufig bestanden und wird nicht verworfen - sie hat sich "bewährt". Es sei explizit darauf hingewiesen, dass die Theorie durch diese Entscheidung nicht als "wahr" oder "bewiesen" gilt, sie wird lediglich aufrecht erhalten, so lange sie nicht durch eine andere Theorie abgelöst wird.

Eine negative Entscheidung bedeutet, dass die Folgerungen falsifiziert werden und somit auch die Theorie, aus der die Folgerungen deduziert wurden.

Diese Schlussfolgerung von der Falsifizierung eines Folgesatzes auf die Falsifizierung der Theorie wird als "Modus tollens" bezeichnet (= aufhebende Schlussweise; eigentlich Modus tollendo tollens = durch Aufheben aufhebende Schlussweise). Er ist eine logische Schlussfigur mit folgender Aussage:

Sei  $A$  eine Theorie und  $B$  ein aus ihr abgeleiteter Folgesatz. Dann gilt " $A$  impliziert  $B$ " oder

$$A \rightarrow B$$

Zeigt sich nun in der empirischen Überprüfung, dass  $B$  nicht zutrifft, so impliziert dies, dass auch  $A$  nicht zutrifft - durch die Falsifizierung von  $B$  wird auch  $A$  falsifiziert:

$$[(A \rightarrow B) \wedge \neg B] \rightarrow \neg A$$

Es sei an dieser Stelle darauf hingewiesen, dass es nicht Ziel dieses Ansatz ist, Experimente durchzuführen, um eine Theorie zu widerlegen, wie es häufig als Kritik geäußert wird (z.B. Hager, 2004), sondern sie dadurch zu stärken, dass sie Falsifizierungsversuchen standhält. Eine Theorie muss sich bewähren - dies kann sie nur, wenn auch eine Möglichkeit des Scheiterns existiert. Je strengeren Überprüfungen eine Theorie standhält, um so stärker hat sie sich bewährt.

Popper betont den Unterschied zwischen der Falsifizierbarkeit einer Theorie als *logische* und somit endgültig entscheidbare Angelegenheit und der *empirischen* Falsifizierung einer

Theorie, die "wie jede empirische Angelegenheit unsicher und nicht endgültig entscheidbar ist". (Popper, Keuth, 2005) In einer empirischen Untersuchung haben viele Faktoren Einfluss auf das Ergebnis, die nicht der Theorie zugeordnet werden können. Insofern liegt es immer im begründeten Ermessen des Wissenschaftlers, inwieweit das Resultat seiner empirischen Überprüfung dann tatsächlich zur Falsifikation der Theorie führt oder auf die Rahmenbedingungen zurück geführt wird.

### 3.1.4 Logischer Empirismus

Empiristische Wissenschaftstheoretiker unterteilen Aussagen in analytisch determinierte und synthetische Aussagen. Die Wahrheit oder Falschheit analytisch determinierter Sätze bestimmt sich allein aus formalen logischen Gründen, d.h. sie sind gegenüber jeder Erfahrung immun, während Wahrheit oder Falschheit synthetischer Sätze allein durch Beobachtung begründet werden kann.

Nach Carnap sind synthetische Aussagen dann wissenschaftlich sinnvoll, wenn sie sich mit Hilfe von Beobachtungen vollständig bestätigen, also verifizieren, lassen (Carnap, Stegmüller, 1959).

Die Beobachtungen werden in Form von Protokollsätzen  $B_1, B_2, \dots, B_n$  dokumentiert. Solange sich eine theoretische Aussage  $H$  aus den Protokollsätzen logisch ableiten lässt, gilt sie als verifizierbar:

$$(B_1 \wedge B_2 \wedge \dots \wedge B_n) \rightarrow H$$

In diesem Ansatz haben Experimente die primäre Funktion, Ergebnisse im Sinne von Protokollsätzen zu liefern. Die darin beschriebenen Beobachtungen sind die Basis jeder Wissenschaft, aus denen dann allgemeinere Aussagen über Gesetzmäßigkeiten abzuleiten sind, die die beobachteten Tatsachen erklären. (Westermann, 1987 b) .

Aufgrund des oben aufgeführten Induktionsproblems können wissenschaftliche Theorien niemals endgültig verifiziert werden, d.h. man kann nicht aus einer endlichen Zahl von Beobachtungen auf die Allgemeingültigkeit von Theorien schließen. Man kann ihnen jedoch auf Grundlage von empirischen Befunden eine gewisse Wahrscheinlichkeit zuordnen. In Carnaps Konzeption ist die die Zuordnung von bedingten Wahrscheinlichkeiten der zentrale Aspekt der induktiven Methode. (Lauth, Sareiter, 2005)

Sei

- $h$  eine wissenschaftliche Hypothese
- $e$  ein empirischer Datensatz, d.h. die Beschreibung der Ereignisse irgendwelcher Beobachtungen oder Experimente

Betrachtet wird nun die bedingte Wahrscheinlichkeit  $p(h/e)$  als Wahrscheinlichkeit für die Hypothese  $h$  unter der Voraussetzung, dass die empirischen Daten  $e$  vorliegen. Dies kann zu folgenden Ergebnissen führen (Lauth, Sareiter, 2005):

- $p(h/e) = 1$       Es gilt  $e \vdash h$ , d.h. die Hypothese  $h$  ist durch die Daten  $e$  verifiziert (die Symbolik bedeutet "aus  $e$  folgt die logische Beweisbarkeit von  $h$ ").
- $p(h/e) = 0$       Es gilt  $e \vdash \neg h$ , d.h. aus  $e$  folgt die logische Beweisbarkeit von  $\neg h$ ; die Hypothese  $h$  ist durch die Daten  $e$  falsifiziert.
- $p(h/e) > p(h)$       Deduktive Bestätigung der Hypothese  $h$  durch die Daten  $e$ , wenn  $h \vdash e$  gilt, d.h. die Daten  $e$  sind aus der Hypothese  $h$  ableitbar.

Die bedingte Wahrscheinlichkeit  $p(h/e)$  wird in der Regel Werte zwischen 0 und 1 annehmen, d.h. die Hypothese kann nicht definitiv verifiziert oder falsifiziert werden. Aber auch wenn die Induktion nicht zu definitiver Erkenntnis führen kann, so kann sie doch über rational begründbare Korrekturen der subjektiven Wahrscheinlichkeiten zu einer graduellen Annäherung an die Wahrheit führen.

Dies kann mit folgender Vorstellung veranschaulicht werden:

Beobachtungen oder Experimente finden zu unterschiedlichen Zeitpunkten  $t = 0, 1, \dots, n$  statt und liefern als Ergebnisse die Daten  $e_1, e_2, \dots, e_n$ . Sei  $p_0(h)$  die als Startverteilung bezeichnete subjektive Wahrscheinlichkeitsverteilung vor der Durchführung der Beobachtungen oder Experimente. Für eine empirische Bestätigung der Hypothese muss nun gelten  $p_0(h/e_1, e_2) > p_0(h/e_1)$ . Für  $p_n(h) = p_0(h/e_1, e_2, \dots, e_n)$  bedeutet dann

$$\begin{aligned} \lim_{n \rightarrow \infty} p_n(h) &= 1 && \text{falls die Hypothese } h \text{ tatsächlich zutrifft} \\ \lim_{n \rightarrow \infty} p_n(h) &= 0 && \text{falls die Hypothese } h \text{ nicht zutrifft} \end{aligned}$$

Solange die bedingte Wahrscheinlichkeit für die Gültigkeit einer Hypothese durch weitere empirische Befunde erhöht wird, spricht man von einer objektiven Bestätigung der Hypothese durch die Daten. (Lauth, Sareiter, 2005)

### 3.1.5 Strukturalismus

"Die strukturalistische Theoriekonzeption soll in erster Linie keine normative Theorie darüber sein, wie Wissenschaftler vorzugehen haben, sondern eine deskriptive Theorie im Sinne einer logisch-systematischen Rekonstruktion der wissenschaftlichen Theorien, der mit ihnen verbundenen empirischen Behauptungen und ihrer Rolle innerhalb der Forschungs- und Anwendungsprozesse." (Westermann, 1987 a)

Während die Methodologien von Carnap und Popper Theorien als Klassen von semantischen Aussagen auffassen und sich mit der Untersuchung der logischen Ableitungsbeziehungen zwischen ihnen beschäftigen, die Theorien somit die Form deterministischer Allsätze mit universeller Gültigkeit haben, wird in der strukturalistischen Theoriekonzeption zwischen dem formalen Kern und den intendierten Anwendungen einer Theorie unterschieden.

Die Theorie wird dabei als Netzwerk von Theorieelementen dargestellt, zwischen denen verschiedene Arten systematischer Beziehungen identifiziert werden können. Dieses Netzwerk kann sich im Laufe des Forschungsprozesses verändern, was sich in theoretischen sowie empirischen Fortschritten und Rückschlägen zeigen kann.

Ziel der theoriegeleiteten Forschung ist also die Abgrenzung derjenigen Klassen von intendierten Anwendungen, auf die die Theorie erfolgreich angewendet werden kann, von

denjenigen, die noch nicht mit Hilfe der Theorie erklärt werden können. Das hat zur Folge, dass eine Theorie nicht mehr an empirischer Erfahrung scheitern kann, sondern allenfalls von einer anderen Theorie verdrängt wird. Es wird nicht mehr über Falsifikation oder Bewährung einer Theorie entschieden sondern über die Zusammensetzung der Menge der intendierten Anwendungen und dabei besonders über die Teilmenge der bewährten Anwendungen.

Im theorieorientierten Forschungsprozess dient jedes Experiment der Überprüfung einer empirischen Hypothese über die erfolgreiche Anwendbarkeit eines bestimmten Theorieelements auf ein bestimmtes empirisches System. Dabei wird das Risiko von Fehlentscheidungen über die Gültigkeit der empirischen Hypothese mithilfe entsprechender experimenteller Planungs-, Durchführungs- und Auswertungstechniken verringert.

Fehlentscheidungen bezüglich der empirischen Hypothese sind in zwei Richtungen möglich:

- I Die empirische Hypothese trifft nicht zu, wird aber als zutreffend bezeichnet.
- II Die empirische Hypothese trifft zu, wird aber als nicht zutreffend bezeichnet.

Die Wahrscheinlichkeiten für die beiden Arten der Fehlentscheidung werden mit  $e_j$  für I und mit  $f_j$  für II bezeichnet (siehe Tabelle 4).

	EH trifft tatsächlich zu	EH trifft tatsächlich nicht zu
EH wird als zutreffend betrachtet	$1 - f_j$	$e_j$
EH wird als nicht zutreffend betrachtet	$f_j$	$1 - e_j$

Tabelle 4: Wahrscheinlichkeiten für richtige und falsche Entscheidungen über das Zutreffen einer empirischen Hypothese (EH) (nach Westermann, 1987 a)

Diese Wahrscheinlichkeiten sind naturgemäß weder bekannt noch können sie abgeschätzt oder in in Relation zueinander oder zu den Wahrscheinlichkeiten anderer Untersuchungen gesetzt werden. Sie dienen lediglich als konzeptueller Rahmen für die weitere Betrachtung der Validität empirischer Untersuchungen.

Der von Popper geforderten strengen Prüfung von Theorien entspricht hier das Bestreben, die Wahrscheinlichkeit  $e_j$  möglichst gering zu halten. Selbstverständlich sollen auch fälschliche Falsifikationen vermieden werden – Prüfungen müssen also ebenso danach streben, zutreffende Hypothesen zu bestätigen, d.h. auch den Wert  $f_j$  möglichst gering zu halten. Um einer Hypothese die Chance zu geben, sich als zutreffend zu erweisen, sollen die Prüfungen demnach nicht nur streng sondern auch wohlwollend sein. (Westermann, 1987 a).

Diese Prüfungen beziehen sich auf die oben beschriebenen intendierten Anwendungen einer Theorie. Um die erfolgreichen Anwendungsbereiche einer wissenschaftlichen Theorie gegen die nicht zum Geltungsbereich gehörenden Anwendungen abgrenzen zu können, werden empirische Untersuchungen durchgeführt, in der die empirische Hypothese geprüft wird,

dass ein bestimmtes Theorie-Element auf bestimmte Personen in einer bestimmten Situation anwendbar ist. (Westermann, 2000)

Hierfür werden aus der wissenschaftlichen Hypothese konkrete Vorhersagen über empirisch zu erwartende Ergebnisse abgeleitet, um anschließend anhand vorher festgelegter Entscheidungskriterien zu entscheiden, ob die erhaltenen Daten systematisch von den erwarteten abweichen. Eine Methode zur Festlegung solcher Entscheidungskriterien ist der Signifikanztest. Westermann beschreibt folgende Vorteile gegenüber anderen möglichen Verfahren (Westermann, 2000):

1. Signifikanztests überprüfen, inwieweit neben der Fehlervarianz durch Zufallsfaktoren sowie die unvermeidliche Abweichung von den Ceteris-paribus-Bedingungen systematische Unterschiede bestehen.
2. Strenge und Wohlwollen der Prüfung können durch explizite Kontrolle der  $\alpha$ - und  $\beta$ -Fehlerwahrscheinlichkeiten systematisch beeinflusst werden.
3. Der Signifikanztest ist unter den Fachwissenschaftlern eine weithin akzeptierte Strategie zur Datenauswertung und Hypothesenprüfung. Ohne solche Konventionen als notwendiger Bestandteil des fachwissenschaftlichen Paradigmas ist eine normalwissenschaftliche Forschung nicht möglich.

In seinen Ausführungen rekurriert Westermann auf zwei miteinander konkurrierende statistische Testtheorien, ohne diese weiter zu differenzieren: Die Theorie des Signifikanztests von Ronald Aylmer Fisher (FST) und die Theorie des statistischen Hypothesentestens von Jerzy Neyman und Egon Sharpe Pearson (NPT). Auch wenn diese von Hager (Hager, 2004) als Hybridisierung bezeichnete Vermischung von Elementen beider Theorien gängige Forschungspraxis ist, sollte sie nicht unreflektiert übernommen werden. Im folgenden Abschnitt sollen daher diese beiden Testtheorien in ihren wesentlichen Merkmalen beschrieben und einander gegenübergestellt werden.

## 3.2 Testtheorien zum statistischen Testen von Hypothesen

### 3.2.1 Theorie des Signifikanztests von R.A. Fisher (FST)

Unter der FST wird einer Forschungshypothese eine testbare Nullhypothese  $H_0$  gegenübergestellt. Diese kann im Rahmen eines Experiments widerlegt, jedoch niemals bewiesen oder bestätigt werden. Das Zurückweisen der Nullhypothese darf keinesfalls gleichgesetzt werden mit dem Beweis der Forschungshypothese, da diese nicht eindeutig formuliert ist: Während die Nullhypothese sich auf einen einzigen Wert bezieht, umfasst die Forschungshypothese alle anderen Werte. Eine Festlegung der Alternativhypothese auf einen einzigen Wert hätte zur Folge, dass sie durch ein einziges Gegenbeispiel widerlegt werden könnte. (Fisher, 1951)

Der entscheidende Kennwert ist der p-Wert, der aussagt, inwieweit das mittels Teststatistik ermittelte Ergebnis durch andere als die experimentell erzeugten Bedingungen, also zufällig, entstanden sein kann. Ist der p-Wert hinreichend klein - in der Regel 5% - wird die Nullhypothese verworfen und damit die Forschungshypothese gestützt (nicht bewiesen!). (Fisher, 1951)

Hager (Hager, 2004) betont in diesem Zusammenhang, dass sich die Nullhypothese als statische Hypothese ausschließlich auf statistische Konzepte wie Zufallsvariablen und ihre Kennwerte beziehen muss und nicht als verbalisierte Negation der Forschungshypothese formuliert sein kann.

### 3.2.2 Theorie des statistischen Hypothesentestens von J. Neyman und E.S. Pearson (NPT)

Grundlage dieser Theorie ist, dass man niemals wissen kann, ob eine Hypothese wahr oder falsch ist, man jedoch bestimmte "Entscheidungsregeln" aufstellen kann, wie zum Beispiel die, eine Hypothese  $H$  zu verwerfen, wenn für ein bestimmtes Merkmal  $x$  gilt  $x > x_0$ , und die Hypothese  $H$  zu akzeptieren, wenn  $x \leq x_0$  (Neyman, Pearson, 1933).

Ziel ist demnach die Entscheidung für eine von zwei entgegengesetzten Hypothesen als Grundlage für das weitere Handeln, also sich anschließend so zu verhalten, als ob die angenommene statistische Hypothese zuträfe (Hager, 2004). Dabei ist nicht auszuschließen, dass die Entscheidung fälschlicherweise zugunsten einer nichtzutreffenden Hypothese fällt oder eine zutreffende Hypothese zurück gewiesen wird. Für diese beiden möglichen Fehlentscheidungen haben Neyman und Pearson folgende Fehlerwahrscheinlichkeiten definiert:

$$\alpha := \pi(\text{Entscheidung: "Die } H_0 \text{ wird abgelehnt und die } H_1 \text{ angenommen"} \mid \text{die } H_0 \text{ trifft zu})$$

$$\beta := \pi(\text{Entscheidung: "Die } H_0 \text{ wird beibehalten" } \mid \text{die } H_1 \text{ trifft zu})$$

Der  $\alpha$ -Fehler wird als Fehler 1. Art oder Typ-1-Fehler bezeichnet, der  $\beta$ -Fehler als Fehler 2. Art oder Typ-2-Fehler.

Neben dem Signifikanzniveau  $\alpha$  als Entscheidungswert für die Annahme der  $H_0$  ist die Teststärke oder Power  $1-\beta$  ein zentraler Wert, der die Wahrscheinlichkeit angibt, mit der eine zutreffende Hypothese  $H_1$  auch tatsächlich angenommen wird. In Tabelle 5 sind die Fehlerwahrscheinlichkeiten dargestellt.

	$H_0$ trifft zu	$H_0$ trifft nicht zu
$H_0$ wird beibehalten	$1 - \alpha$	$\beta$
$H_0$ wird abgelehnt	$\alpha$	$1 - \beta$

Tabelle 5: Wahrscheinlichkeiten für richtige und falsche Entscheidungen über das Beibehalten einer statistischen Hypothese  $H_0$

Unter der strukturalistischen Theoriekonzeption ist es unverzichtbar, die Wahrscheinlichkeit für beide Fehler möglichst gering zu halten, denn  $\alpha$  entspricht der Strenge des Prüfversuchs, während  $\beta$  das Wohlbefinden beeinflusst. (Hager, 2004)

Selbstverständlich sind beide Parameter ebenso wie die Effektgröße a priori festzusetzen, so dass sie über den Stichprobenumfang kontrolliert werden können.



### 3.2.3 Hybridisierung von FST und NPT

In Tabelle 6 wird die Vorgehensweise der beiden Testtheorien in der praktischen Anwendung vergleichend gegenübergestellt.

