
Does Direct-care Staffing Impact Nursing Home Outcomes? A Systematic Review



Vanessa Schlage and Carl Rudolf Blankart

Rollenmodelle; Pflegemanagement; Langzeitpflege; Pflegeressourceneinsatz; Prozess- und Ergebnisqualität

staff-quality relationship; long-term care; nurse staffing; direct care workforce; nursing role model; nursing home management



Empirische Evidenz zum Zusammenhang von Pflegepersonaleinsatz und Qualität in US Pflegeheimen wurde systematisch aufbereitet. Identifizierte Studien wurden an Hand einer 21-Faktoren-Skala beurteilt und die Ergebnisse nach Qualifikation der Pflegekräfte stratifiziert. Die meiste Evidenz wurde zu examiniertem Pflegepersonal identifiziert. Ein Einfluss einer Erhöhung von weniger qualifiziertem Pflegepersonal auf die Qualität kann nicht belegt werden. Das Pflegeheimmanagement sollte diesen Zusammenhang bei der Substitution von besser zu schlechter ausgebildetem Pflegepersonal berücksichtigen. Zwar können durch die Substitution kurzfristige finanzielle Ziele erreicht werden, allerdings führt eine sinkende Qualität langfristig zu schlechteren Geschäftsperspektiven. Vor dem Hintergrund der Harmonisierungsbemühungen der EU, wären weitere Forschungsarbeiten mit europäischen Daten wünschenswert.

The empirical evidence on the relationship between nurse staffing levels and quality in US nursing homes was reviewed. Studies were assessed using a newly developed 21-item quality framework. The findings were systematically collected and stratified by nurse type. Most evidence identified was on highly educated registered nurses. According to our review, less qualified staff cannot be claimed to improve quality. Nursing home managers contemplating switching from highly to less educated staff to improve short-term financial performance should consider that this negatively impacts quality outcomes and therefore long-term business opportunities. Evidence on this relationship in the European environment would be desirable, especially in the context of currently discussed health workforce policy harmonization efforts or health workforce qualification initiatives.

1. Introduction

Nursing home (NH) managers face a trade-off between boosting profits and providing high-quality services in their management decisions. As the labor cost of direct-care staff, i.e., registered nurses (RNs), licensed practical nurses (LPNs) and nurses' aides (NAs), contributes to the largest cost categories in NHs, programs on efficiency often modify the number and type of staff to increase profit margins. Federal and state governments there-

fore try to regulate the number and qualifications of direct-care staff, as they worry that lowering staff ratios will result in poor outcomes (*Bowblis* 2011).

At first glance, the causal relationship between higher direct-care staffing levels and improved outcomes seems obvious. In line with this assumption, *Bostick et al.* (2006) claimed, based on their review of 87 government documents and scientific articles, that the association between higher total staffing levels and improved quality of care has been proven. However, reviews published a few years later showed that the scientific evidence on this relationship is ambiguous. *Castle* (2008) found that most studies support the assumption of an association between NH staffing levels and outcomes, but he also emphasized that only a few studies seem methodologically robust and that none of the studies analyzed represented a gold standard. Therefore, the author concluded that the evidence for this relationship was too scarce to draw further conclusions and that more research is needed (*Castle* 2008). During the same year, another group of authors confirmed that hospitalizations—an important NH outcome—are often preventable as well as related to facility practices and state Medicaid policies. However, they also could not provide definitive evidence for the assumed relationship regarding staffing levels, and they noted that most studies were characterized by potentially confounded research designs (*Grabowski et al.* 2008).

While NH managers have to find an answer to address the trade-off between efficiency and outcomes at the facility level, policy makers face a trade-off at the health care system level: although increased regulations may shift NH managers' focus toward better outcomes, regulation always requires a higher administrative effort, and therefore less efficiency at the system level. It is thus highly important to obtain a good understanding of the assumed causal relationship before enacting regulatory measures, such as minimum staffing ratios, wage pass-through legislation and minimum qualification requirements. Although this review focuses on NHs in the US, the insights into the direct-care staffing-quality relationship are also highly relevant from an international perspective, especially because empirical analyses that focus on other large heterogeneous nursing home markets, such as the diverse European markets, are unavailable due to a lack of harmonized data. However, those analyses are needed to advance large initiatives, such as the workforce policy harmonization efforts preceded by the EU Joint Action on Health Workforce Planning & Forecasting or the more recent health workforce qualification draft bill that has been discussed in Switzerland since 2015.