FST	NPT
1) Spezifiziere die $H_0$ , d.h. die zurückzuweisende Negation einer Forschungshypothese.	1) Spezifiziere die $H_0$ als Hypothese einer Nulldifferenz oder eines fehlenden Zusammenhangs sowie die ihr komplementäre $H_1$ .
2) Spezifiziere eine geeignete Teststatistik ( $TS$ ) für die Testung der $H_0$ und die zugehörige Testverteilung.	2) Spezifiziere eine geeignete Teststatistik ( $TS$ ) für die Testung der $H_0$ und die zugehörige Testverteilung.
3) A-priori-Spezifizierung des Signifikanzniveaus $\alpha$ möglich, jedoch nicht zwingend erforderlich	3) Spezifiziere a priori das Signifikanzniveau $\alpha$ , die Teststärke $1 - \beta$ und eine nachzuweisende Effektgröße.
	4) Bestimme a priori den benötigten Stichprobenumfang relativ zu $\alpha$ , zu $\beta$ und zum Wert der Effektgröße.
4) Erhebe die Daten und bestimme den empirischen Wert von $TS$ , $TS_{emp}$	5) Erhebe die Daten und bestimme den empirischen Wert von $TS$ , $TS_{emp}$
5) Berechne den p-Wert zu $TS_{emp}$ .	6) Bestimme den kritischen Wert $TS$ , $TS_{krit}$ , und den Ablehnungsbereich.
6) Beurteile die Größe des empirischen p-Wertes. Ist er hinreichend klein ( $p \leq 0,05$ ; optional $p \leq \alpha$ ), weise $H_0$ zurück und bezeichne die Forschungshypothese als zutreffend. Verallgemeinere ggf. auf die Population/en. Ist der p-Wert nicht hinreichend klein, ziehe keine Schlüsse.	7) Vergleiche $TS_{emp}$ und $TS_{krit}$ : Gilt, dass $ TS_{emp}  \geq  TS_{krit} $ , weise $H_0$ zurück und nimm $H_1$ an; andernfalls behalte $H_0$ bei. Verhalte dich so, als ob die angenommene statistische Hypothese zuträfe.

Tabelle 6: Das Vorgehen unter den beiden konkurrierenden Testtheorien FST und NPT (nach Hager, 2004)

Auch wenn es nicht explizit Bestandteil der Testtheorien ist, sollte bei beiden Ansätzen selbstredend der Wert der tatsächlichen Effektgröße ermittelt und unter der NPT mit dem Kriteriumswert  $TS_{krit}$  verglichen werden.

Unter der FST ist der p-Wert die einzige zulässige Statistik; dieser beantwortet lediglich die Frage, ob ein überzufälliger Effekt aufgetreten ist oder nicht und somit die Nullhypothese  $H_0$  beibehalten oder die Alternativhypothese  $H_1$  angenommen wird. Aus der Nicht-Ablehnung der Nullhypothese kann jedoch nicht auf ihr Zutreffen geschlossen werden; eine Annahme der Nullhypothese ist in diesem Ansatz nicht möglich. Somit spielt auch die Teststärke bei der Planung keine Rolle. Ebenso wenig wird die erwartete Effektgröße berücksichtigt – es stellt sich erst im Nachhinein durch einen zu großen p-Wert heraus, wenn die Stichprobe zu klein war um nachzuweisen, dass der eingetretene Effekt nicht nur zufällig entstanden ist. Die Größe des Effekts ist dabei irrelevant.

Unter der NPT kann die Nullhypothese  $H_0$  nicht nur verworfen sondern auch angenommen werden, und zwar ausschließlich unter Kontrolle der Teststärke  $1 - \beta$ . Die beiden Risiken  $\alpha$

und  $\beta$  werden im Rahmen der Versuchsplanung festgelegt und bestimmen zusammen mit der erwarteten Effektgröße den erforderlichen Stichprobenumfang.

Da beide Theorien mit den gleichen zentralen Konzepten und Tests operieren, ist eine Hybridisierung der beiden Testtheorien in der Form möglich, dass im Rahmen der Umsetzung eines Tests unter der NPT zusätzlich der p-Wert bestimmt wird, um ihn mit dem vorher festgelegten Signifikanzniveau  $\alpha$  zu vergleichen.

Eine Vermischung der FST mit einzelnen Elementen der NPT, z.B. die Bestimmung eines zu erwartenden Effekts und/oder der Teststärke ohne diese mit der Berechnung des erforderlichen Stichprobenumfangs zu verbinden, ist weder sinnvoll noch zulässig. (Hager, 2004)

### 3.3 Fallzahlschätzung und Teststärkenanalyse

Im Folgenden wird davon ausgegangen, dass in der Pflegeforschung einseitig gerichtete Effekte untersucht werden sollen, d.h. die durch die zu untersuchende Intervention erreichte Veränderung soll in eine bestimmte Richtung gehen wie z.B. Erhöhung der Mobilität oder des BMI, Verringerung eines Risikos, Schmerzreduzierung etc. Die dargestellten Formeln beziehen sich daher auf einseitige Tests.

Die im Beispiel in Abschnitt 2.2.1 verwendete Formel lässt sich in allgemeiner Form wie folgt darstellen (Bredenkamp, 1969):

$$n = \frac{(z_\alpha + z_\beta)^2 2\sigma^2}{\delta^2}$$

wobei  $\delta$  für die erwartete Mittelwertsdifferenz steht.

Diese Formel setzt voraus, dass die Varianz  $\sigma^2$  vor dem Experiment bekannt ist. Dies ist jedoch nicht immer der Fall. Hierfür leitet Bredenkamp eine Formel ab, in der die praktisch bedeutsame Mittelwertsdifferenz als Vielfaches der Standardabweichung der Fehler festgelegt wird:

$$n = \frac{(z_\alpha + z_\beta)^2 (1 - \omega^2)}{2\omega^2}$$

Dabei entspricht  $\omega^2$  demjenigen Anteil an der Gesamtvarianz, der durch die Variation der experimentellen Bedingungen erklärt werden soll. (Bredenkamp, 1969)

Signifikanzniveau und Teststärke sind für eine valide Testplanung im Vorfeld festzulegen. Diese sollten auf inhaltlichen Kosten-Nutzen-Überlegungen basieren – welche Folgen hätte eine falsche Entscheidung? Das Risiko einer Fehlentscheidung zugunsten einer Maßnahme, die nicht wirkt (Fehler 1. Art), soll möglichst gering gehalten werden, wenn die Maßnahme mit erheblichen Kosten verbunden ist, z.B. Nebenwirkungen oder Zeitaufwand; ist die Fehlentscheidung mit keinen Einschränkungen verbunden, kann ein höheres Risiko toleriert werden. Umgekehrt wird man auch das Risiko einer Fehlentscheidung gegen eine Maßnahme, die jedoch gewirkt hätte (Fehler 2. Art), z.B. prophylaktische Maßnahmen, umso höher einschätzen, je schwerwiegender die Folgen des Unterlassens wären.

Der in Abschnitt 3.1.5 beschriebene Vorteil des Hypothesentestens als konventionell anerkannte Methode unter Forschern beinhaltet die Diskussion und den Vergleich von Forschungsergebnissen in Metaanalysen.

Somit ergeben sich zwei Möglichkeiten für die Festsetzung der beiden Fehlerwahrscheinlichkeiten: Eine Möglichkeit ist die individuelle Festsetzung der Fehlerwahrscheinlichkeiten als nach Ansicht des Forschenden tolerierbare Risiken oder im Rahmen einer Teststärkeanalyse. Die zweite Möglichkeit ist das Einsetzen der konventionell verwendeten Werte, wobei in der Regel  $\alpha = 0,05$ , in Ausnahmefällen auch  $\alpha = 0,1$  oder  $\alpha = 0,01$  und  $\beta = 0,2$  gesetzt werden (z.B. in Donner, Klar, 2000).

Die für die Teststärkeanalyse zugrunde zu legenden Effektstärken werden bei Erdfelder (Erdfelder et al., 2004) mit  $\varepsilon$ , bei Cohen mit  $d$  bezeichnet und für den Vergleich von Mittelwerten  $m$  wie folgt bestimmt (Cohen, 1988):

Für den Fall  $\sigma_A = \sigma_B$ ,  $n_A = n_B$  wird

$$d = \frac{m_B - m_A}{\sigma}$$

gesetzt (wenn die Alternativhypothese lautet  $m_B > m_A$ , sonst umgekehrt).

Für den Fall  $\sigma_A \neq \sigma_B$ ,  $n_A = n_B$  wird  $\sigma$  ersetzt durch

$$\sigma' = \sqrt{\frac{\sigma_A^2 + \sigma_B^2}{2}}$$

Cohen legt für die Effektgrößen folgende Konventionen fest (Cohen, 1988):

- $d = 0,2$  als kleine Effektgröße
- $d = 0,5$  als mittlere Effektgröße
- $d = 0,8$  als große Effektgröße.

Die hier für den Mittelwertsvergleich vorgestellten Berechnungen und Konventionen finden sich bei Cohen (1988) für viele weitere Teststatistiken. An diesen Konventionen können sich Forschende für die Berechnung der Stichprobengröße orientieren, wenn Mittelwerte und/oder Standardabweichung nicht vorliegen.

Cohen (1988) stellt vier Formen der Teststärkeanalyse vor; die Bezeichnungen sind Erdfelder et al. (2004) entnommen, die als weitere Analysemöglichkeit die Kompromiss-Analyse aufführen.

1. Mit der *A-priori-Analyse* wird bei bekanntem  $\alpha$ ,  $\beta$  und erwarteter Effektgröße  $\varepsilon$  der Stichprobenumfang  $N$  ermittelt.
2. Die *Post-hoc-Analyse* liefert die Teststärke  $1 - \beta$  oder den  $\beta$ -Fehler zu einem gegebenen Stichprobenumfang  $N$ , dem gewählten  $\alpha$  und der bestimmten Effektgröße  $\varepsilon$ . Mit dieser Analyse kann vor der Untersuchung ermittelt werden, ob eine zur Verfügung stehende Stichprobe ausreicht, den gewünschten Effekt zu erzielen. Sie kann aber

auch genutzt werden, um nach einer Untersuchung die Studienpower zu ermitteln. Diese retrospektive Vorgehensweise ist umstritten, da die Effektstärke aus denselben Stichprobendaten geschätzt wird, für die auch die Teststärke ermittelt werden soll (Erdfelder et al., 2004), und die Stichprobenwerte besonders bei kleinen Stichproben stark von den Populationsparametern abweichen können (Erdfelder et al., 2005).

3. Wenn die verfügbare Stichprobe nicht das erforderliche Ausmaß erreicht, kann die *Kompromiss-Analyse* eingesetzt werden, um zu ermitteln, welches Signifikanzniveau  $\alpha$  und welche Power  $1 - \beta$  erreicht werden können. Diese werden aus dem vom Forscher zu bestimmenden Verhältnis  $\beta/\alpha$ , der Stichprobengröße  $N$  und der erwarteten Effektgröße  $\varepsilon$  ermittelt.
4. Die *Sensitivitäts-Analyse* zeigt, welcher Effekt mit einem begrenzten Stichprobenumfang  $N$  bei den festgelegten Fehlerwahrscheinlichkeiten  $\alpha$  und  $\beta$  noch nachgewiesen werden kann. Ist der erwartete Effekt  $\varepsilon$  kleiner, müssen entweder der Stichprobenumfang erhöht oder  $\alpha$  und  $\beta$  im Rahmen einer Kompromiss-Analyse angepasst werden.
5. Die *Kriteriumsanalyse* liefert das Signifikanz-Niveau  $\alpha$  bei gegebenem Stichprobenumfang  $N$  sowie gegebener Effektstärke  $\varepsilon$  und Teststärke  $1 - \beta$ . Dies kann sinnvoll sein, wenn beispielsweise nachgewiesen werden soll, dass eine neue Intervention zu günstigeren Konditionen den gleichen Effekt erzielt wie die herkömmliche Maßnahme. In dem Fall besagt die Nullhypothese  $H_0$ , dass die beiden Gruppen nicht äquivalent sind.

Diese Teststärkeanalysen können zum Beispiel mit den Power-Tabellen von Cohen (Cohen, 1988) oder einem open-source-Programm wie G\*Power (Faul et al., 2007) durchgeführt werden.

## 3.4 Cluster-Randomisierung und ICC

### 3.4.1 Clusterrandomisierung

Für die Prüfung von Interventionen wird die randomisierte kontrollierte Studie (RCT) als besonders geeignete Methode betrachtet, da die Randomisierung eine gleichmäßige Verteilung der bekannten und unbekanntenen Störgrößen auf die Interventionsgruppe und die Kontrollgruppe garantiert und somit kausale Aussagen über Wirkungszusammenhänge ermöglicht (Kuß et al., 2009).

Dies ist naturgemäß am ehesten bei individual-randomisierten Studiendesigns gewährleistet. Aus verschiedenen Gründen kann es jedoch notwendig sein, zwar die Zielgröße an einzelnen Probanden zu erheben, diese jedoch nicht individuell sondern in Gruppen, sogenannten Clustern, auf die Interventions- und Kontrollgruppe zu verteilen. Solche Gründe können sein (Kuß et al., 2009):

1. Bei der Intervention handelt es sich um eine üblicherweise auf Gruppen angewendete Maßnahme, beispielsweise eine Schulung oder eine Änderung der Organisationsform.

2. Eine Kontamination soll verhindert werden. Würde man beispielsweise zufällig ausgewählte Patienten aus mehreren Krankenhäusern in einer bestimmten Technik schulen, andere jedoch nicht, bestünde die Gefahr, dass die Patienten sich austauschen und auch diejenigen aus der Kontrollgruppe die Technik angewendet haben, von der sie nicht durch das Programm sondern durch andere Patienten erfahren haben. Dies lässt sich vermeiden, indem die Randomisierung über die Krankenhäuser statt über die einzelnen Patienten erfolgt.
3. Es ist effektiver, eine Maßnahme in Gruppen zu testen. Es wäre sehr aufwändig, das Pflegepersonal vieler Pflegeeinrichtungen in einer bestimmten Technik zu schulen und die Auswirkungen dann nur an jeweils einem Bewohner zu testen.

Mit der Randomisierung von Clustern auf die Interventions- und Kontrollgruppe ergibt sich das Problem, dass von einer Unabhängigkeit der Ergebnisse von Probanden im selben Cluster nicht mehr ausgegangen werden kann, d.h. die Ergebnisse innerhalb eines Clusters variieren weniger stark als die der gesamten Stichprobe. Wenn die Untersuchungseinheit beispielsweise ein Krankenhaus ist, kann das Ergebnis zusätzlich durch die Personalsituation, die bauliche Struktur, das Leitbild, das Patientenkontingent usw. beeinflusst worden sein. Diese Korrelation zwischen den Probanden innerhalb eines Clusters wird als Intra-Cluster-Correlation (ICC)  $\rho$  bezeichnet und ist nicht nur bei der Auswertung der Daten sondern bereits bei der Planung der Studie zu berücksichtigen.

### 3.4.2 ICC

Wie bereits im Beispiel in Abschnitt 2.2.1 gezeigt wurde, hat die ICC einen nicht vernachlässigbaren Einfluss auf die Fallzahlschätzung. Hierbei ist zu berücksichtigen, dass der Design-Effekt wesentlich geringer ausfällt, wenn die erforderliche Stichprobe auf viele kleine Cluster verteilt werden kann. In Tabelle 7 wird an zwei Beispielen dargestellt, wie sich der Design-Effekt mit abnehmender Clustergröße (und zunehmender Clusterzahl) verändert:

Beispiel 1: Es wurde eine Fallzahl von  $n = 30$  Probanden pro Studienarm ermittelt.

Beispiel 2: Es wurde eine Fallzahl von  $n = 100$  Probanden pro Studienarm ermittelt.

Stichprobengröße $n$	30				100			
Clustergröße $m$	15	10	5	3	50	25	20	10
Anzahl Cluster $k$	2	3	6	10	2	4	5	10
ICC $\rho = 0,01$	1,14	1,09	1,04	1,02	1,49	1,24	1,19	1,09

Tabelle 7: Design-Effekt  $DE$  für  $\rho = 0,01$  für unterschiedliche Clustergrößen

Die Entscheidung zur Clusterrandomisierung, in der Pflegeforschung meist über Stationen oder Einrichtungen, ist in der Regel bedingt durch die Art der Intervention; der Möglichkeit, die Clustergröße zugunsten der Anzahl der Cluster zu verringern, sind somit enge Grenzen gesetzt. Die Berücksichtigung des Design-Effekts bei der Berechnung der Stichprobengröße gewinnt damit umso mehr an Bedeutung.

Dass die Leitlinie für Berichte randomisierter Studien "CONSORT 2010 statement" (Schulz et al., 2010) um eine Erweiterung für clusterrandomisierte Studien (Campbell et al., 2012) ergänzt wurde, zeigt die Bedeutung, die den methodischen Unterschieden zu individual-randomisierten Studien beigemessen wird. Hier wird gefordert, neben der Berechnung der Stichprobengröße auch die Anzahl der Cluster, ihre Größe und die berücksichtigte ICC in die Veröffentlichung aufzunehmen.

Zwar schreibt Kuß (Kuß et al., 2009), die in pflegewissenschaftlichen Studien zur erwartenden ICC seien "glücklicherweise" klein und bezieht sich dabei auf eine Studie aus der Allgemeinmedizin, die in einer Analyse von 1039 ICC einen medianen ICC von 0,01 ergeben hatte (Adams et al., 2004). Abgesehen davon, dass bezweifelt werden darf, inwieweit diese Ergebnisse auf pflegerisch relevante Endpunkte übertragbar sind, kann wie in Tabelle 3 dargestellt bereits eine ICC von 0,01 erheblichen Einfluss auf die erforderliche Stichprobengröße haben.

## 4 Datenanalyse

Die geplante Sekundäranalyse von Daten der in der Literaturrecherche ermittelten Studien (siehe Anhang A) war nicht möglich, da diese aus ethischen Gründen nicht zur Verfügung gestellt wurden. Die Probanden hatten einer Sekundär-Analyse nicht zugestimmt oder eine Weitergabe der Daten an andere Institutionen war explizit ausgeschlossen worden.

Für eine Datenanalyse stand der Datensatz der im Rahmen der am Lehrstuhl für Statistik und standardisierte Methoden für Pflegeforschung der Philosophisch-Theologischen Hochschule Vallendar im Auftrag der Saarländischen Pflegegesellschaft e.V. durchgeführten PiSaar-Studie zur Verfügung.

Ziel der PiSaar-Studie war es, über eine Erhebung der Gesamtpflege- und -betreuungszeit angemessene statistische Methoden zu identifizieren, die die Haupteffekte derjenigen Variablen entdecken, die den Pflegezeitaufwand erklären (Brühl, Planer 2013). Hierfür wurden neben dem Zeitaufwand über 400 Variablen aus den Bereichen Organisation, personenbezogene Merkmale der Pflegenden sowie Bewohnermerkmale erfasst.

Für die hier durchgeführte Untersuchung wurden lediglich die bewohnerbezogenen Daten benötigt und zur Verfügung gestellt.

### 4.1 Datenerhebung

#### 4.1.1 Stichprobe

An der Untersuchung beteiligten sich 62 Einrichtungen aus der Grundgesamtheit aller Altenheime des Saarlandes. Die Einrichtungen haben sich freiwillig für die Teilnahme an der Untersuchung gemeldet, so dass es sich um eine Gelegenheitsstichprobe handelt.

Für die Berechnung der ICC konnten die Daten von 2028 Bewohner/innen (Bew.) von 68 Wohnbereichen (WB) heran gezogen werden. Die Wohnbereiche haben eine Größe von 12

bis 56 Bewohner/innen, die sich wie in Tabelle 8 dargestellt verteilen.

Anzahl Bew./WB	<20	20-29	30-39	40-49	> 50
Anzahl WB	3	34	21	7	3

Tabelle 8: Verteilung der Größe der Wohnbereiche

#### 4.1.2 Variablenauswahl

Von jedem Bewohner/jeder Bewohnerin wurden 119 bewohnerbezogene Merkmale erhoben. Dem Ziel der PiSaar-Studie entsprechend handelt es sich um solche Merkmale, die den Zeitaufwand der Pflegenden beeinflussen (Brühl, Planer 2013).

Die Variablen wurden in Anlehnung an das ABEDL-Strukturmodell von Monika Krohwinkel (2013) strukturiert, ergänzt um allgemeine Merkmale sowie Risiken, medizinische Diagnosen und behandlungspflegerische Tätigkeiten. (In der PiSaar-Studie wird der Begriff "AEDL" verwendet und so mit den zugehörigen Daten für diese Arbeit übernommen.)

Abgesehen von den allgemeinen Variablen sind die Merkmale dichotom ausgeprägt.

Die Erhebung der hier verwendeten Daten war als Stichtagserhebung konzipiert und fand am 31. März 2012 statt und erfolgte je nach Größe der Einrichtung durch Pflegedienstleitungen, Wohnbereichsleitungen oder Pflegefachkräfte.

Für die Berechnung der ICC wurden 53 Variablen ausgewählt, die im Zusammenhang mit den in 2.2.2 identifizierten Forschungsthemen als Untersuchungsgegenstand oder Endpunkt in Frage kommen könnten.

Im Folgenden werden die ausgewählten Variablen mit ihrer im Erhebungsbogen hinterlegten Bedeutung dargestellt (Brühl, Planer 2013), wobei die behandlungspflegerischen Maßnahmen bei den zugehörigen AEDL eingeordnet sind.

Die Variable wird mit "1" codiert, wenn die angegebene Frage mit "ja" zu beantworten ist.

<b>Medizinische Diagnosen</b>
2.3 Demenz
Leidet der Bewohner unter folgender Erkrankung des Nervensystems: Demenz?
1 – wenn der Bewohner unter einer ärztlich diagnostizierten Demenz leidet. Hierbei ist nicht relevant, ob es sich um eine vaskuläre Demenz, eine Alzheimer-Demenz oder eine andere Demenzform handelt. Es ist ebenfalls nicht relevant, welcher Arzt die Diagnose gestellt hat (Facharzt, Hausarzt).
2.11 Dekubitus
Leidet der Bewohner an folgender Erkrankung der Haut: Dekubitus?
1 – wenn der Bewohner an einem Dekubitus Grad 2-4 leidet. Dekubitus Grad 1 ist hier nicht relevant, da die Diagnose häufig nicht eindeutig zu stellen ist. Sollte differentialdiagnostisch auch eine andere Hauterkrankung möglich sein und die Diagnosestellung diesbezüglich nicht eindeutig sein, so wird diese nicht hier sondern in der Frage "andere Erkrankung der Haut" angegeben.
2.15 Kontraktur/Spastik
Leidet der Bewohner an folgender Erkrankung des Bewegungsapparates: Kontraktur/Spastik?
1 – wenn Sie nach Ihrem pflegerischen Verständnis der Meinung sind, dass der Bewohner eine Kontraktur und/oder eine Spastik hat. Diese Diagnose wird durch Pflegekräfte gestellt. In einem systematischen Literaturreview konnten Gnass et al. (Gnass 2010) keine verbindliche Definition von "Kontraktur" formulieren. Eine mögliche Definition: "Als Kontraktur wird eine anhaltende Gelenksteifigkeit bzw. ein mehr oder weniger stark ausgeprägter Verlust der physiologischen Mobilität eines Gelenkes bezeichnet" aus Thieme's Pflege (Juchli 2000, S. 304).
2.23 MRSA/ESBL
Leidet der Bewohner an einer der folgenden Infektionskrankheiten: MRSA, ESBL
1 – wenn eine ärztliche Diagnosestellung zum Zeitpunkt der Datenerhebung vorliegt. Es werden auch diejenigen Bewohner aufgenommen, bei denen derzeit keine Sanierungsversuche unternommen werden.
<b>AEDL Kommunizieren können</b>
3.3 inhaltliches Verstehen
Lässt die Reaktion des Bewohners auf Äußerungen, Hin- und Anweisungen des Pflegepersonals den Rückschluss zu, dass er diese verstanden hat?



1 – wenn die Pflegekraft aus der Erinnerung heraus weiß, dass der Bewohner hierzu in der Lage ist oder die Pflegekraft dies testet, indem sie ihm etwas sagt und merkt, dass er dies verstanden hat. Diese Frage bezieht sich nur auf "verstehen" – hier ist nicht die Umsetzung in Handlung gemeint sondern die Aufmerksamkeit des Bewohners. Verstehen kann z.B. auch ein Tetraplegiker, der dies durch Lidschluss mitteilt.
3.4 gutes Kurzzeitgedächtnis
Kann sich der Bewohner nachmittags daran erinnern, was es zum Mittagessen gab?
1 – wenn die Pflegekraft aus der Erinnerung heraus weiß, dass der Bewohner dazu in der Lage ist, oder sie dies kurz testet, indem sie ihm diese Frage stellt und merkt, dass er sich richtig erinnert. Sollte der Bewohner kein Mittagessen zu sich genommen haben, so wird das Kurzzeitgedächtnis (innerhalb der letzten 12 Stunden) durch eine andere Frage getestet.
3.5 gutes Langzeitgedächtnis
Kann sich der Bewohner an Details früherer Lebensphasen erinnern?
1 – wenn die Pflegekraft aus der Erinnerung heraus weiß, dass der Bewohner hierzu in der Lage ist, oder sie dies testet, indem sie ihm eine Frage zu einer Begebenheit stellt, die längere Zeit zurück liegt. Beispiel: "Wo sind Sie zur Schule gegangen?"
3.6 zeitlich orientiert
Ist der Bewohner zur Zeit orientiert?
1 – wenn der Bewohner "... Tages- und Jahreszeiten und die chronologische Abfolge des eigenen Lebens benennen kann" (Bensch 2009). Dies kann aus der Erinnerung heraus beantwortet werden oder die Pflegekraft testet den Bewohner mit einer entsprechenden Frage z.B. nach Wochentag oder Uhrzeit.
3.7 örtlich orientiert
Ist der Bewohner zum Ort orientiert?
1 – wenn der Bewohner "... weiß, in welcher Stadt, in welchem Gebäude bzw. in welchem Zimmer ..." [er] "... sich befindet" (Bensch 2009). Dies kann aus der Erinnerung heraus beantwortet werden oder die Pflegekraft testet den Bewohner mit einer entsprechenden Frage z.B. nach dem Aufenthaltsort.
3.8 zur Person orientiert
Ist der Bewohner zur Person orientiert?
1 – wenn Personen des näheren Umfelds von dem Bewohner erkannt werden. "... dies können Familienangehörige, Nachbarn und Pflegekräfte etc. sein. In der Bewertung zählt nicht, ob jemand blind ist. Hier wird gewertet, inwieweit er jemanden an der Stimme erkennt." (Bensch 2009)
3.9 verständliche Äußerung möglich
Kann sich der Bewohner verständlich äußern, auch nonverbal?

1 – wenn der Bewohner sich durch Sprache oder Gesten äußern kann. Diese Frage zielt auf die kognitive Fähigkeit zur Sprache (wenn nicht vorhanden: Aphasie) und/oder der Mitteilungsfähigkeit ab; somit ist nicht relevant, wenn ein Bewohner eine fremde Sprache spricht und diese vom Pflegepersonal nicht verstanden wird. In diesem Fall ist ebenfalls "1" anzugeben.
3.10 verbale Kommunikation möglich
Ist verbale Kommunikation mit dem Bewohner möglich?
1 – wenn der Bewohner sich verbal äußern kann und verbale Äußerungen des Pflegepersonals versteht. Hier wird das "sprechen und verstehen können" in den Vordergrund gestellt. Sollte er sich nur äußern können, versteht sein Gegenüber jedoch nicht, so ist nicht mit "1" zu antworten. Bei Anderssprachigkeit ist die "Möglichkeit" zu betrachten, wenn das Pflegepersonal die gleiche Sprache spräche.
3.11 Nutzung Rufanlage
Macht der Bewohner von der Rufanlage Gebrauch?
1 – wenn der Bewohner im Zeitraum der Datenerhebung von der Rufanlage Gebrauch macht. Diese Variable ist ein Ausdruck dessen, dass ein Bewohner kognitiv in der Lage ist, ein Bedürfnis in einen Handlungsimpuls zu übersetzen und dass er motorisch in der Lage ist, den Knopf zu finden und zu drücken. Darüber hinaus gibt sie Aufschluss über Einschränkungen, die ihn veranlassen fremde Hilfe zu beanspruchen. Die Beantwortung ist unabhängig davon, wie oft jemand "klingelt" und wie intensiv die einzelnen Kontakte sind.
3.12 Handlungsanweisung umsetzen
Kann der Bewohner eine Handlungsanweisung des Pflegepersonals umsetzen?
1 – wenn der Bewohner eine Handlungsanweisung (z.B. "Bitte geben Sie mir die Hand") versteht und der Aufforderung nachkommen kann. Hierbei geht es um die kognitive Fähigkeit "verstehen" und Verknüpfung zu einem Handlungsimpuls (sog. Exekutivfunktion). Darüber hinaus geht es nicht um motorische Fähigkeiten. Eine Aufforderung könnte auch heißen: "Bitte bleiben Sie liegen". Im Vordergrund steht die Umsetzung einer Anweisung. Wenn der Bewohner dies nicht möchte, grundsätzlich jedoch dazu in der Lage wäre, ist die Antwort "1".
3.13 Entscheidungen treffen
Ist der Bewohner in der Lage, alltägliche Entscheidungen zu treffen?
1 – wenn sich im alltäglichen Handeln zeigt, dass der Bewohner Entscheidungen selbst treffen kann. Beispiel: ... was er anzieht; womit er sich beschäftigt; was er isst... Die Fähigkeit eine Entscheidung zu treffen geht über das Ausführen von Handlungsanweisungen hinaus und beinhaltet eine individuelle Willensäußerung und die Fähigkeit ein Urteil zu fällen (zwischen Alternativen abwägen zu können). Die Sinnhaftigkeit einer Entscheidung spielt in dieser Frage keine Rolle.

<b>AEDL Sich bewegen können</b>
4.1 Transfer aus Bett selbstständig
Ist der Bewohner in der Lage, den Transfer aus dem Bett selbstständig durchzuführen?
1 – wenn der Bewohner sich komplett selbstständig von einer liegenden Position in einen Rollstuhl oder Stuhl alleine transferieren kann und umgekehrt sich auch wieder hinlegen kann. Hierbei ist entscheidend, dass dies ohne personelle Hilfe möglich ist.
4.2 ohne Hilfe von Stuhl aufstehen
Ist der Bewohner in der Lage, ohne personelle Hilfe von einem Stuhl/Rollstuhl aufzustehen?
1 – wenn der Bewohner alleine ohne Hilfe einer anderen Person in eine stehende Position gelangen kann. Entscheidend ist, ob die Hilfe einer Pflegenden notwendig ist. Wenn der Bewohner sich alleine aufstellen kann und nutzt dabei Hilfsmittel, so wird diese Frage dennoch mit "1" beantwortet.
4.3 ohne Hilfe stehen bleiben
Ist der Bewohner in der Lage, ohne personelle Hilfe zu stehen?
1 – wenn der Bewohner in der Lage ist, ohne Hilfe einer Person stehen zu bleiben. Sollte er sich hierzu festhalten, so ist diese Frage dennoch mit "1" zu beantworten. Auch beim Aufsehen darf er personelle Hilfe benötigen und die Frage kann dennoch mit "1" beantwortet werden, wenn er alleine stehen kann.
4.4 sich ohne Hilfe hinsetzen
Ist der Bewohner in der Lage, sich ohne personelle Hilfe hinzusetzen?
1 – wenn der Bewohner sich alleine an die Bettkante oder auf den Stuhl hinsetzen kann. Hierbei ist entscheidend, dass dies ohne personelle Hilfe möglich ist.
4.5 ohne Hilfe sitzen bleiben
Ist der Bewohner in der Lage, ohne personelle Hilfe sitzen zu bleiben?
1 – wenn der Bewohner in der Lage ist, ohne Hilfe einer Person sitzen zu bleiben. Beim "sitzen bleiben" ist entscheidend, wie stabil die Rumpfmuskulatur ist. Sollte ein Kissen im Rücken genügen, so ist diese Frage dennoch mit "1" zu beantworten. Auch beim Aufsetzen darf er personelle Hilfe benötigen und die Frage kann dennoch mit "1" beantwortet werden, wenn er alleine sitzen kann.
4.6 ohne Hilfe gehen
Ist der Bewohner in der Lage, ohne personelle Hilfe zu gehen?
1 – wenn sich der Bewohner gehend im Zimmer oder im Wohnbereich fortbewegen kann und dabei keine personelle Hilfe benötigt. Es ist unerheblich, welche Entfernungen er zurücklegen kann oder ob er einen Rollator oder Rollstuhl selbstständig nutzt.
4.7 umfängliche Lageveränderung im Bett

Kann der Bewohner seine Lage im Bett selbst umfänglich verändern?
1 – wenn der Bewohner sich alleine vom Rücken auf die Seite oder umgekehrt drehen kann. Diese Frage soll nur bei überwiegend bettlägerigen Bewohnern ausgefüllt werden.
4.8 geringfügige Lageveränderung im Bett
Ist der Bewohner in der Lage, im Bett geringfügige Lageveränderungen vorzunehmen?
1 – wenn der Bewohner dazu in der Lage ist, auf dem gleichen Niveau die Lage zu verändern, z.B. "auf dem Rücken hin- und herrutschen". Diese Frage soll nur bei bettlägerigen Bewohnern ausgefüllt werden.
<b>AEDL Essen und trinken können</b>
6.1 mundgerechte Zubereitung selbstständig
Ist der Bewohner in der Lage, Nahrungsmittel mundgerecht zuzubereiten?
1 – wenn der Bewohner selbstständig ohne personelle Hilfe die Nahrung in mundgerechte Stücke zerkleinern kann und dies nach ein- bis zweimaliger Aufforderung auch tut. Diese Bewohner benötigen keine Hilfe beim Essen.
6.2 essen selbstständig
Ist der Bewohner in der Lage, zubereitete Nahrung selbstständig zu sich zu nehmen?
1 – wenn der Bewohner vorbereitete und zerkleinerte Nahrung selbstständig essen kann und dies nach ein- bis zweimaliger Aufforderung auch tut. Selbstständig bedeutet in diesem Zusammenhang, dass die Pflegekraft den Bewohner nicht anleiten muss sondern währenddessen das Zimmer verlassen kann. Es ist durchaus möglich, dass sowohl die vorherige als auch diese Frage mit "1" beantwortet werden.
6.3 trinken selbstständig
Ist der Bewohner in der Lage, aus einem bereit gestellten Gefäß selbstständig zu trinken?
1 – wenn der Bewohner motorisch dazu in der Lage ist und dies nach ein- bis zweimaliger Aufforderung auch tut. Selbstständig bedeutet in diesem Zusammenhang, dass die Pflegekraft nicht neben dem Bewohner stehen bleiben muss, sondern währenddessen das Zimmer verlassen kann.
6.4 hohe Gewichtsabnahme
Hat der Bewohner ungewollt sehr viel Gewicht verloren?
1 – wenn der Bewohner nach Maßgabe des "Expertenstandard zur Sicherstellung und Förderung der oralen Ernährung in der Pflege" (Schiemann 2009) als untergewichtig eingestuft wird. Dies bezieht sich in dieser Frage maßgeblich auf eine Gewichtsabnahme im Zeitverlauf: > 5% in einem Monat, > 10% in 6 Monaten. Eine starke Gewichtsabnahme infolge Diuretika-Therapie ist nicht mit "1" zu bewerten.
6.6 Dehydratation

Ist der Bewohner dehydriert?
1 – wenn der Bewohner nach Maßgabe des "Expertenstandard zur Sicherstellung und Förderung der oralen Ernährung in der Pflege" (Schiemann 2009) als dehydriert eingestuft wird oder der Bewohner deutliche Exsikkosezeichen zeigt. Beispiele von Exsikkosezeichen: Hautfalte, Oligurie, Obstipation, Verwirrtheit...
12.7 PEG
Hat der Bewohner eine PEG-Sonde (perkutan endoskopische Gastrostomie)?
6.7 Sondenernährung
Wird der Bewohner ganz oder überwiegend über Sonde ernährt?
1 – wenn die überwiegende Nahrung des Bewohners aus Sondenkost besteht. Hier ist die Verabreichung von Sondenkost über eine Sonde gemeint (PEG, Nasensonde).
<b>AEDL Ausscheiden können</b>
7.1 ungewollter Urinverlust
Verliert der Bewohner mehrmals pro Woche Urin?
1 – wenn der Bewohner mehr als einmal pro Woche ungewollt Urin verliert und/oder eine der folgenden Einstufungen des Kontinenzprofils nach dem "Expertenstandard zur Förderung der Harnkontinenz in der Pflege" (Schiemann 2007) vorliegt: - unabhängig kompensierte Inkontinenz - abhängig kompensierte Inkontinenz - nicht kompensierte Inkontinenz Dies kann sich in Einnässen von Inkontinenzmaterial, Bettwäsche oder Kleidung zeigen. Hierbei ist unerheblich, um welche Form von Inkontinenz es sich handelt (z.B. Stressinkontinenz, Dranginkontinenz)
7.2 selbstständige Toilettennutzung
Ist der Bewohner in der Lage, selbstständig die Toilette bzw. den Toilettenstuhl zu benutzen?
1 – wenn der Bewohner den Toilettengang selbstständig und ohne personelle Hilfe durchführt. Selbstständig bedeutet in dieser Frage ohne personelle Hilfe. Nutzt der Bewohner zum Toilettengang andere Hilfsmittel, so wird diese Frage dennoch mit "1" beantwortet. Diese Frage zielt auf die Unterstützung ab, die üblicherweise tagsüber nötig ist. Sollte der Bewohner ausnahmsweise einmal in der Nacht Unterstützung benötigen, so kann er in dieser Frage dennoch mit "1" – also "selbstständig" bewertet werden.
12.5 Blasendauerkatheter
Hat der Bewohner einen Blasendauerkatheter?
12.6 suprapubischer Katheter
Hat der Bewohner einen suprapubischen Blasenkatheeter?