The purpose of this study is to (i) review the scientific literature on the relationship between nurse staffing and quality, (ii) assess the strengths and weaknesses of the studies published, and (iii) draw conclusions regarding this relationship considering the study quality.

2. Methods

2.1. Literature search

The aim of the analysis was to find original empirical research analyzing the potential causal relationships between direct-care staffing levels and outcomes in US NHs. Both free text and subject heading searches in the PubMed, Web of Science, and EBSCO databases were conducted to identify relevant original research articles. Databases were searched from their start dates through June 01, 2016. We additionally used ScienceDirect to search

the databases of the following journals that have a specific nursing focus: Geriatric Nursing, Journal of Nursing Regulations, International Journal of Nursing Studies, Applied Nursing Research, Journal of Health Economics, and International Journal of Nursing Sciences. Electronic searches were supplemented by manual screening of reference lists. Only full-text original research articles published in English were considered.

2.2. Study selection and data collection

We followed the PRISMA minimum set of items to ensure that the study selection was transparent and complete and to minimize selection bias (*Liberati et al. 2009*). The two reviewers focused on original empirical research that presumably evaluated the causal effects of staffing levels on outcomes in US NHs. We therefore excluded studies that did not focus on US NHs, studies that were solely descriptive or theoretical and non-peer-reviewed literature. We also did not include letters, editorials, commentaries, conference and poster abstracts, methodological papers, reviews, and case studies.

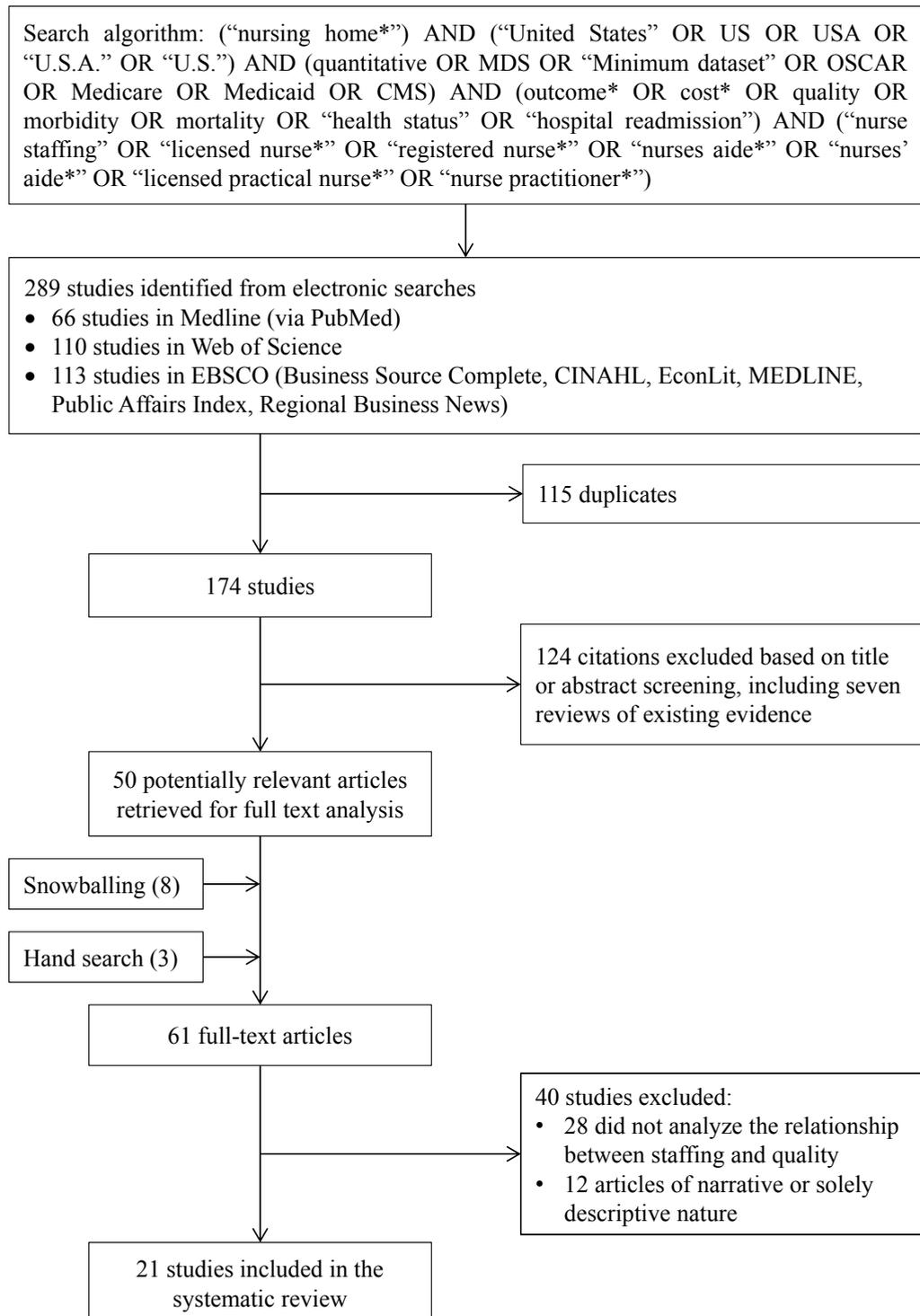
We first independently screened the electronic search results and excluded irrelevant publications based on title and abstract according to the defined criteria. We obtained the full texts of all potentially appropriate studies and assessed their eligibility. In addition, we manually searched the reference sections of eligible articles and previous systematic reviews for potentially relevant articles. The selection process is displayed in *Figure 1*.

Information about the study setting (e.g., state/region, number of observations), NH characteristics (e.g., sample size, type of NH), data analyzed (e.g., Online Survey Certification and Reporting (OSCAR) or state-level data), statistical approach (e.g., panel, cross-section), and main outcomes were extracted using a standardized form. In addition, we extracted the dependent and independent variables including their parameter estimates and the reported levels of significance.

2.3. Assessment of study quality

Quality assessments of the included studies are essential to ensuring a reasonable appraisal of the evidence. Until now, an appropriate instrument for quality assessment has not been available, especially because the methodological approaches, data sources, and outcomes are rather heterogeneous in type and nature across the studies included in reviews.

Figure 1: Selection of publications to be reviewed



We therefore designed a checklist based mainly on existing quality assessment tools, i.e., the British Medical Journal (BMJ) Check-list (Drummond/Jefferson 1996), Quality of Health Economic Studies (QHES) Instrument (Ofman et al. 2003) and Consensus Health Economic Criteria (CHEC) List (Evers et al. 2005; Gerkens et al. 2008). Our final quality assessment tool included 21 questions grouped in six categories (see Table 3): (i) study design, (ii) data, (iii) statistical analysis, (iv) dependent variables and risk adjustment, (v) interpretation of the results, and (vi) conflicts of interests. A combined global score of a minimum of zero and a maximum of 38 points was calculated by allocating points to each question: zero points for a question answered “no”, one point if answered “partially” and two points if answered “yes”. This approach enabled the calculation of a global score for each study, thus allowing for direct comparisons. A higher score can be interpreted to indicate a higher study quality. We considered a score greater than 30 points (> 80%) as “excellent”, greater than 24 (> 65% – 80%) as “good” and greater than 20 (> 50% – 65%) as “satisfactory”. Scores lower than or equal to 19 ($\leq 50\%$) were considered “unacceptable”. For the interpretative analysis, we stratified the results by staff type and plotted bar charts to visualize the relationship. To avoid an overrepresentation of studies reporting more than one model, we aggregated all findings and normalized the results to one.

3. Results

A total of 21 studies that empirically analyzed the relationship between staffing and NH outcomes were included in our review. Table 1 summarizes the general characteristics of the studies that were published between 1990 and 2014. The 40 studies that did not meet the inclusion criteria were off-topic (28 studies) or were solely descriptive/narrative in nature (12 studies) (see Figure 1).

3.1. Data and statistical analyses

The studies varied regarding data sources, which resulted in different sample sizes, quality indicators used, type of direct-care staff analyzed, confounding factors controlled for, and statistical models employed. In addition, the studies varied in NH characteristics, time covered, and states analyzed. Table 1 provides an overview of the data and variables used. In the US, 96% of all nursing homes are certified by the CMS and therefore required by law to report essential facility-level data (OSCAR). Most studies relied on this comprehensive database. Five studies used OSCAR as their only data source because the database also comprises detailed staffing-level data, resident characteristics and deficiencies. However, depending on their research question, researchers often supplemented OSCAR data with additional information from the Area Resource File (ARF), the Minimum Data Set (MDS), or other additional databases. While the ARF includes aggregate information on demographics, socioeconomic and health status at the county level, the MDS contains health status at the individual resident level.