7.4 ungewollter Stuhlverlust
Verliert der Bewohner mehrmals pro Woche ungewollt Stuhl?
1 – wenn der Bewohner mehr als einmal pro Woche ungewollt Stuhl verliert. Dies kann sich durch Einkoten in Inkontinenzmaterial, Bettwäsche oder Kleidung zeigen.
7.6 Obstipation
Leidet der Bewohner unter Obstipation?
1 – wenn die Pflegekraft weiß, dass der Bewohner im Zeitraum der Datenerhebung zu Obstipation neigt. Dies kann sich darin zeigen, dass der Bewohner dies äußert oder dass Maßnahmen zur Verbesserung der Ausscheidungsfrequenz durchgeführt werden. Von pflegerischer Seite wird auf regelmäßige Verdauung geachtet (ballaststoffreiche Kost, viel Flüssigkeit, evtl. Gabe von Laxantien).
<b>AEDL Ruhen, schlafen / existentielle Erfahrungen</b>
8.1 Nachtruhe
Verhält sich der Bewohner in der Nacht ruhig?
1 – wenn der Bewohner im Erhebungszeitraum in der Nacht keine pflegerische Betreuung außerhalb der Routinerundgänge benötigt.
8.2 > 3 Stunden Schlaf am Tag
Schläft der Bewohner tagsüber mehr als 3 Stunden insgesamt?
1 – wenn der Pflegekraft bekannt ist, dass der Bewohner im Erhebungszeitraum tagsüber mehr als drei Stunden geschlafen hat.
8.3 benötigt emotionale Unterstützung
Wirkt der Bewohner als benötige er besondere emotionale Unterstützung?
1 – wenn die Pflegekraft die Situation so einschätzt, dass der Bewohner emotionale Unterstützung benötigt. Diese Einschätzung kann an dieser Stelle nicht standardisiert werden und es obliegt der Pflegefachkraft, die Beantwortung dieser Frage einzuschätzen.
<b>AEDL Sich beschäftigen / soziale Kontakte</b>
9.1 herausforderndes Verhalten
Zeigt der Bewohner herausforderndes Verhalten?
1 – wenn die Pflegekraft sich durch das Verhalten des Bewohners herausgefordert fühlt. Hiermit sind Verhaltensweisen gemeint, „...die die Umgebung herausfordern, die also auch bestimmte Anforderungen an das Verhalten der Pflegenden stellen... Der Begriff herausforderndes Verhalten legt den Fokus auf diejenigen, die sich durch ein bestimmtes Verhalten von demenziell erkrankten Menschen herausgefordert fühlen“ (Bartholomeyczik et al., 2006). Beispiele: Umherwandern, rufen, schreien, aggressives Verhalten gegenüber Dritten...

9.2 ablehnen Pflegemaßnahmen
Lehnt der Bewohner vorgesehene Pflegemaßnahmen zeitweise ab?
1 – wenn der Bewohner im Zeitraum der Datenerhebung Pflegemaßnahmen ablehnt. Die Willensäußerung kann sowohl verbal geäußert werden als sich auch in ablehnendem Verhalten zeigen.
9.3 ablehnen sozialer Betreuung
Lehnt der Bewohner Angebote der sozialen Betreuung ab?
1 – wenn der Bewohner im Zeitraum Angebote der sozialen Betreuung ablehnt, die dem Bewohner seitens der Einrichtung aktiv angeboten werden. Die Willensäußerung kann sowohl verbal geäußert werden als sich auch in ablehnendem Verhalten zeigen.
<b>AEDL Sichere Umgebung</b>
10.2 freiheitsentziehende Maßnahmen
Werden bei dem Bewohner mechanische freiheitsentziehende Maßnahmen eingesetzt?
1 – wenn der Bewohner im Zeitraum der Datenerhebung mechanisch fixiert wurde. Wenn dies im Zeitraum der Datenerhebung vorgekommen ist, so sollte die Frage mit "1" beantwortet werden – unabhängig von der Dauer der Fixierung oder einem richterlichen Beschluss.
10.3 zentral wirksame Medikamente
Erhält der Bewohner Medikamente, die einen Einfluss auf das zentrale Nervensystem haben?
1 – wenn der Bewohner an den Tagen der Datenerhebung Medikamente erhält, die einer der folgenden Gruppen zugeordnet werden können (Steurer 2009): - Antidepressiva - Neuroleptika/Antipsychotika - Benzodiazepine/Sedativa Die Beantwortung der Frage ist unabhängig von der Dosis oder der eingeschätzten Wirkung.
10.4 Hinlauftendenz
Zeigt der Bewohner eine Hinlauftendenz?
1 – wenn die Pflegefachkraft den Bewohner als potentiell besonders gefährdet einstuft aufgrund einer Hinlauftendenz. Dies zeigt sich z.B. in vermehrter Unruhe und gesteigertem Bewegungsdrang oder er ist im Zeitraum der Datenerhebung "weggelaufen".
<b>Risiken</b>
11.1 Dekubitusrisiko
Hat der Bewohner ein Dekubitusrisiko?

<p>1 – wenn der Bewohner von der Pflegefachkraft als dekubitusgefährdet eingestuft wird. Grundlage der Einschätzung ist der "Expertenstandard Dekubitusprophylaxe in der Pflege (1. Aktualisierung 2010)" (Schiemann 2010). Neben eingesetzten Assessmentinstrumenten steht die klinische Einschätzung der Pflegefachkraft im Fokus. Die Entscheidung muss nachvollziehbar sein (aus der Dokumentation muss hervorgehen, dass der Bewohner entsprechende Risikofaktoren aufweist – z.B. vermehrte Druck- bzw. Scherkräfte, schlechte Durchblutungssituation, schlechte Ernährungssituation, bereits vorhandene Hautrötungen. . . .)</p>
<p>11.2 Thromboserisiko</p>
<p>Hat der Bewohner ein Thromboserisiko?</p>
<p>1 – wenn der Bewohner von der Pflegefachkraft als thrombosegefährdet eingestuft wird. Grundlage ist die klinische Einschätzung der Pflegefachkraft. Die Entscheidung muss nachvollziehbar sein (aus der Dokumentation muss hervorgehen, dass der Bewohner entsprechende Risikofaktoren aufweist). Virchow'sche Trias:</p> <ul style="list-style-type: none"> <li>- veränderte Gefäßwände (z.B. durch Entzündungen, Degeneration, Trauma. . .)</li> <li>- erhöhte Gerinnungsneigung (z.B. durch gestörte Gerinnung, Exsikkose, Wunden. . .)</li> <li>- verlangsamter Blutfluss (z.B. durch Immobilität, Herzinsuffizienz, Varizen. . .)</li> </ul>
<p>11.3 Kontrakturrisiko</p>
<p>Hat der Bewohner ein Kontrakturrisiko?</p>
<p>1 – wenn der Bewohner von der Pflegefachkraft als kontrakturgefährdet eingestuft wird. Grundlage der Einschätzung ist die klinische Einschätzung der Pflegefachkraft. Die Entscheidung muss nachvollziehbar sein (aus der Dokumentation muss hervorgehen, dass der Bewohner entsprechende Risikofaktoren aufweist, z.B. Immobilität, fixierte Gelenke, Hemiparese, Spastiken, andere Bewegungseinschränkungen).</p>
<p>11.4 Sturzrisiko</p>
<p>Hat der Bewohner ein Sturzrisiko?</p>
<p>1 – wenn der Bewohner von der Pflegefachkraft als sturzgefährdet eingestuft wird. Grundlage der Einschätzung ist die klinische Einschätzung der Pflegefachkraft. Grundlage der Entscheidung ist der "Expertenstandard Sturzprophylaxe in der Pflege" (Schiemann 2006) und die klinische Einschätzung der Pflegefachkraft. Die Entscheidung muss nachvollziehbar sein (aus der Dokumentation muss hervorgehen, dass der Bewohner entsprechende Risikofaktoren aufweist, z.B. Einschränkungen der Mobilität, Sehbeeinträchtigungen, Sturzereignis, Hypoglykämie, kognitive Einschränkungen, Medikamente. . .).</p>
<p>11.5 Intertrigoisrisiko</p>
<p>Hat der Bewohner ein Intertrigoisrisiko?</p>



1 – wenn der Bewohner von der Pflegefachkraft als intertrigo-gefährdet eingestuft wird. Grundlage für die Einschätzung ist die klinische Einschätzung der Pflegefachkraft. Die Entscheidung muss nachvollziehbar sein (aus der Dokumentation muss hervorgehen, dass der Bewohner entsprechende Risikofaktoren aufweist, z.B. Adipositas, starkes Schwitzen, Inkontinenz. . .).
11.6 Pneumonierisiko
Hat der Bewohner ein Pneumonierisiko?
1 – wenn der Bewohner von der Pflegefachkraft als pneumoniegefährdet eingestuft wird. Die Entscheidung muss nachvollziehbar sein und aus der Dokumentation muss hervorgehen, dass der Bewohner entsprechende Risikofaktoren aufweist: z.B. mangelnde Belüftung der Lunge durch Immobilität, durch andere Lungenerkrankungen oder Herzkrankungen, Aspirationsgefahr. . .
11.7 Risiko Mangelernährung
Hat der Bewohner ein Risiko zur Mangelernährung?
1 – wenn der Bewohner von der Pflegefachkraft als gefährdet im Sinne einer Mangelernährung eingestuft wird. Grundlage der Einschätzung ist der "Expertenstandard zur Sicherstellung und Förderung der oralen Ernährung in der Pflege" (Schiemann 2009) und die klinische Einschätzung der Pflegefachkraft. Die Entscheidung muss nachvollziehbar sein (aus der Dokumentation muss hervorgehen, dass der Bewohner entsprechende Risikofaktoren aufweist, z.B. Mobilitätseinschränkung, Demenz, Depression, Diarrhoe, Erkrankungen des Rachens. . .).
11.8 Risiko Dehydratation
Hat der Bewohner ein Risiko zur Dehydratation?
1 – wenn der Bewohner von der Pflegefachkraft als gefährdet im Sinne einer Dehydratation eingestuft wird. Grundlage der Einschätzung ist der "Expertenstandard zur Sicherstellung und Förderung der oralen Ernährung in der Pflege" (Schiemann 2009) und die klinische Einschätzung der Pflegefachkraft. Die Entscheidung muss nachvollziehbar sein (aus der Dokumentation muss hervorgehen, dass der Bewohner entsprechende Risikofaktoren aufweist, z.B. Mobilitätseinschränkung, Demenz, Depression, Fieber, Diarrhoe. . .).

Die Daten lagen vollständig vor, d.h. es gab keine fehlenden Angaben.

Zu Beginn der Untersuchung wiesen die Variablen 4.7 "umfängliche Lageveränderung im Bett" und 4.8 "geringfügige Lageveränderung im Bett" starke Abweichungen der Werte im Vergleich zu den anderen Variablen auf. Die inhaltliche Analyse der Variablen zeigte, dass diese beiden Items nur dann ausgefüllt werden sollten, wenn der Bewohner überwiegend bettlägerig ist. Das Programm war jedoch so ausgelegt, dass die Pflegekräfte nur bei Zustimmung eine "1" eingeben konnten oder nichts tun, das Programm ergänzte dann automatisch die "0". Da auch bei nicht überwiegend bettlägerigen Bewohnern eine Eingabe erfolgt war, kann hier ein systematischer Bias nicht ausgeschlossen werden. Daher wurden

diese beiden Variablen in der weiteren Untersuchung nicht berücksichtigt.

## 4.2 Datenauswertung

### 4.2.1 Das Common-Correlation-Modell

$k$	Anzahl der Gruppen $i$
$n_i$	Anzahl an Individuen $j$ in der $i$ -ten Gruppe
$\bar{n} = \sum_{i=1}^k n_i/k$	durchschnittliche Gruppengröße
$N = \sum_{i=1}^k n_i$	Gesamtzahl an Individuen
$X_{ij}(i = 1 \dots k; j = 1 \dots n_i)$	dichotome Antwort des $j$ -ten Individuums der $i$ -ten Gruppe
$Y_i = \sum_{j=1}^{n_i} X_{ij}$	Anzahl der "Erfolge" in der $i$ -ten Gruppe
$\pi = \Pr(X_{ij} = 1)$	Wahrscheinlichkeit für alle $i, j$ , dass das $j$ -te Individuum der $i$ -ten Gruppe das Merkmal 1 aufweist
$\hat{\pi} = (\sum_{i=1}^k Y_i)/N$	Schätzer für die Gesamt-Erfolgsrate
$\rho$	Korrelation zwischen jedem möglichen Antwortpaar $(X_{ij}, X_{il})(j \neq l)$ - diese ist in allen Clustern identisch und unabhängig von der Clustergröße = intraclass correlation

### 4.2.2 Varianzanalyse-Schätzer

Betrachtet man die ICC als den Anteil an der Gesamtvarianz, der durch die Cluster-Randomisierung entsteht, kann sie aus den mittleren Abweichungsquadraten zwischen  $MS_b$  und innerhalb  $MS_w$  der Cluster geschätzt werden:

$$\hat{\rho}_{AOV} = \frac{MS_b - MS_w}{MS_b + (n_0 - 1)MS_w}$$

mit

$$n_0 = \frac{1}{k-1} \sum_1^k n_i \quad \text{für Cluster mit identischer Größe}$$

bzw.

$$n_0 = \frac{1}{k-1} \left( N - \sum_{i=1}^k \frac{n_i^2}{N} \right) \quad \text{und} \quad N = \sum_{i=1}^k n_i \quad \text{für Cluster unterschiedlicher Größe.}$$

Für binäre Daten sind  $MS_b$  und  $MS_w$  wie folgt definiert (Ridout et al., 1999):

$$MS_b = \frac{1}{k-1} \left[ \sum_{i=1}^k \frac{Y_i^2}{n_i} - \frac{1}{N} \left( \sum_{i=1}^k Y_i \right)^2 \right]$$

$$MS_w = \frac{1}{N - k} \left( \sum_{i=1}^k Y_i - \sum_{i_1}^k \frac{Y_i^2}{n_i} \right)$$

Ridout et al. stellen auch einen von Fleiss modifizierten Schätzer vor (Fleiss, 1981). Dieser Schätzer soll hier nicht weiter verwendet werden, da ihm ein anderes Modell zugrunde liegt: Fleiss untersucht die Interrater-Reliabilität und stellt deren Übereinstimmung mit einer Modifikation des Varianz-Analyse-Schätzers dar.

### 4.2.3 Pearson-Schätzer

Nach Donner (1986) ist die älteste Methode für die Schätzung der ICC die Berechnung des Pearson-Korrelations-Koeffizienten. Da dieser großen Clustern zu großes Gewicht geben würde, schlagen Karlin, Cameron und Williams (1981) einen gewichteten Schätzer vor. Dabei stellen sie drei verschiedene Arten der Gewichtung vor, so dass die Clustergröße unterschiedlich berücksichtigt werden kann. Ridout, Demétrio und Firth (1999) beschreiben die unterschiedlich gewichteten Schätzer für dichotome Daten. Danach lautet die Grundform des gewichteten Pearson-Schätzers

$$\hat{\rho}_{PW} = \frac{\sum_{i=1}^k w_i Y_i (Y_i - 1) - \hat{\mu}^2}{\hat{\mu}(1 - \hat{\mu})}$$

mit

$$\hat{\mu} = \sum_{i=1}^k w_i (n_i - 1) Y_i$$

und der Voraussetzung, dass alle Gewichte  $w_i$  der Bedingung genügen

$$\sum_{i=1}^k n_i (n_i - 1) w_i = 1$$

Je nach Wahl des Gewichts  $w_i$  ergeben sich drei verschiedene Schätzer für die ICC.

- (i) Jedes Paar von Beobachtungen erhält das gleiche Gewicht.

Dies wird erreicht durch die Gewichtung

$$w_{PEQ} = \frac{1}{\sum_{i=1}^k n_i (n_i - 1)}$$

Mit dem sich daraus ergebenden

$$\hat{\mu}_{PEQ} = \frac{\sum_{i=1}^k (n_i - 1) Y_i}{\sum_{i=1}^k (n_i - 1) n_i}$$

lautet der Schätzer

$$\hat{\rho}_{PEQ} = \frac{\sum_{i=1}^k w_{PEQ} Y_i (Y_i - 1) - \hat{\mu}_{PEQ}^2}{\hat{\mu}_{PEQ} (1 - \hat{\mu}_{PEQ})}$$

(ii) Jedes Cluster erhält unabhängig von seiner Größe das gleiche Gewicht.

Dies wird erreicht durch die Gewichtung

$$w_{PGP} = \frac{1}{kn_i(n_i - 1)}$$

Mit dem sich daraus ergebenden

$$\hat{\mu}_{PGP} = \frac{1}{k} \sum_{i=1}^k \frac{Y_i}{n_i}$$

lautet der Schätzer

$$\hat{\rho}_{PGP} = \frac{\sum_{i=1}^k w_{PGP} Y_i (Y_i - 1) - \hat{\mu}_{PGP}^2}{\hat{\mu}_{PGP} (1 - \hat{\mu}_{PGP})}$$

(iii) Jedes Paar an Beobachtungen wird entsprechend der Anzahl an Paaren gewichtet, in denen jeweils eines der Individuen erscheint. Damit erhalten größere Gruppen ein kleineres Gewicht.

Dies wird erreicht durch die Gewichtung

$$w_{PPR} = \frac{1}{N(n_i - 1)}$$

Mit dem sich daraus ergebenden

$$\hat{\mu}_{PPR} = \frac{1}{N} \sum_{i=1}^k Y_i$$

lautet der Schätzer

$$\hat{\rho}_{PPR} = \frac{\sum_{i=1}^k w_{PPR} Y_i (Y_i - 1) - \hat{\mu}_{PPR}^2}{\hat{\mu}_{PPR} (1 - \hat{\mu}_{PPR})}$$

#### 4.2.4 Moment-Schätzer

Ridout et al. (1999) schlagen in Anlehnung an Kleinmann (1973) verschiedene Momentenschätzer für binäre Daten vor, die sich zum einen in der Gewichtung, zum anderen in der Berechnung des zweiten Moments unterscheiden.

Ausgehend von dem von Kleinmann (1973) für variierende Clustergrößen vorgeschlagenen Momentenschätzer

$$\hat{\rho} = \frac{s_w - \tilde{\pi}_w(1 - \tilde{\pi}_w) \sum_{i=1}^k \frac{w_i(1-w_i)}{n_i}}{\tilde{\pi}_w(1 - \tilde{\pi}_w) \left( \sum_{i=1}^k w_i(1 - w_i) - \sum_{i=1}^k \frac{w_i(1-w_i)}{n_i} \right)} \quad (1)$$

mit

$$\begin{aligned} \tilde{\pi}_i &= \frac{Y_i}{n_i} && \text{als Erfolgsrate im } i\text{-ten Cluster} \\ \tilde{\pi}_w &= \sum_{i=1}^k w_i \tilde{\pi}_i && \text{als Moment erster Ordnung} \\ s_w &= \sum_{i=1}^k w_i (\tilde{\pi}_i - \tilde{\pi}_w)^2 && \text{als Moment zweiter Ordnung} \end{aligned}$$

ergeben sich die Schätzer

$$\begin{aligned} \hat{\rho}_{KEQ} & \text{ bei gleicher Gewichtung aller Cluster mit } w_i = \frac{1}{k} \quad \text{und} \\ \hat{\rho}_{KPR} & \text{ bei Gewichtung der Cluster entsprechend ihrer Größe mit } w_i = \frac{n_i}{N}. \end{aligned}$$

Für beta-binomial-verteilte Daten mit einem erwarteten  $\rho < 0,8$  schlägt Kleinmann (1973) eine Variante vor, die bei der Schätzung des Moments zweiter Ordnung eine geringere quadratische Abweichung ergibt. Die modifizierten Schätzer  $\hat{\rho}_{KEQ}^*$  und  $\hat{\rho}_{KPR}^*$  werden berechnet, indem das  $s_w$  in der obigen Gleichung (1) ersetzt wird durch

$$s_w^* = (k - 1) \frac{s_w}{k} \quad .$$

#### 4.2.5 probabilistischer Schätzer

Unter dem zugrunde liegenden Modell ist die Wahrscheinlichkeit, dass zwei Individuen eines Clusters das gleiche Merkmal aufweisen, entweder "1" oder "0".

$$p_s = 1 - 2\pi(1 - \pi)(1 - \rho)$$

Stammen die Individuen aus unterschiedlichen Clustern, gilt  $\rho = 0$  (bedingt durch die Voraussetzung der Unabhängigkeit), und somit ergibt sich für die Wahrscheinlichkeit gleicher Merkmale bei Individuen unterschiedlicher Cluster

$$p_0 = 1 - 2\pi(1 - \pi)$$

Für  $\rho$  gilt dann (Mak, 1988)

$$\rho = \frac{p_s - p_0}{1 - p_0}$$

Für die Berechnung der ICC verwendet Mak die von Fleiss (1971) vorgestellten Schätzer für identische Clustergrößen  $n_i = n$  für alle  $i$

$$\hat{p}_s = 1 - \frac{2}{k} \sum_{i=1}^k \frac{Y_i(n - Y_i)}{n(n - 1)}$$

$$\hat{p}_0 = 1 - 2\hat{\pi}(1 - \hat{\pi})$$

mit

$$\hat{\pi} = \frac{\sum_{i=1}^k Y_i}{\sum_{i=1}^k n_i}$$

und somit

$$\hat{\rho} = 1 - \frac{\sum_{i=1}^k Y_i(n_i - Y_i)}{kn(n - 1)\hat{\pi}(1 - \hat{\pi})}$$

Für variierende Clustergrößen haben Fleiss und Cuzak (1979) den Schätzer adaptiert, indem sie  $\hat{p}_s$  proportional zu  $(n_i - 1)$  gewichten:

$$\hat{\rho}_{FC} = 1 - \frac{1}{(N - k)\hat{\pi}(1 - \hat{\pi})} \sum_{i=1}^k \frac{Y_i(n_i - Y_i)}{n_i}$$

Für große  $k$  ( $k \rightarrow \infty$ ) entspricht dieser Schätzer dem Varianz-Analyse-Schätzer.

Mak (1988) schätzt  $p_0$  mit einem erwartungstreuen Schätzer, der die Daten von zwei Gruppen einschließt

$$\hat{p}_0 = \frac{Y_i Y_j + (n_i - Y_i)(n_j - Y_j)}{n_i n_j} = 1 - \frac{Y_i(n_j - Y_j) + Y_j(n_i - Y_i)}{n_i n_j}$$

Dadurch wird berücksichtigt, dass die Individuen aus unterschiedlichen Clustern stammen. Als ungewichteter Durchschnitt über alle  $k(k-1)/2$  möglichen Paare ergibt sich daraus

$$\begin{aligned}\hat{\rho}_0^* &= 1 - \frac{2}{k-1} \sum_{i=1}^k \sum_{j=1}^{i-1} \frac{Y_i(n_j - Y_j) + Y_j(n_i - Y_i)}{n_i n_j} \\ &= 1 - \frac{2}{k(k-1)} \left( \sum_{i=1}^k \frac{Y_i}{n_i} \sum_{i=1}^k \frac{n_i - Y_i}{n_i} - \sum_{i=1}^k \frac{Y_i(n_i - Y_i)}{n_i^2} \right)\end{aligned}$$

Für  $p_s$  verwendet Mak den Schätzer von Fleiss (1971), wobei er  $n$  durch die jeweilige Clustergröße  $n_i$  ersetzt. Daraus ergibt sich der Schätzer

$$\hat{\rho}_{MAK} = 1 - \frac{(k-1) \sum_{i=1}^k \frac{Y_i(n_i - Y_i)}{n_i(n_i - 1)}}{\sum_{i=1}^k \frac{Y_i^2}{n_i^2} + \left( \sum_{i=1}^k \frac{Y_i}{n_i} \right) \left( k - 1 - \sum_{i=1}^k \frac{Y_i}{n_i} \right)}$$

Nach Ridout et al. (1999) konnte in Simulationsstudien gezeigt werden, dass eine weitere Adaption durch Verbindung des Mak-Schätzers für  $\hat{\rho}_0$  mit dem erweiterten Schätzer von Fleiss und Cuzak den beiden vorgestellten unterlegen ist, insbesondere bei kleinen Stichproben. Daher soll diese Möglichkeit hier nicht weiter betrachtet werden.

## 5 Ergebnis

Für die ausgewählten Variablen wurden mit der Statistik-Software R die oben beschriebenen verschiedenen ICC-Schätzer berechnet.

Da die Anzahl der Bewohner pro Wohnbereich zwischen unter 15 und über 50 schwankt, wurden die ICC zusätzlich getrennt nach Wohnbereichsgröße berechnet (Abschnitt 5.1).

Um zu zeigen, inwieweit die verschiedenen Schätzer voneinander abweichen, wird in Abschnitt 5.2 die Streuung der ICC-Werte über die Schätzer dargestellt.

In Abschnitt 5.3 werden diejenigen Schätzer aufgeführt, die die jeweils kleinste oder größte Ausprägung aufweisen.

Um zu untersuchen, inwiefern die Erfolgsrate der Variablen Einfluss auf die ICC hat und somit bei der Wahl des ICC-Schätzers zu berücksichtigen ist, wurden die Schätzer auch in Bezug auf diese Größe verglichen (Abschnitt 5.4).

Zuletzt wurde mit den aus der PiSaar-Studie vorliegenden Daten eine Simulationsstudie durchgeführt um zu untersuchen, inwieweit die ICC für  $k = 2$  oder  $k = 3$  zufällig ausgewählte Wohnbereiche variiert (Abschnitt 5.5).

## 5.1 ICC der einzelnen Variablen

Für die unter 4.2 beschriebenen Schätzer wurde die ICC sowohl über alle Wohnbereiche als auch getrennt für die unterschiedlichen Wohnbereichsgrößen entsprechend Tabelle 8 berechnet (siehe 4.1.1). Die Ergebnisse sind in Anhang ?? dargestellt.

Es zeigt sich, dass die ICC für die einzelnen Variablen höchst unterschiedlich ausfällt; abhängig von der Clustergröße und der Wahl des Schätzers liegt sie beispielsweise zwischen negativen Werten und 0,7459 für die Variable *Sturzrisiko* in der Gruppe der kleinen Cluster mit  $n_i \leq 20$ .

## 5.2 Streuung der ICC-Werte

Für die einzelnen Variablen wurde die Streuung der ICC über die untersuchten Schätzer ermittelt. Die Ergebnisse sind in Anhang ?? aufgeführt.

Auch in dieser Übersicht wird deutlich, dass es Variablen gibt, bei denen die Gruppenzugehörigkeit eine bedeutende Rolle spielt und somit bei der Fallzahlschätzung nicht unberücksichtigt bleiben kann.

Es entsteht der Eindruck, dass die Momentschätzer eher hohe ICC produzieren, während die Pearson-Schätzer die ICC unterschätzen. Da die Werte jeweils nah beieinander liegen, kann diese Schlussfolgerung jedoch nicht ohne Weiteres gezogen werden.

Im nächsten Schritt soll die Performanz der Schätzer in Abhängigkeit von der Clustergröße untersucht werden.

## 5.3 Vergleich der Schätzer

In der folgenden Übersicht ist aufgeführt, wie häufig die jeweiligen Schätzer unter Berücksichtigung der Clustergröße die kleinste bzw. größte ICC produziert haben.



$\hat{\rho}_{min}$	$12 \leq n_i \leq 56$	$n_i < 20$	$20 \leq n_i \leq 29$	$30 \leq n_i \leq 39$	$40 \leq n_i \leq 49$	$n_i \geq 50$	Summe
<i>AOV</i>							0
<i>PEQ</i>	20		32	24	30	14	120
<i>PGP</i>	20	24	10	19	14	35	122
<i>PPR</i>	2	20			2		22
<i>KEQ</i>							0
<i>KPR</i>	2						2
<i>KEQ*</i>	8		9	12	15	4	48
<i>KPR*</i>		6	2	2	4	1	15
<i>FC</i>	3	3	2	2	5	1	15
<i>MAK</i>							0

Tabelle 10: kleinste ICC-Schätzer nach Wohnbereichs-Größe

$\hat{\rho}_{max}$	$12 \leq n_i \leq 56$	$n_i < 20$	$20 \leq n_i \leq 29$	$30 \leq n_i \leq 39$	$40 \leq n_i \leq 49$	$n_i \geq 50$	Summe
<i>AOV</i>	2				2		4
<i>PEQ</i>	14	4	3	5			26
<i>PGP</i>							0
<i>PPR</i>							0
<i>KEQ</i>	18	3	26	24	32	14	117
<i>KPR</i>	14	24	16	21	21	36	132
<i>KEQ*</i>		21					21
<i>KPR*</i>							0
<i>FC</i>							0
<i>MAK</i>	7		12	1	3	1	24

Tabelle 11: größte ICC-Schätzer nach Wohnbereichs-Größe

Hier wird ersichtlich, dass die beiden Pearson-Schätzer  $\hat{\rho}_{PEQ}$  und  $\hat{\rho}_{PGP}$  die ICC besonders häufig unterschätzen, während die Momentschätzer  $\hat{\rho}_{KEQ}$  und  $\hat{\rho}_{KPR}$  auffallend oft den höchsten Schätzer liefern. Einzig der Varianzanalyseschätzer  $\hat{\rho}_{AOV}$  zeigt sich stabil, ebenso der Pearson-Schätzer  $\hat{\rho}_{PPR}$ , der die ICC lediglich für sehr kleine Cluster mit  $n_i \leq 20$

unterschätzt.

## 5.4 Einfluss der Erfolgsrate auf die ICC

Im Folgenden wird untersucht, welchen Einfluss die Erfolgsrate  $EQ_i = \frac{Y_i}{n_i}$  der einzelnen Variablen auf die Performanz der Schätzer hat, d.h. ob sich zeigt, dass ein Schätzer bei kleinen bzw. großen Erfolgsraten eher zu Extremwerten neigt als bei ausgewogenen.

Hierfür wurde für jede Variable der Mittelwert der Erfolgsraten  $\overline{EQ} = \frac{\sum_{i=1}^k EQ_i}{k}$  der einzelnen Wohnbereiche berechnet (siehe Anhang ??). Um eine klare Abgrenzung zu erhalten, wurde als kleine Erfolgsrate  $\overline{EQ} \leq 0,25$  zugrunde gelegt, für die mittlere Erfolgsrate  $0,375 \leq \overline{EQ} \leq 0,625$  und für eine große Erfolgsrate  $\overline{EQ} \geq 0,75$  gewählt.

Die Tabellen 12 und 13 zeigen, wie häufig die jeweiligen Schätzer die kleinste bzw. die größte ICC liefern. Zusätzlich ist in den Tabellen angegeben, wie viele Variablen jeweils die entsprechende Erfolgsrate aufweisen.

Hier wird deutlich, dass die Erfolgsrate keinen Einfluss auf die Performanz der Schätzer hat. Bei der Wahl eines geeigneten Schätzers für die Berechnung der ICC kann die erwartete Erfolgsrate somit unberücksichtigt bleiben.

$\hat{\rho}_{min}$	$12 \leq n_i \leq 56$ $k = 68$	$n_i < 20$ $k = 3$	$20 \leq n_i \leq 29$ $k = 34$	$30 \leq n_i \leq 39$ $k = 21$	$40 \leq n_i \leq 49$ $k = 7$	$n_i \geq 50$ $k = 3$	$\Sigma$
$\overline{EQ} \leq 0,25$	AOV 0 PEQ 5 PGP 7 PPR 0 KEQ 0 KPR 0 KEQ* 3 KPR* 0 FC 1 MAK 0 15 Variablen	AOV 0 PEQ 0 PGP 4 PPR 7 KEQ 0 KPR 0 KEQ* 0 KPR* 1 FC 2 MAK 0 14 Variablen	AOV 0 PEQ 7 PGP 1 PPR 0 KEQ 0 KPR 0 KEQ* 6 KPR* 1 FC 1 MAK 0 15 Variablen	AOV 0 PEQ 8 PGP 3 PPR 0 KEQ 0 KPR 0 KEQ* 5 KPR* 1 FC 1 MAK 0 15 Variablen	AOV 0 PEQ 9 PGP 5 PPR 1 KEQ 0 KPR 0 KEQ* 3 KPR* 1 FC 1 MAK 0 15 Variablen	AOV 0 PEQ 9 PGP 9 PPR 0 KEQ 0 KPR 0 KEQ* 0 KPR* 1 FC 1 MAK 0 19 Variablen	0 38 29 8 0 0 17 5 7 0

$\hat{\rho}_{min}$	$12 \leq n_i \leq 56$ $k = 68$	$n_i < 20$ $k = 3$	$20 \leq n_i \leq 29$ $k = 34$	$30 \leq n_i \leq 39$ $k = 21$	$40 \leq n_i \leq 49$ $k = 7$	$n_i \geq 50$ $k = 3$	$\Sigma$
$0,375 \leq \overline{EQ} \leq 0,625$	AOV 0	AOV 0	AOV 0	AOV 0	AOV 0	AOV 0	0
	PEQ 7	PEQ 0	PEQ 15	PEQ 10	PEQ 15	PEQ 0	47
	PGP 8	PGP 14	PGP 4	PGP 10	PGP 5	PGP 11	52
	PPR 2	PPR 4	PPR 0	PPR 0	PPR 1	PPR 0	7
	KEQ 0	KEQ 0	KEQ 0	KEQ 0	KEQ 0	KEQ 0	0
	KPR 0	KPR 0	KPR 0	KPR 0	KPR 0	KPR 0	0
	KEQ* 4	KEQ* 0	KEQ* 2	KEQ* 4	KEQ* 5	KEQ* 4	19
	KPR* 0	KPR* 2	KPR* 0	KPR* 0	KPR* 1	KPR* 0	3
	FC 1	FC 1	FC 0	FC 0	FC 2	FC 0	4
	MAK 0	MAK 0	MAK 0	MAK 0	MAK 0	MAK 0	0
	21 Variablen	19 Variablen	20 Variablen	21 Variablen	20 Variablen	11 Variablen	
$\overline{EQ} \geq 0,75$	AOV 0	AOV 0	AOV 0	AOV 0	AOV 0	AOV 0	0
	PEQ 4	PEQ 0	PEQ 5	PEQ 4	PEQ 1	PEQ 2	16
	PGP 1	PGP 5	PGP 2	PGP 1	PGP 0	PGP 4	13
	PPR 0	PPR 4	PPR 0	PPR 0	PPR 0	PPR 0	4
	KEQ 0	KEQ 0	KEQ 0	KEQ 0	KEQ 0	KEQ 0	0
	KPR 0	KPR 0	KPR 0	KPR 0	KPR 0	KPR 0	0
	KEQ* 1	KEQ* 0	KEQ* 1	KEQ* 1	KEQ* 5	KEQ* 0	8
	KPR* 0	KPR* 2	KPR* 0	KPR* 1	KPR* 2	KPR* 0	5
	FC 1	FC 0	FC 0	FC 1	FC 2	FC 0	4
	MAK 0	MAK 0	MAK 0	MAK 0	MAK 0	MAK 0	0
	7 Variablen	10 Variablen	7 Variablen	6 Variablen	6 Variablen	6 Variablen	

Tabelle 12: kleinste ICC-Schätzer nach Erfolgsrate

$\hat{\rho}_{max}$	$12 \leq n_i \leq 56$ $k = 68$	$n_i < 20$ $k = 3$	$20 \leq n_i \leq 29$ $k = 34$	$30 \leq n_i \leq 39$ $k = 21$	$40 \leq n_i \leq 49$ $k = 7$	$n_i \geq 50$ $k = 3$	$\Sigma$
$\overline{EQ} \leq 0,25$	AOV 2	AOV 0	AOV 0	AOV 0	AOV 2	AOV 0	4
	PEQ 2	PEQ 3	PEQ 1	PEQ 3	PEQ 0	PEQ 0	9
	PGP 0	PGP 0	PGP 0	PGP 0	PGP 0	PGP 0	0
	PPR 0	PPR 0	PPR 0	PPR 0	PPR 0	PPR 0	0
	KEQ 3	KEQ 1	KEQ 3	KEQ 8	KEQ 10	KEQ 9	34
	KPR 8	KPR 3	KPR 5	KPR 3	KPR 4	KPR 9	32
	KEQ* 0	KEQ* 7	KEQ* 0	KEQ* 0	KEQ* 0	KEQ* 0	7
	KPR* 0	KPR* 0	KPR* 0	KPR* 0	KPR* 0	KPR* 0	0
	FC 0	FC 0	FC 0	FC 0	FC 0	FC 0	0
	MAK 2	MAK 0	MAK 6	MAK 1	MAK 2	MAK 1	12
	15 Variablen	14 Variablen	15 Variablen	15 Variablen	15 Variablen	19 Variablen	
$0,375 \leq \overline{EQ} \leq 0,625$	AOV 0	AOV 0	AOV 0	AOV 0	AOV 0	AOV 0	0
	PEQ 9	PEQ 1	PEQ 1	PEQ 1	PEQ 0	PEQ 0	12
	PGP 0	PGP 0	PGP 0	PGP 0	PGP 0	PGP 0	0
	PPR 0	PPR 0	PPR 0	PPR 0	PPR 0	PPR 0	0
	KEQ 7	KEQ 1	KEQ 14	KEQ 10	KEQ 15	KEQ 0	47
	KPR 3	KPR 15	KPR 5	KPR 10	KPR 7	KPR 11	51
	KEQ* 0	KEQ* 3	KEQ* 0	KEQ* 0	KEQ* 0	KEQ* 0	3
	KPR* 0	KPR* 0	KPR* 0	KPR* 0	KPR* 0	KPR* 0	0
	FC 0	FC 0	FC 0	FC 0	FC 0	FC 0	0
	MAK 4	MAK 0	MAK 0	MAK 0	MAK 0	MAK 0	4
	21 Variablen	19 Variablen	20 Variablen	21 Variablen	20 Variablen	11 Variablen	
$\overline{EQ} \geq 0,75$	AOV 0	AOV 0	AOV 0	AOV 0	AOV 0	AOV 0	0
	PEQ 1	PEQ 0	PEQ 1	PEQ 0	PEQ 0	PEQ 0	2
	PGP 0	PGP 0	PGP 0	PGP 0	PGP 0	PGP 0	0
	PPR 0	PPR 0	PPR 0	PPR 0	PPR 0	PPR 0	0
	KEQ 4	KEQ 5	KEQ 5	KEQ 4	KEQ 3	KEQ 2	23
	KPR 1	KPR 5	KPR 0	KPR 2	KPR 4	KPR 4	16
	KEQ* 0	KEQ* 0	KEQ* 0	KEQ* 0	KEQ* 0	KEQ* 0	0
	KPR* 0	KPR* 0	KPR* 0	KPR* 0	KPR* 0	KPR* 0	0
	FC 0	FC 0	FC 0	FC 0	FC 0	FC 0	0
MAK 1	MAK 0	MAK 2	MAK 0	MAK 0	MAK 0	4	
	7 Variablen	10 Variablen	7 Variablen	6 Variablen	6 Variablen	6 Variablen	

Tabelle 13: größte ICC-Schätzer nach Erfolgsrate

## 5.5 Simulationsstudie

In einer realen Forschungssituation wird man zur Bestimmung der ICC im Rahmen einer Vorstudie keine 62 Stationen oder Wohnbereiche untersuchen. Daher soll im Rahmen einer Simulationsstudie untersucht werden, in welchem Rahmen sich die ICC bewegt, wenn zufällig zwei oder drei Wohnbereiche ausgewählt werden.