The included studies analyzed the relationship between staffing levels and outcomes from as early as 1985 (Munroe 1990) to as late as 2007 (Wagner et al. 2013; King et al. 2013). The times analyzed in longitudinal studies varied from a few months (Intrator et al. 2004) up to seven years (Wagner et al. 2013). Eight studies focused on a single state, while four studies analyzed more than one state. The remaining nine studies did not explicitly include information about the states analyzed, but it can be assumed that they focused on

the national level. The dataset with the smallest number of NHs (90 NHs) was used by *Akinci/Krolikowski* (2005), while *Wagner et al.* (2013) reported the largest sample size, at 16,745 different NHs.

Table 1: Data sources and analytical approach

Study	Data source(s)	State(s)	Year(s)	Number/unit of observations	Type of data	Explanatory statistical model	Sensitivity analysis
<i>Munroe (1990)</i>	OSHPD, HCFA	CA	1985-1986	455 Medicare-certified SNFs	Cross-sectional	OLS regression	No
<i>Cohen/Spector (1996)</i>	Institutional population component of the NMEs, ARF	National	1987	658 Medicaid-certified NHs; 2,633 residents	Cross-sectional	2SLS regression; logistic regression	No
<i>Bliesmer et al. (1998)</i>	Not reported	MIN	1988-1990	Not reported	Cross-sectional	OLS regression; Probit regression	No
<i>Harrington et al. (2000)</i>	OSCAR	National	1995-1996	13,770 Medicaid- and/or Medicare-certified NHs (excluding ICFs)	Cross-sectional	OLS regression; logistic regression	Yes, consistent results
<i>Castle (2000)</i>	OSCAR	National	1992/1997	12,193 Medicaid- and/or Medicare-certified NHs	Longitudinal	Multinomial logistic regression	Yes, consistent results
<i>Castle (2002)</i>	OSCAR	National	1996-1999	14,042 Medicaid and/or Medicare-certified NHs	Longitudinal	Multinomial logistic regression	No
<i>Carter/Porell (2003)</i>	State quarterly Medicaid case-mix reimbursement data, Medicare inpatient hospital claims, MMQ data, MEDPAR, MRSC	MA	1991-1994	527 Medicaid- and/or Medicare-certified NHs and ICFs; 72,319 person-quarter observations	Longitudinal	Multinomial logistic regression	Yes, consistent results
<i>Intrator et al. (2004)</i>	MDS, OSCAR, ARF	ME, KS, NY, SD, (52 urban counties)	1997	663 freestanding urban NHs; 631 long-stay residents	Cross-sectional	Multilevel multinomial; logistic regression	No

Study	Data source(s)	State(s)	Year(s)	Number/unit of observations	Type of data	Explanatory statistical model	Sensitivity analysis
Zhang, X./Grabowski (2004)	1987 MIMACS, 1993 OSCAR, ARF	AR, AZ, CA, DE, ID, LA, MD, MN, NJ, NM, NV, NY, OH, OK, OR, PA, TN, TX, VA, WA, WI, WV	1987-1993	5,092 Medicaid- and/or Medicare-certified NHs	Longitudinal	First-difference FE regression	Yes
Akincil Krolikowski (2005)	OSCAR, NH standard output reports and NH performance profiles database from the Pennsylvania Department of Health	Northeast PA	03/2000-09/2002	90 licensed NHs	Cross-sectional	OLS regression	No
Zhang, N. J. et al. (2006)	OSCAR	National	2002-2003	13,663 Medicaid- and/or Medicare-certified NHs	Cross-sectional	OLS regression; logistic regression	No
Dellefield (2006)	OSCAR, OSHPD	CA	1996	897 Medicaid- and/or Medicare-certified NHs	Cross-sectional	OLS regression; logistic regression	No
Decker (2006)	1999 NNHS	National	1999	6,386 Medicaid-, Medicare-certified or state licensed NHs	Cross-sectional	Multinomial logistic regression	No
Decker (2008)	1999 NNHS, ARF, NNHS universal file	National	1999	Medicaid- or Medicare-certified or state licensed NHs; 6,623 discharges	Cross-sectional	Multinomial logistic regression	No
Koneitzka et al. (2008)	MDS, OSCAR	OH, KS, ME, MS, SD	1997-2000	1,366 freestanding Medicaid- and/or Medicare-certified NHs; 399,206 residents	Longitudinal	Conditional logit regression with FE (with and without IV)	Yes