Hierfür wurden aus der gesamten Gruppe jeweils alle möglichen Kombinationen aus  $m = 2$  bzw.  $m = 3$  Wohnbereichen gebildet; wo mehr als 1000 Kombinationen möglich sind, wurden per Zufallsgenerator 1000 Kombinationen ausgewählt (siehe Tabelle 14). Für alle Kombinationen wurde die ICC bestimmt und daraus jeweils Mittelwert, Standardabweichung sowie das 95%-Konfidenzintervall ermittelt.

Kombinationen	$m = 2$	$m = 3$
$12 \leq n_i \leq 56$ (k=62)	1000	1000
$n_i \leq 20$ (k=3)	6	entfällt
$20 \leq n_i \leq 29$ (k=34)	561	1000
$30 \leq n_i \leq 39$ (k=21)	210	1000
$40 \leq n_i \leq 49$ (k=7)	21	35
$n_i \geq 50$ (k=3)	6	entfällt

Tabelle 14: Anzahl der Kombinationen

Die Ergebnisse sind im Anhang ?? dargestellt.

Die Simulationsstudie zeigt ein sehr indifferentes Ergebnis. So liegt für das Item "Sturzrisiko" der Mittelwert der ICC für Wohnbereiche mit weniger als 20 Bewohnern bei  $\mu = 0,3831$  und für Wohnbereich mit 50 und mehr Bewohnern bei  $\mu = 0,0394$ . Für das Item "benötigt emotionale Unterstützung" liegt die ICC für Wohnbereiche mit weniger als 20 Bewohnern mit  $\mu = 0,0183$  weit unter dem Wert für alle Wohnbereiche von  $\mu = 0,2839$ .

Diese Schwankungen entstehen durch einzelne Wohnbereiche mit abweichenden Werten, z.B. spezialisierte Wohnbereiche. So können Bewohner auf Wohnbereichen, die nach dem Werdenfeller Weg arbeiten, ein höheres Sturzrisiko aufweisen als diejenigen Bewohner, die fixiert werden. In einer gerontopsychiatrischen Einrichtung leben vielleicht mehr Bewohner, die emotionale Unterstützung benötigen, oder der Bedarf wird von den gerontopsychiatrisch geschulten Pflegekräften eher diagnostiziert. Durch die vielen Kombinationsmöglichkeiten wird jeder Wohnbereich mehrfach in der Simulationsstudie berücksichtigt und kann

somit in einer kleinen Stichprobe (bezogen auf die Anzahl der Wohnbereiche) zu extremen Werten der ICC führen.

Die insgesamt hohe Streuung verdeutlicht, dass es sowohl sehr ähnliche als auch sehr unterschiedliche Cluster gibt, so dass eine allgemeingültige Aussage zu den ICC für die hier untersuchten Erkrankungen und pflegerelevanten Phänomene nicht möglich ist.

Es wird jedoch ersichtlich, dass die ICC sehr hohe Werte annehmen kann und daher bei der Stichprobenberechnung unbedingt zu berücksichtigen ist.

Der Vergleich der durchschnittlichen ICC von  $m = 2$  und  $m = 3$  Wohnbereichen zeigt, dass die mittlere ICC zwischen drei Wohnbereichen höher liegt als diejenige zwischen zwei Wohnbereichen, dafür jedoch weniger streut und ein engeres 95%-Konfidenzintervall aufweist. Die einzige Ausnahme stellen die Momentschätzer  $\hat{\rho}_{KPR}$  und  $\hat{\rho}_{KPQ}$  dar, die für die Wohnbereiche mit 20 bis 29 Bewohnern die ICC zwischen drei Wohnbereichen geringer schätzen als die zwischen zwei Wohnbereichen bei ebenfalls geringerer Streuung und engerem 95%-Konfidenzintervall.

## 6 Ergebnisse anderer Studien

In diesem Abschnitt werden drei Studien vorgestellt, die die ICC berichten.

Kron (2003) untersucht auf Grundlage der Ulmer Sturzstudie den Einfluss der ICC auf die Inzidenzrate.

Zwei Studien (Meyer et al. (2009) und Köpke et al. (2012)) untersuchen Fixierungen in deutschen Altenheimen.

### 6.1 Ulmer Sturz-Studie

Die Ulmer Sturzstudie untersucht den Einfluss einer multifaktoriellen nicht-pharmazeutischen Intervention auf die Sturz- und Fraktur-Inzidenzrate in Ulmer Altenheimen.

Das Interventionsprogramm umfasste

- Personalschulung (60-minütiger Unterricht mit schriftlichem Material über Inzidenz und Konsequenzen von Stürzen, Risikofaktoren, präventive Maßnahmen sowie Möglichkeit zur Diskussion von Problemen während der Besuche der Study-Nurse, Einrichtung einer Telefon-Hotline)

- Bewohnerinformation (Ausgabe von schriftlichem Informationsmaterial zur Sturzprävention, persönliches Gespräch mit Study-Nurse oder Trainer über Sturzprävention)
- Umgebungsanpassung (Durchführung eines Umgebungsschecks unter besonderer Berücksichtigung der Punkte Beleuchtung, Stuhl- und Betthöhen, Oberflächen der Fußböden, Haltegriffe an Toiletten und Badewannen sowie Gehhilfen mit Diskussion der Ergebnisse seitens des Personals und der Heimleitung)
- Training (Programm bestehend aus einer Kombination von Balanceübungen und Krafttraining als Gruppenprogramm mit Wiederholungsmöglichkeiten)
- Hüftprotektoren (Ausgabe von fünf Hüftprotektoren pro Bewohner mit der Empfehlung, einen Hüftprotektor vom Aufstehen bis zum Schlafengehen zu tragen)

In die Studie eingeschlossen wurden 981 Alten- und Pflegeheimbewohner aus sechs Ulmer Alten- und Pflegeheimen. Um Kontaminationseffekte zu vermeiden, wurden nicht die Bewohner randomisiert, sondern die Einrichtungen. So wurden zufällig je drei Heime der Interventions- und der Kontrollgruppe zugeteilt. Den Bewohnern der Kontrollheime wurden keine Interventionen angeboten; die der Interventionsheime konnten 12 Monate lang an allen angebotenen Interventionsmaßnahmen teilnehmen.

In die Studie aufgenommen wurden Bewohner, die mindestens 60 Jahre alt waren und während der Laufzeit der Studie mindestens vier Wochen in einem der Alten- und Pflegeheime wohnten. Auch Bewohner, die erst während der Laufzeit in eines der Heime einzogen, wurden in die Studie eingeschlossen.

In die Auswertung aufgenommen wurden lediglich diejenigen Bewohner, bei denen eine Beobachtungsdauer von mindestens 90 Tagen erreicht wurde, wobei die Anzahl der beobachteten Ereignisse auf ein Jahr hochgerechnet wurde, wenn der Beobachtungszeitraum unter einem Jahr lag. Somit wurden die Daten von 891 Bewohnern ausgewertet (Beschreibung der Studienteilnehmer siehe Tabelle 15).

Die Studienteilnehmer waren sich nicht nur hinsichtlich Alter, Geschlecht und Mortalität sehr ähnlich sondern auch hinsichtlich der Risikofaktoren für Stürze. Einzig der Anteil der Bewohner mit mindestens vier Medikamenten war in der Interventionsgruppe (55,3%) deutlich höher als in der Kontrollgruppe (39,2%).

	Interventionsgruppe	Kontrollgruppe	Gesamt
Anzahl Bewohner	448	443	891
Bewohnerjahre	384,1	382,3	766,4
durchschnittliche Aufenthaltsdauer pro Bewohner [Tage]	329,9	315,0	314,0
Alter $\bar{x} \pm s$	83,7 $\pm$ 7,4	84,5 $\pm$ 6,8	84,1 $\pm$ 7,1
Median (Min-Max)	85 (60-99)	85 (60-104)	85 (60-104)
weibliches Geschlecht	360 (80,4%)	343 (77,5%)	703 (78,9%)
Mortalität während der Studie	71 (15,8%)	69 (15,6%)	140 (15,7%)

Tabelle 15: Beschreibung der Studienteilnehmer (aus Kron, 2003)

Im Rahmen der Studie wurden die Anzahl der Stürze, der Stürzer, der Bewohner mit mehr als zwei Stürzen sowie Hüftfrakturen und Nicht-Hüftfrakturen erfasst. Da die ICC in der Studie lediglich für Sturz- und Frakturinzidenz dargestellt wird, werden auch nur deren Ergebnisse in Tabelle 16 aufgeführt.

	Interventionsgruppe		Kontrollgruppe		RR	95% KI	p-Wert* (2-seitig)
	n=509	Inzidenz pro 1000 BJ (ges. 391 BJ)	n=472	Inzidenz pro 1000 BJ (ges. 383 BJ)			
Stürze	547	1399	980	2558	0,55	[0,41 ; 0,73]	< 0,001
Hüft-Frakturen	17	43	15	39	1,11	[0,49 ; 2,51]	0,801
Nicht-Hüftfrakturen	16	41	20	52	0,78	[0,57 ; 1,07]	0,128

BJ = Bewohnerjahre

RR = Relatives Risiko

KI = Konfidenz-Intervall

Tabelle 16: Inzidenz von Stürzen und Frakturen in der Studienpopulation (aus Kron, 2003)

Im Folgenden werden die Ergebnisse der Schätzungen der ICC der Stürze sowie der Frakturen für in dieser Arbeit vorgestellte Schätzer übernommen (Tabellen 17 und 18):



	Interventionsgruppe (n=448)	Kontrollgruppe (n=443)	Gesamt (n=891)
$\hat{\rho}_{AOV}$	-0,003	0,013	0,028
$\hat{\rho}_{PEQ}$	-0,004	0,007	0,024
$\hat{\rho}_{PGP}$	-0,004	0,006	0,020
$\hat{\rho}_{PPR}$	-0,004	0,007	0,023
$\hat{\rho}_{KEQ}$	0,007	0,109	0,212

(aus Kron (2003))

Tabelle 17: Schätzungen für die ICC der Stürze (nur Bewohner mit mindestens 90 Tagen Beobachtungszeit) (aus Kron, 2003)

	Interventionsgruppe (n=448)	Kontrollgruppe (n=443)	Gesamt (n=891)
$\hat{\rho}_{AOV}$	-0,002	0,004	0,000
$\hat{\rho}_{PEQ}$	-0,003	0,001	-0,001
$\hat{\rho}_{PGP}$	-0,003	0,001	-0,001
$\hat{\rho}_{PPR}$	-0,003	0,001	-0,001
$\hat{\rho}_{KEQ}$	-0,001	0,006	0,002

Tabelle 18: Schätzungen für die ICC der Frakturen (nur Bewohner mit mindestens 90 Tagen Beobachtungszeit) (aus Kron, 2003)

## 6.2 Fixierung

Meyer et al. (2009) untersuchten im Rahmen einer Querschnittstudie die Prävalenz von Fixierungen und Psychopharmaka in Hamburger Altenheimen sowie in einer prospektiven Studie deren neue Anordnung innerhalb eines Jahres nach der Ersterhebung.

In die Erhebung eingeschlossen wurden 30 Heime (Merkmale siehe Tabelle 19) mit 2367 Bewohner/innen für die Querschnittstudie bzw. 2353 Bewohner/innen für die prospektive Studie.

Für die Querschnittstudie wurden die Prävalenzdaten erhoben, indem für jeden Bewohner/jede Bewohnerin an einem Tag zu drei Zeitpunkten (10 Uhr, 15 Uhr und 20 Uhr) von geschulten externen Beobachtern dokumentiert wurde, ob eine Fixierung vorlag in Form

von beidseitigen Bettgittern (oder einseitigen Bettgittern, wenn das Bett mit der anderen Seite an der Wand stand), einem Gurt im Stuhl, einem Gurt im Bett, einem festen Rollstuhl-Therapietisch sowie andere Maßnahmen wie Decken oder Laken und Ellbogen- oder Fußfixierungen.

	(n=30)
Träger der Einrichtung	
- privat	12 (40)
- städtisch	1 (3)
- kirchlich	6 (20)
- gemeinnützig	11 (37)
Einrichtungen mit mindestens einer gerontopsychiatrischen Abteilung	10 (33)
Bewohner pro Heim	80 ± 51 (19-210)
Bewohner pro Pflegekraft	2,4 ± 0,7 (1,1-4,0)
Fachkraft-Quote	58 ± 7 (43-69)
Bewohner pro Nachtschwester	36 ± 14 (16-62)
standardisierte Dokumentation von Fixierungen	
hausinterner Pflegestandard zu Fixierungen	18 (60)
spezieller Dokumentationsbogen	24 (40)

Die Werte sind absolute (relative) Häufigkeiten oder Mittelwert ± Standardabweichung (Spannweite).

Tabelle 19: Merkmale der Altenheime zu Beginn der Querschnitt-Studie (aus Meyer et al., 2009)

In der anschließenden zwölfmonatigen Prospektiv-Studie wurde für jeden Bewohner/jede Bewohnerin vier Wochen lang in jeder Schicht auf einem speziellen Bogen die Anwendung von Fixierungen dokumentiert, wobei hier lediglich Bettgitter, Gurte und Therapietische erfasst wurden, da die Pilot-Studie gezeigt hatte, dass andere Maßnahmen erheblichen Interpretationsspielraum boten und somit die Validität gefährdeten. Bedarfsmedikation in Form von Psychopharmaka wurde von den Wohnbereichsleitungen aus den Bewohner-Akten erhoben.

Die ICC wird angegeben für die Häufigkeit von mechanischen Fixierungen sowie Psychopharmaka als Bedarfsmedikation (siehe Tabellen 20 und 21).

	% (95% Konfidenz-Intervall)	ICC
<b>Querschnittstudie (n = 2367)</b>		
Bewohner mit ≥ 1 Fixierung	26,2 (21,3-31,1)	0,0818

Bewohner mit beobachteter Anwendung von		
- restriktiven Bettgittern	24,5 (19,5-29,5)	0,0906
- Taillengurt im Stuhl oder im Bett	2,7 (1,6-3,9)	0,0230
- Stuhl mit Tisch	2,1 (1,3-2,9)	0,0115
- andere Maßnahme	2,3 (0,8-3,9)	0,0593
<b>12-monatige Anschlussstudie (n = 2353)</b>		
Bewohner mit $\geq 1$ Fixierung	39,8 (33,6-45,9)	0,1074
Bewohner mit dokumentierter Anwendung von		
- restriktiven Bettgittern	38,5 (32,2-44,8)	0,1170
- Taillengurt im Stuhl oder im Bett	8,9 (5,8-12)	0,0647
- Stuhl mit Tisch	9,9 (7-12,7)	0,0569
Beobachtungstage mit dokumentierter Anwendung von		
- restriktiven Bettgittern (n = 870)	27,7 (22,5-33)	0,1065
- Taillengurt im Stuhl oder im Bett (n = 233)	3,1 (1,7-4,5)	0,0475
- Stuhl mit Tisch (n = 227)	2,2 (1,3-3,1)	0,0312
Beobachtungstage innerhalb der Untergruppe fixierter Bewohner mit dokumentierter Anwendung von		
- restriktiven Bettgittern	70,4 (65,9-74,8)	0,0593
- Taillengurt im Stuhl oder im Bett	33,8 (24,3-43,2)	0,2204
- Stuhl mit Tisch	22,9 (16,1-29,7)	0,1794

Tabelle 20: Häufigkeit mechanischer Fixierungen (aus Meyer et al., 2009)

	% (95% Konfidenz-Intervall)	ICC
<b>Querschnitt-Daten der verordneten Medikation (n = 2367)</b>		
Bewohner mit $\geq 1$ Psychopharmakon	52,4 (48,7-56,1)	0,026
Bewohner mit $\geq 1$		
- antipsychotischen Medikament	28,4 (24,2-32,7)	0,0554
- atypischen antipsychotischen Medikament	10,7 (8,5-13)	0,0250
- Anxiolytikum	12,9 (10,6-15,1)	0,0219
- Hypnotikum	10,7 (8,9-12,5)	0,0116

-Antidepressivum	20,1 (17,1-23,6)	0,0096
Bewohner mit zwei oder mehr Medikamenten	20,3 (17-23,6)	0,0388
Verordnungen pro Bewohner*	0,8 ± 2,1 (0-5)	0,0510
Verordnungen pro Bewohner mit Psychopharmaka*	1,5 ± 1,2(1-5)	0,0372
<b>12-monatige Anchlusserhebung der Bedarfsmedikation (n = 2353)</b>		
Bewohner mit ≥ 1 Medikament	5 (2,7-7,2)	0,0816
Gesamt-Zahl an Verordnungen	704	
Verordnungen pro Bewohner*	0,3 ± 5,7 (0-228)	0,0029

\* Mittelwert ± Standardabweichung (Spannweite).

Tabelle 21: Psychopharmaka (nach Meyer et al., 2009)

Köpke et al. (2012) veröffentlichen die ICC für die im Rahmen einer RCT erhobenen Prävalenz von Fixierungen (Tabelle 23). In der Studie wurden die Auswirkungen einer multifaktoriellen Intervention auf die Anwendung von direkten Fixierungen in Altenheimen untersucht. Dies beinhaltet beidseitige Bettgitter, Gurte, feste Rollstuhl-Therapietische sowie andere physikalische Fixierungen.

	Interventionsgruppe (n=18)	Kontrollgruppe (n=18)
Teilnehmer pro Altenheim	107 ± 44 (47-184)	99 ± 40 (53-215)
Träger der Einrichtung		
- gemeinnützig	8 (44)	5 (28)
- kirchlich	5 (28)	9 (50)
- privat	5 (28)	4 (22)
Fachkraft-Quote	58 ± 7 (49-76)	57 ± 11 (40-76)
Altenheime mit gerontopsychiatrischer Abteilung	9 (50)	8 (44)
Teilnehmer pro Pflegekraft	3,1 ± 1,3 (1,7-7,0)	3,7 ± 7 (1,0-4,5)
standardisierte Dokumentation von Fixierungen	15 (83)	16 (89)
hausinterner Standard zu Fixierungen	10 (56)	17 (94)

Die Werte sind absolute (relative) Häufigkeiten oder Mittelwert ± Standardabweichung (Spannweite).

Tabelle 22: Merkmale der Altenheime zu Studienbeginn (nach Köpke et al., 2012)

In die Studie eingeschlossen wurden 36 Altenheime aus Hamburg und Witten mit einer Fixierungsrate von über 20 %, wovon jeweils 18 Einrichtungen zufällig der Interventions- und der Kontrollgruppe zugeordnet wurden (Tabelle 22).

Für die Durchführung wurde zunächst eine 290-seitige evidenzbasierte Richtlinie entwickelt. Neben den Empfehlungen und einer umfassenden Beschreibung der Evidenz enthält sie Hintergrundinformationen über freiheitsentziehende Maßnahmen, Stürze und damit zusammenhängende Verletzungen sowie die rechtlichen Hintergründe.

Die Intervention beinhaltete folgende Komponenten:

- Erklärung der Einrichtung, sich für die Ziele der Studie, also die Vermeidung direkter Fixierungen, einzusetzen.
- strukturierte 90-minütige Schulung für alle Pflegekräfte
  - Definition direkter Fixierung
  - erwünschte und unerwünschte Wirkungen von direkten Fixierungen
  - Entwicklung der Richtlinie und die daraus resultierenden Empfehlungen
  - Einstellungen und Erfahrungen der Pflegekräfte
  - Alternative Maßnahmen zur Vermeidung von Fixierungen
- externer strukturierter Tagesworkshop für ausgewählte Fachkräfte verschiedener Einrichtungen
  - Austausch und Diskussion
  - Vertiefendes Arbeiten mit der Richtlinie
  - Entwicklung einer hauseigenen Agenda für die Verringerung von Fixierungen
- strukturierte Unterstützung für die ausgewählten Fachkräfte
  - Tagebuch zur Dokumentation der täglichen Aktivitäten und Probleme der Maßnahmen-Implementierung
  - monatlicher Kontakt durch Projekt-Mitarbeiter in den ersten drei Monaten
- unterstützendes Material
  - je ein Exemplar der Richtlinie für die ausgewählte Fachkraft sowie die Pflegedienstleitung

- 16-seitige Kurzversion der Richtlinie für alle Pflegekräfte sowie die gesetzlichen Betreuer und Angehörigen, die sich um die rechtlichen Belange kümmern
  - Flyer in Form eines Leporello für alle Pflegekräfte und Angehörigen mit Informationen über die Ziele des Projekts
  - Material mit Logo und/oder Slogan "Riskier mehr Freiheit", Poster, Kugelschreiber und Post-its für die Pflegekräfte, die an der Schulung teilgenommen haben, Becher für die ausgewählten Fachkräfte
- Öffentlichkeitsarbeit
- Die Fachkräfte wurden ermutigt, die Erklärung zusammen mit Informationen im Foyer der Einrichtung auszuhängen.
  - Poster mit Projektlogo und Slogan wurden überall in den Einrichtungen aufgehängt.
  - Förderung individueller Aktivitäten, z.B. Informationsabende für Angehörige

In der Kontrollgruppe erhielten die Pflegedienstleitungen eine schriftliche Information in Form von drei 12- bis 24-seitigen Broschüren über den Einsatz von Fixierungen und Methoden zu ihrer Vermeidung. Ergänzend wurde das Thema Fixierung im Rahmen einer kurzen Präsentation durch einen der Forscher diskutiert.

Die Prävalenz von Fixierungen wurde zu Beginn des Projekts ( $T_0$ ), nach drei Monaten ( $T_1$ ) sowie nach sechs Monaten ( $T_2$ ) erhoben. In Tabelle 23 sind die Ergebnisse sowie die dazu ermittelten ICC dargestellt.

	Interventionsgruppe	Kontrollgruppe	Differenz	95%-CI	ICC
$T_0$	31,5 %	30,6 %			
$T_1$	23,9 %	30,5 %	6,6	[0,6 % ; 12,6 %]	0,029
$T_2$	22,6 %	29,1 %	6,5	[0,6 % ; 12,4 %]	0,029

Tabelle 23: Prävalenz von Fixierungen (nach Köpke et al., 2012)

## 7 Diskussion

### 7.1 Studiendesign

Es wurde eine Sekundäranalyse der im Rahmen der PiSaar-Studie erhobenen personenbezogenen Daten durchgeführt. Die Daten lagen dichotomisiert vor. Über die ICC sollte gezeigt werden, welchen Einfluss die Cluster-Zugehörigkeit auf die Ausprägung der untersuchten Variablen hat.

Allerdings spiegeln die Erfolgsraten der Variablen keine objektiven Inzidenzen wieder, sondern jeweils die Einschätzung einer Pflegefachkraft, inwieweit eine Fähigkeit, ein Risiko oder ein Zustand der pflegebedürftigen Person vorliegen. Es konnte gezeigt werden, dass dieser Faktor großen Einfluss auf die Erfolgsrate der Variablen hat. So liegt die ICC bei objektiv überprüfbareren Variablen wie "Hat der Bewohner einen Blasendauerkatheter?", "Hat der Bewohner einen suprapubischen Blasenkateter?" oder "Hat der Bewohner eine PEG?" konstant bei unter 0,06, während sie bei der Variablen "Wirkt der Bewohner als benötigte er besondere emotionale Unterstützung?" Werte bis 0,5 annimmt.

Hierin zeigt sich zum einen der Nachteil einer Sekundäranalyse – die Daten wurden für einen anderen Zweck erhoben und die Variablen entsprechend kodiert. Zum anderen ergibt sich daraus, dass die ICC von Variablen, die auf der subjektiven Einschätzung einer Pflegekraft beruhen, nicht ohne weiteres aus anderen Studien übernommen werden kann.

Sicherlich wäre ein Item "Wirkt der Bewohner als benötigte er besondere emotionale Unterstützung?" niemals Bestandteil einer durch Hypothesentesten zu überprüfenden Fragestellung. Etliche Instrumente, z.B. die Beobachtungsskala zur Beurteilung von Schmerzen bei Demenz (BESD) (Warden et al., 2003; Basler et al., 2006), vermitteln jedoch durch Punktvergabe den Eindruck eines objektivierbaren Ergebnisses, beruhen dabei aber auf der subjektiven Beobachtung und Einschätzung der Pflegefachkraft. Hier wäre beispielsweise von einer hohen ICC auszugehen, die neben Organisationsvariablen vor allem auf die einschätzende Person zurück zu führen ist.

### 7.2 Hypothesentesten - Möglichkeiten und Grenzen

Das Hypothesentesten ist eine von vielen Forschungsmethoden. Sie ist schnell und einfach durchzuführen und dann geeignet, wenn die Wirksamkeit einer Intervention an objektiv messbaren oder objektivierbaren Endpunkten festgemacht werden kann, z.B. biometrische Daten oder Inzidenzen.

Hierbei ist von entscheidender Bedeutung, dass die Methode richtig verstanden und angewendet wird: Die Wirksamkeit der Intervention kann nicht bewiesen werden. Es wird lediglich die Wahrscheinlichkeit  $\alpha$  ermittelt, mit der die untersuchte Gruppe einer Grundgesamtheit angehört, auf die die  $H_0$ -Hypothese zutrifft. Die Power  $1 - \beta$  besagt, mit welcher Wahrscheinlichkeit die Stichprobe einer Grundgesamtheit angehört, auf die die  $H_1$ -Hypothese zutrifft.

Die Entscheidung, die statistische Hypothese  $H_1$  beizubehalten, ist kein Beweis der Forschungshypothese, sondern unterstützt lediglich die Entscheidung davon auszugehen, dass die Beobachtungen der Untersuchung auf die gesamte Population übertragbar sind. Daraus ergibt sich keinesfalls die Verpflichtung, sie umzusetzen, sondern vielmehr sie als Regelwissen in die Entscheidungsfindung im jeweiligen Einzelfall einzubeziehen.

Es kann kritisch hinterfragt werden, inwiefern dabei die konventionellen Grenzwerte der Bezugswissenschaften für  $\alpha = 0,05$  oder  $0,1$  und  $\beta = 0,2$  übernommen werden sollen. Vielmehr scheint es sinnvoll, aus einer Kosten-Nutzen-Abwägung heraus für den jeweiligen Forschungsgegenstand festzulegen, bis zu welchen Werten die untersuchte Intervention als wirksam zu betrachten ist. Hierbei sind mit Kosten selbstverständlich nicht nur monetärer Aufwand sondern auch mögliche Risiken und Nebenwirkungen gemeint. Insbesondere wenn die für eine entsprechende Power erforderliche Stichprobengröße nicht erreicht werden kann, bietet sich eine Kompromiss-Analyse an (siehe Kapitel 3.3). Unabhängig davon sind die gewählten Werte für das Signifikanzniveau sowie für die Power der Untersuchung immer im Vorfeld festzulegen und zu begründen.

Gerade im Prozess der Etablierung der Pflegewissenschaft als eigenständige Wissenschaftsdisziplin scheint es angebracht, sich von den wissenschaftlich nicht belegten konventionellen Werten der Medizin zu lösen. Konventionelle Werte, die allen Studien gleichermaßen zu Grunde gelegt werden, haben zwar den Vorteil einer besseren Vergleichbarkeit im Rahmen von Meta-Analysen; es wäre jedoch wünschenswert, dass die pflegewissenschaftliche Community im Rahmen wissenschaftlicher Diskurse eigene Konventionen entwickelt.

### 7.3 ICC-Schätzer

#### Varianzanalyseschätzer

Der Varianzanalyse-Schätzer  $\hat{\rho}_{AOV}$  erweist sich als stabiler Schätzer, der die ICC in den Untersuchungen weder über- noch unterschätzt.



## Pearson-Schätzer

Der Pearson-Schätzer  $\hat{\rho}_{PEQ}$ , der allen Paaren das gleiche Gewicht gibt, schätzt die ICC bei Clustern gleicher Größe regelmäßig zu niedrig. Lediglich für die gesamte Stichprobe mit unterschiedlichen Clustergrößen wurde die ICC stabil geschätzt. Er ist somit eher für variierende Clustergrößen geeignet.

Der Schätzer  $\hat{\rho}_{PGP}$ , der jedem Cluster das gleiche Gewicht gibt, unterschätzt die ICC bei größeren Clustern und ist daher eher für die Berechnung der ICC bei sehr kleinen Clustern geeignet.

Als stabil sowohl bei variierenden als auch bei allen homogenen Clustergrößen hat sich der Schätzer  $\hat{\rho}_{PPR}$  gezeigt, der größeren Clustern ein kleineres Gewicht gibt.

## Moment-Schätzer

Die Untersuchung belegt, dass die ICC im diesem Kontext der Clusterbildung Werte annimmt, die weit unter  $\rho = 0,8$  liegen. Kleinmann (1973) empfiehlt für diese Situation die modifizierten Schätzer  $\hat{\rho}_{KEQ}^*$  und  $\hat{\rho}_{KPR}^*$ , die im Moment zweiter Ordnung die Clustergröße berücksichtigen. Die beiden Schätzer zeigen sich auch in dieser Untersuchung stabil, während die nicht modifizierten Moment-Schätzer  $\hat{\rho}_{KEQ}$  und  $\hat{\rho}_{KPR}$  die hier vorliegenden kleinen ICC regelmäßig überschätzen.

## probabilistische Schätzer

Während  $\hat{\rho}_{FC}$  unabhängig von der Clustergröße stabile ICC liefert, überschätzt  $\hat{\rho}_{MAK}$  die ICC in der Regel. Lediglich bei den Wohnbereichen mit weniger als 20 Bewohner/innen sowie bei der Berechnung über alle Wohnbereiche wird die ICC unterschätzt.  $\hat{\rho}_{MAK}$  kann somit für die Anwendung in diesem Kontext nicht empfohlen werden.

## 7.4 ICC-Werte

Die Ergebnisse zeigen, dass die ICC Werte annehmen kann, die in der Fallzahlschätzung bei Cluster-randomisierten Studien nicht unberücksichtigt bleiben können.

Es konnte gezeigt werden, dass die ICC unabhängig sowohl von der Wohnbereichsgröße als auch von der Erfolgsrate innerhalb der einzelnen Cluster ist.

Die Wohnbereichszugehörigkeit ist ein multifaktorielles Geschehen. Solche Faktoren können zum Beispiel sein: Träger der Einrichtung (privat, freie Wohlfahrtspflege), Personalausstattung – sowohl quantitativ als auch Qualifikationen, persönliche Eigenschaften der Pflegekräfte, Konzepte (z.B. der von Kirsch und Wassermann (2009) entwickelte Werdenfelser Weg) oder Hausstandards (z.B. ob jemand als dehydratationsgefährdet eingestuft wird, der eine Flasche nicht öffnen kann). Hinzu kommt, dass die Ergebnisse der Befragung auf der zwar fachlichen, aber dennoch subjektiven Einschätzung der Pflegefachkraft beruhen. Zu Beginn der Untersuchung wurde davon ausgegangen, dass alle diese Faktoren zusammen genommen letztlich den Einfluss der Wohnbereichszugehörigkeit auf die Ausprägung der untersuchten Variablen ergeben und dieser als ICC dargestellt werden kann.

Die indifferenten Ergebnisse insbesondere der Simulationsstudie zeigen jedoch, dass diese Unterschiede für Daten, die auf der subjektiven Einschätzung einer Pflegekraft beruhen, nicht standardisiert dargestellt werden können.

Daher kann eine Verallgemeinerung oder Übertragung der ICC-Werte dieser Studie auf andere Settings nicht empfohlen werden. Das Ziel, ICC-Werte für die Fallzahl-Schätzung künftiger Pflegestudien zur Verfügung zu stellen, konnte somit nicht erreicht werden.

Dennoch ergibt sich folgender Erkenntnisgewinn:

1. Die Wohnbereichs- oder Stationszugehörigkeit kann einen nicht zu vernachlässigenden Einfluss auf Bewohnermerkmale haben und darf daher bei der Fallzahlschätzung cluster-randomisierter Studien nicht unberücksichtigt bleiben.
2. Auf Veröffentlichungen der ICC aus anderen Studien sollte nur dann zurückgegriffen werden, wenn es sich um objektiv messbare Daten handelt. Da andernfalls nicht beurteilt werden kann, welche Faktoren letztlich welchen Effekt hervorrufen, ist eine Übertragung der Werte einer Untersuchung auf ein anderes Vorhaben kritisch zu betrachten.
3. Sofern zeitliche und finanzielle Ressourcen sowie das Forschungsdesign es zulassen, sollte zur Ermittlung der ICC eine eigene Vorstudie dort durchgeführt werden, wo auch die eigentliche Untersuchung stattfinden soll.

## 7.5 Empfehlungen

### Hypothesentesten als Methode

Die der Sozialwissenschaft entlehnte und in der Medizin häufig praktizierte Methode wird auch in der Pflegewissenschaft angewendet und ist dann geeignet, wenn der Wirksamkeitsnachweis einer Intervention erbracht werden soll, deren Erfolg an sogenannten harten Endpunkten wie beispielsweise Inzidenzen, Prävalenzen oder biometrischen Daten gemessen werden kann.

Dabei sind im Vorfeld das Signifikanz-Niveau  $\alpha$  sowie die Power  $1 - \beta$  als Bedingungen festzulegen, unter denen die Alternativhypothese  $H_0$  respektive die Forschungshypothese  $H_1$  zu verwerfen oder beizubehalten sind.

Hierbei kann auf die konventionellen Werte  $\alpha = 0,05$  oder  $0,1$  und  $\beta = 0,2$  zurückgegriffen werden. Dies ist vor allem dann zu empfehlen, wenn die Ergebnisse der Studie für die Veröffentlichung einem Peer-Review-Verfahren unterzogen werden sollen. Vor allem bei kleinen nachzuweisenden Effekten sind hier jedoch große Stichproben erforderlich, um das festgelegte Signifikanzniveau bei entsprechender Power zu erreichen.

Diese Werte sind nicht zwingend vorgeschrieben. Pflegeforschende sollten vielmehr den Mut aufbringen, Signifikanzniveau und Power am Forschungsgegenstand auszurichten bzw. im Rahmen einer Kompromiss-Analyse den Ressourcen und dem Forschungsdesign Rechnung zu tragen und entsprechend zu begründen.

Führen die Untersuchungsergebnisse zu dem Schluss, die statistische Hypothese  $H_1$  beizubehalten, so ist dies nicht gleichzusetzen mit einem "Beweis" der Forschungshypothese. Stattdessen wird das Resultat als Empfehlung für das weitere Handeln betrachtet; in diesem Fall wird so gehandelt, als ob die statistische Hypothese zuträfe - immer in dem Bewusstsein, dass weitere Studien zu einem anderen Ergebnis kommen könnten.

### Berücksichtigung der ICC bei der Bestimmung des Stichprobenumfangs

Die in dieser Studie sowie auch die von Kron (2003), Meyer et al. (2009) und Köpke et al. (2012) ermittelten ICC nehmen Werte an, die bei der Fallzahlschätzung nicht unberücksichtigt bleiben können, wenn kleine Effekte bei hohem Signifikanzniveau und unter Berücksichtigung der Power nachgewiesen werden sollen.

Die Verwendung der Werte dieser Untersuchung kann nur dann empfohlen werden, wenn die Datenerhebung wie hier auf Grundlage einer Einschätzung durch die Pflegefachkraft erfolgt.

Abhängig von der Anzahl  $k$  der Cluster, die in die Studie einbezogen werden sollen, kann dabei für große  $k$  auf die Berechnungen über die Wohnbereiche zurückgegriffen werden, für kleine  $k$  auf die Werte der Simulationsstudie.

Die sicherste Methode zur Schätzung der ICC bleibt letztlich die Durchführung einer Vorstudie, die jedoch aufgrund des hohen Einflusses durch das Pflegepersonal nur dann zuverlässige Werte liefern kann, wenn sie unter denselben Bedingungen - einschließlich einschätzende Pflegekräfte - durchgeführt wird, unter denen die eigentliche Untersuchung stattfindet. Dies ist methodisch nicht immer möglich.

### **Wahl des Schätzers**

Diese Arbeit war darauf ausgelegt, Pflegeforschenden mit geringen zeitlichen und finanziellen Ressourcen ein einfaches Verfahren zur Ermittlung des benötigten Stichprobenumfangs bei kleinen clusterrandomisierten Studien zur Verfügung zu stellen.

Die Analyse zeigt, dass die Werte der hier untersuchten Schätzer in der Regel nah beieinander liegen.

Als einer der stabilen Schätzer basiert der Varianzanalyse-Schätzer  $\rho_{AOV}$  auf einem Verfahren, das auch aus anderen Kontexten bekannt und einfach durchzuführen ist. Dieser Schätzer kann daher für die Bestimmung der ICC empfohlen werden.

Bei Verwendung eines Pearson-Schätzers kann der Schätzer  $\hat{\rho}_{PPR}$  für balancierte und unbalancierte Designs empfohlen werden, während  $\hat{\rho}_{PEQ}$  nur für variierenden Clustergrößen eingesetzt werden sollte.  $\hat{\rho}_{PGP}$  hat in dieser Untersuchung die ICC nur bei den sehr kleinen Clustern mit  $m < 20$  nicht unterschätzt und wird daher nicht empfohlen.

Soll ein Moment-Schätzer eingesetzt werden, sollten die modifizierten Schätzer  $\hat{\rho}_{KEQ}^*$  für homogene oder  $\hat{\rho}_{KPR}^*$  für variierende Clustergrößen verwendet werden, da die erwarteten Werte für die ICC weit unter  $\rho = 0,8$  liegen und die nicht modifizierten Schätzer  $\hat{\rho}_{KEQ}$  und  $\hat{\rho}_{KPR}$  die ICC dann regelmäßig überschätzen.

Von den probabilistischen Schätzern kann lediglich  $\hat{\rho}_{FC}$  empfohlen werden, der in allen Settings stabile ICC liefert. Da  $\hat{\rho}_{MAK}$  die ICC sowohl über- als auch unterschätzt, sollte von der Verwendung abgesehen werden.

## 7.6 Ausblick

In dieser Arbeit werden die ICC für die Inzidenz verschiedener Erkrankungen sowie für Zustände oder Verhaltensweisen von Bewohner/innen in Altenheimen vorgestellt. Dabei zeigt sich, welche unterschiedlichen Ausprägungen dieser Wert annehmen kann und welche Relevanz er bei der Fallzahlschätzung haben kann.

Dass die Ergebnisse dieser Untersuchung indifferent sind und zeigen, dass die in einer Studie ermittelten ICC nicht ohne weiteres auf eine andere Studie übertragen werden können, macht den weiteren Forschungsbedarf in dieser Thematik deutlich.

Forschende sollten in ihren Veröffentlichungen von cluster-randomisierten Studien die Empfehlungen der Erweiterung für clusterrandomisierte Studien (Campbell et al. 2012) der Leitlinie für Berichte randomisierter Studien "CONSORT 2010 statement" (Schulz et al. 2010) umsetzen und neben der Berechnung der Stichprobengröße auch die Anzahl der Cluster, ihre Größe und die berücksichtigte ICC berichten.

In der Folge bedeutet dies, dass auch die aus den eigenen Daten ermittelten ICC zu veröffentlichen sind, so dass sie für Meta-Analysen zur Verfügung stehen und gegebenenfalls als Ausgangswert für weitere Studien genutzt werden können.

In dieser Untersuchung zeigt sich, dass die ICC größtenteils sehr kleine Werte annimmt, was bedeutet, dass die Wohnbereichszugehörigkeit nur einen geringen Anteil an der Ausprägung der untersuchten Variablen hat. Dies führt zur Frage, inwieweit es sinnvoll ist, beispielsweise die Dekubitus- oder die Dehydratations-Inzidenz als Kriterien zur Bemessung von Pflegequalität heran zu ziehen, wenn dabei gute von schlechten Einrichtungen abgegrenzt werden sollen. Vielmehr sollte untersucht werden, welche Kriterien die Lebensqualität der Bewohner/innen abbilden könnten und inwiefern diese sich in ihren Ausprägungen in Abhängigkeit von der Einrichtungs- oder Wohnbereichszugehörigkeit unterscheiden.

## 8 Zusammenfassung

Pflegeforschung ist ein weites Feld, in dem sowohl qualitative als auch quantitative Methoden ihre Anwendung finden, je nach Art des Forschungsgegenstandes.

Nach § 113a SGB XI dienen Expertenstandards Sicherheit und Weiterentwicklung der Qualität in der Pflege und werden auch als Instrument zur Überprüfung der Pflegequalität eingesetzt. Sie sind somit handlungsleitend. Die darin enthaltenen Empfehlungen werden

überwiegend auf der Grundlage quantitativer Studien erstellt, für deren Bewertung die konventionellen Werte für das Signifikanzniveau  $\alpha$  und die Power  $1 - \beta$  herangezogen werden. Häufig lassen die Studienergebnisse keine eindeutigen Empfehlungen zu, unter anderem weil die Stichprobe zu klein war, um die Ergebnisse auf die Gesamtheit übertragen zu können.

Im Theorie-Teil dieser Arbeit werden die Methode des Hypothesentestens in ihrer Entwicklung sowie die Bedeutung von Signifikanzniveau und Power beschrieben. Die den Untersuchungen meist zugrunde liegenden konventionellen Werte für Signifikanzniveau und Power werden kritisch hinterfragt. Es wird empfohlen, sich für pflegewissenschaftliche Studien von den Konventionen der Disziplin Medizin zu lösen und im wissenschaftlichen Diskurs der dem Forschungsgegenstand angemessene eigene Konventionen zu entwickeln.

In Interventionsstudien werden im Feld der Pflegeforschung häufig Cluster (Stationen, Wohnbereiche) anstelle von Individuen randomisiert. Dies hat zur Folge, dass die untersuchten Variablen nicht nur von der Intervention sondern auch von der Cluster-Zugehörigkeit beeinflusst werden. Der daraus resultierende Power-Verlust der Studie wird durch eine Erhöhung des Stichprobenumfangs um den Design-Effekt ausgeglichen, welcher auf Grundlage der Intraclass-Correlation ICC ermittelt wird. Dieser Wert ist jedoch vor der Untersuchung nicht bekannt.

Die ICC wird nur selten im Rahmen der Veröffentlichung einer Studie dokumentiert. Die in dieser Arbeit in der Literaturrecherche gefundenen Werte beziehen sich auf Sturzinzidenz und Sturzfolgen sowie auf freiheitsentziehende Maßnahmen und werden in Kapitel ?? dargestellt.

In dieser Untersuchung wurden im Rahmen einer Sekundäranalyse der am Lehrstuhl für Statistik und standardisierte Methoden für Pflegeforschung der Philosophisch-Theologischen Hochschule Vallendar durchgeführten PiSaar-Studie erhobenen Daten die ICC für verschiedene personenbezogene Variablen ermittelt. Die Daten lagen binär-codiert vor. Auf Grundlage des Common-Correlation-Modells wurde die ICC mit unterschiedlichen Schätzern bestimmt und verglichen. Die Schätzer wurden sowohl auf die Grundgesamtheit sowie auf homogene Clustergrößen angewendet. In einem weiteren Schritt wurde untersucht, inwiefern die Eignung der Schätzer von der Erfolgsrate der untersuchten Variablen abhängt.

In Form einer Simulationsstudie wurden die ICC mit den untersuchten Schätzern über alle möglichen Kombinationen aus  $k = 2$  und  $k = 3$  Clustern sowohl über alle Wohnbereiche als auch nach gruppierten Wohnbereichsgrößen berechnet.

Die Ergebnisse zeigen zum einen, dass die ICC Werte annehmen kann, die in der Fallzahlsschätzung nicht unberücksichtigt bleiben dürfen. Zum anderen wird deutlich, dass die ICC insbesondere dort hoch ausfallen, wo die Einschätzung des Bewohnerzustandes nicht auf objektiven Daten sondern auf fachlicher Expertise der Pflegefachkräfte beruht. Dies kann darauf zurück geführt werden, dass pflegerische Phänomene abhängig vom zugrunde liegenden Begründungsrahmen unterschiedlich bewertet und argumentiert werden können. Somit sind die ICC-Werte für solche Variablen nur sehr begrenzt aus anderen Studien übertragbar.

Als einfaches und stabiles Verfahren für die Ermittlung der ICC zum Beispiel im Rahmen einer Vorstudie hat sich der Varianzanalyse-Schätzer  $\hat{\rho}_{AOV}$  erwiesen.

## Literaturverzeichnis

- Pflege ohne Freiheitsentzug ist machbar (2012). In: Die Schwester/Der Pfleger 51 (8).
- Adams, Geoffrey; Gulliford, Martin C.; Ukoumunne, Obioha C.; Eldridge, Sandra; Chinn, Susan; Campbell, Michael J. (2004): Patterns of intra-cluster correlation from primary care research to inform study design and analysis. In: J Clin Epidemiol 57 (8), S. 785–794.
- Bartholomeyczik, Sabine; Halek, Margareta; Riesner, Christine (2006): Rahmenempfehlungen zum Umgang mit herausforderndem Verhalten bei Menschen mit Demenz in der stationären Altenhilfe. Hg. v. Bundesministerium für Gesundheit. Berlin.
- Basler, Heinz-Dieter; Hüger, D.; Kunz, Roland; Luckmann, Judith; Lukas, Albert; Nikolaus, T.; Schuler, M. (2006). Beurteilung von Schmerz bei Demenz (BESD): Untersuchung zur Validität eines Verfahrens zur Beobachtung des Schmerzverhaltens. Der Schmerz 20 (6), S. 519-526.
- Bensch, Sandra (2009): Handbuch Subskalen "Mobilität" und "Kognitive und kommunikative Fähigkeiten" des neuen Begutachtungsassessments.
- Bredenkamp, J. (1969): Über die Anwendung von Signifikanztests bei theorie-testenden Experimenten. In: Psychologische Beiträge 11, S. 275–285.
- Brühl, Albert; Planer, Katharina (2013): PiSaar – Pflegebedarf im Saarland. Abschlussbericht. Philosophisch-Theologische Hochschule Vallendar. Online verfügbar unter [http://opus.bsz-bw.de/kidoks/volltexte/2013/117/pdf/PiSaar'\\_Abschlussbericht'\\_2013.pdf](http://opus.bsz-bw.de/kidoks/volltexte/2013/117/pdf/PiSaar'_Abschlussbericht'_2013.pdf), zuletzt aktualisiert am 2013, zuletzt geprüft am 19.08.2015.
- Campbell, M. K.; Piaggio, G.; Elbourne, D. R.; Altman, D. G. (2012): Consort 2010 statement: extension to cluster randomised trials. In: BMJ 345 (sep04 1), e5661.
- Carnap, Rudolf; Stegmüller, Wolfgang (1959) Induktive Logik und Wahrscheinlichkeit. Wien: Springer.
- Cohen, Jacob (1988): Statistical power analysis for the behavioral sciences. 2. Aufl. Hillsdale N.J: Lawrence Erlbaum Associates.
- Donner, A. (1986): A Review of Inference Procedures for the Intraclass Correlation Coefficient in the One-Way Random Effects Model. In: Int Stat Rev (International Statistical Review) 54 (1), S. 67–82.
- Donner, Allan; Klar, Neil (2000): Design and analysis of cluster randomization trials in health research. Chichester [England]: John Wiley & Sons.



- Erdfelder, E.; Buchner, A.; Faul, F.; Brandt, M. (2004): GPOWER: Teststärkeanalysen leicht gemacht. In: Edgar Erdfelder und Joachim Funke (Hg.): Allgemeine Psychologie und deduktivistische Methodologie. Göttingen: Vandenhoeck & Ruprecht, S. 148–166.
- Erdfelder, E.; Faul, F.; Buchner, A. (2005): Power analysis for categorical methods. In: Brian Everitt und David C. Howell (Hg.): Encyclopedia of statistics in behavioral science, Bd. 3. 4 Bände. Hoboken, N.J: John Wiley & Sons, S. 1565–1570.
- Faul, Franz; Erdfelder Edgar; Lang Albert-Georg; Buchner Axel (2007): G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. In: Behavior Research Methods 39 (2), S. 175–191.
- Fisher, Ronald Aylmer (1951): The Design of experiments. Edinburgh [u.a.]: Oliver & Boyd.
- Fleiss, Joseph L. (1971): Measuring nominal scale agreement among many raters. In: Psychological Bulletin 76 (5), S. 378–382.
- Fleiss, Joseph L. (1981): Statistical methods for rates and proportions. 2. Aufl. New York: Wiley (Wiley series in probability and mathematical statistics).
- Fleiss, Joseph L.; Cuzick, Jack (1979): The Reliability of Dichotomous Judgments: Unequal Numbers of Judges per Subject. In: Applied Psychological Measurement 3 (4), S. 537–542.
- Gnass, Irmela (2010): Erworbene Kontrakturen der Gelenke im höheren Lebensalter. eine systematische Literaturanalyse. In: Zeitschrift für Gerontologie und Geriatrie 43 (3), S. 147–157.
- Hager, Willi (2004): Testplanung zur statistischen Prüfung psychologischer Hypothesen. Die Ableitung von Vorhersagen und die Kontrolle der Determinanten des statistischen Tests. Göttingen, Seattle: Hogrefe.
- Juchli, Liliane (2000): Thiemes Pflege. Entdecken - erleben - verstehen - professionell handeln. 8. Aufl. u.d.T.: Juchli, Liliane: Pflege. 9. Aufl. Hg. v. Edith Kellnhauser. Stuttgart, New York: Thieme.
- Karlin, S.; Cameron, E. C.; Williams, P. T. (1981): Sibling and parent–offspring correlation estimation with variable family size. In: Proc. Natl. Acad. Sci. U.S.A. 78 (5), S. 2664–2668.
- Kirsch, Sebastian; Wassermann, Josef (2009): 'Der Werdenfelser Weg' - Eine Initiative zur Reduzierung von Fixierungsmaßnahmen mit verfahrensrechtlichem Ansatz. In: Bt-Prax (3), S. 109-112.
- Kleinman, Joel C. (1973): Proportions with Extraneous Variance: Single and Independent Samples. In: Journal of the American Statistical Association 68 (341), S. 46–54.

Köpke, Sascha; Mühlhauser, Ingrid; Gerlach, Anja; Haut, Antonie; Haastert, Burkhard; Möhler, Ralph; Meyer, Gabriele (2012): Effect of a Guideline-Based Multicomponent Intervention on Use of Physical Restraints in Nursing Homes. In: JAMA 307 (20).

Kottner, Jan; Tannen, Antje (2015): Anlage zur Literaturstudie zum Expertenstandard „Dekubitusprophylaxe in der Pflege“, 1. Aktualisierung. Hg. v. Deutsches Netzwerk für Qualitätsentwicklung in der Pflege.

Krohwinkel, Monika (2013): Fördernde Prozesspflege mit integrierten ABEDLs. Forschung, Theorie und Praxis. Bern: Hans Huber, Hogrefe.

Kron, Martina (2003): Statistische Testverfahren zur Auswertung cluster-randomisierter Studien mit der Inzidenzrate als Zielgröße. Habilitationsschrift. Universität Ulm, Ulm. Abteilung Biometrie und Medizinische Dokumentation.

Kuß, Oliver; Jahn, Patrick; Renz, Petra; Landenberger, Margarete (2009): Cluster-randomisierte Studien in der Pflegewissenschaft. In: Hallesche Beiträge zu den Gesundheits- und Pflegewissenschaften 8 (1), S. 302–310.

Lauth, Bernhard; Sareiter, Jamel (2005): Wissenschaftliche Erkenntnis. Eine ideengeschichtliche Einführung in die Wissenschaftstheorie. 2. Aufl. Paderborn: Mentis.

Lüer, Gerd (Hg.) (1987): Allgemeine experimentelle Psychologie. Eine Einführung in die methodischen Grundlagen mit praktischen Übungen für das Experimentelle Praktikum. Stuttgart: Fischer.

Mak, Tak K. (1988): Analysing Intraclass Correlation for Dichotomous Variables. In: Applied Statistics 37 (3), S. 344.

Meyer, Gabriele; Köpke, Sascha; Haastert, Burkhard; Mühlhauser, Ingrid (2009): Restraint use among nursing home residents: cross-sectional study and prospective cohort study. In: Journal of Clinical Nursing 18 (7), S. 981–990.

Neyman, J. and Pearson E. S. (1933): On the Problem of the Most Efficient Tests of Statistical Hypotheses. In: Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character (231), S. 289–337.

Popper, Karl R.; Keuth, Herbert (2005): Logik der Forschung. 11. Aufl. Tübingen: Mohr Siebeck.

Ridout, M. S.; Demétrio, C. G.; Firth, D. (1999): Estimating intraclass correlation for binary data. In: Biometrics 55 (1), S. 137–148.

- Sackett, L. David; Richardson, W. Scott; Rosenberg, William; Haynes, R. Brian (1999): Evidenzbasierte Medizin. EBM-Umsetzung und -Vermittlung. Dt. Ausg.: Kunz, Regina; Lutz Fritsche. Bern: Zuckerschweidt.
- Schiemann, Doris (2006): Expertenstandard Sturzprophylaxe in der Pflege. Entwicklung, Konsentierung, Implementierung. Osnabrück: Fachhochschule Osnabrück (Schriftenreihe des Deutschen Netzwerks für Qualitätsentwicklung in der Pflege).
- Schiemann, Doris (2007): Expertenstandard Förderung der Harnkontinenz in der Pflege. Osnabrück: Fachhochschule Osnabrück (Schriftenreihe des Deutschen Netzwerks für Qualitätsentwicklung in der Pflege).
- Schiemann, Doris (Hg.) (2009): Expertenstandard Ernährungsmanagement zur Sicherstellung und Förderung der oralen Ernährung in der Pflege. Osnabrück: DNQP (Schriftenreihe des Deutschen Netzwerks für Qualitätsentwicklung in der Pflege).
- Schiemann, Doris (2010): Expertenstandard Dekubitusprophylaxe in der Pflege. Entwicklung - Konsentierung - Implementierung. 1. Aktualisierung 2010. Osnabrück: DNQP (Schriftenreihe des Deutschen Netzwerks für Qualitätsentwicklung in der Pflege).
- Schlömer, Gabriele (1999): RCTs und systematic reviews in der Pflegeliteratur. Ein Vergleich zwischen deutscher und internationaler Pflegeforschung. In: Pflege 12 (4), S. 250–258.
- Schulz, K. F.; Altman, D. G.; Moher, D. (2010): CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. In: BMJ 340 (mar23 1), c332.
- Sedlmeier, Peter; Renkewitz, Frank (2011): Forschungsmethoden und Statistik in der Psychologie. [Nachdr.]. München: Pearson Studium (PS Psychologie).
- Steurer, Johann (2009): Vor allem Antidepressiva, Neuroleptika, Benzodiazepine und Sedativa erhöhen das Sturzrisiko bei älteren Menschen. Horten-Zentrum für praxisorientierte Forschung und Wissenstransfer. Online verfügbar unter <http://www.evimed.ch/AGORA/HTZ000/downloads/sturzmedik.pdf>, zuletzt aktualisiert am 30.11.2009, zuletzt geprüft am 20.08.2015.
- Warden, V., Hurley, A.C. Volicer, L. (2003). Development and Psychometric Evaluation of the Pain Assessment in Advanced Dementia (PAINAD) Scale. In: JAMDA 4 (1), 9 –15.
- Westermann, Rainer (1987 a): Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie. Eine Fallstudie. Berlin, New York: Springer (Lehr- und Forschungstexte Psychologie, 25).

Westermann, Rainer (1987 b): Wissenschaftstheoretische Grundlagen der experimentellen Psychologie. In: Gerd Lüer (Hg.): Allgemeine experimentelle Psychologie. Eine Einführung in die methodischen Grundlagen mit praktischen Übungen für das Experimentelle Praktikum; Stuttgart: Fischer, S. 5–42.

Westermann, Rainer (2000): Wissenschaftstheorie und Experimentalmethodik : ein Lehrbuch zur psychologischen Methodenlehre. Göttingen [u.a.]: Hogrefe, Verl. für Psychologie.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt habe.

Ich versichere, dass die vorliegende Arbeit bisher nicht veröffentlicht wurde und nicht bereits einmal an einer anderen Hochschule als Dissertation eingereicht wurde.

Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium (CD-ROM).

Remagen, 22. Juli 2019