Study	Data source(s)	State(s)	Year(s)	Number/unit of observations	Type of data	Explanatory statistical model	Sensitivity analysis
<i>Kim et al. (2009a)</i>	California's long-term care annual cost report, ACLAIMS	CA	1999-2003	Two groups: 201/210 freestanding Medicaid- and Medicare NHs; 850/ 910 annual observations	Longitudinal	Poisson regression with RE	No
<i>Kim et al. (2009b)</i>	ACLAIMS, OSCAR, MDS, U.S. BEA, cost report of California, OSHPD	CA	1999-2003	1,099 freestanding Medicaid- and Medicare- certified NHs; 4,933 observations	Longitudinal	Poisson regression with RE; Probit regression with RE	No
<i>Xing et al. (2013)</i>	CCW, MDS, ARF, provider of service file and Brown University's long-term care facts website (http://ltcfocus.org)	National	2007	11,999 Medicaid- and/or Medicare-certified NHs; 319,270 NH stays; 274,774 residents	Cross-sectional	Poisson regression with RE; logistic regression model with FE	No
<i>Wagner et al. (2013)</i>	OSCAR	National	2000-2007	16,745 Medicaid- and/or Medicare-certified NHs; ~120,000 observations	Longitudinal	Generalized estimating equations with a logit link	No
<i>Lin (2014)</i>	OSCAR, 2004 ARF, CMS, population census	AR, CA, DE, FL, IA, ME, MS, OH	1996-2003	3,275 Medicaid- and/or Medicare-certified NHs (excluding SNFs)	Longitudinal	OLS and IV regression of the first-difference with FE	Yes, consistent results
<i>Lee et al. (2014)</i>	OSCAR, MDS version 2.0, quarterly staffing data from the Colorado State inspections, state Medicaid reimbursement data, ARF	CO	2000	195 freestanding NHs (excluding hospital-based NHs, veteran's homes and specialized care facilities)	Cross-sectional	2SLS regression with IV; OLS regression	No

ACLAIMS: Automated Certification and Licensing Administrative Information and Management System; ARF: area resource file; BEA: Bureau of Economic Analysis; CCW: Chronic Care Data Warehouse; CMS: Centers for Medicare and Medicaid Services; FE: fixed effects; HCFA: Health Care Financing Administration; ICF: intermediate care facilities for individuals with developmental disabilities; IV: instrumental variable; MDS: minimum data set; MEDPAR: Medicare Provider Analysis and Review File; MIMACS: Medicare/Medicaid Automated Certification System; MIMQ: Management Minutes Questionnaire; MRSC: Massachusetts Rate Setting Commission; NH: nursing home; NMES: National Medical Expenditure Survey; NNHS: national nursing home survey; OSCAR: Online Survey Certification and Reporting; OLS: ordinary least square; OSHPD: Office of Statewide Health Planning and Development; RE: random effects; SNF: skilled nursing facility; 2SLS: two-stage least squares;

Table 2: Association and direction of relationship between staffing levels and outcomes

Study	Quality indicators	Registered nurses (RNs)		Licensed practical nurses (LPNs)		Nurses' aides (NAs)	
		Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)
<i>Murowe (1990)</i>	Number of health-related deficiencies (C)	RN/LVN ratio HPRD	+ (0.1)				
<i>Cohen/Spector (1996)</i>	Probability of death (O)	RN/100 residents	+ (0.05)	LPNs/100 residents	n.s.	NAs/100 residents	n.s.
	Probability of bedsores (O)		n.s.		n.s.		n.s.
	1: TDS, 1 year following admission (including deaths and discharges) (O) a: Cohort 1988 b: Cohort 1989 c: Cohort 1990					Non-licensed nursing hours (NA)	1a: + (0.001) 1b: + (0.001) 1c: + (0.001)
<i>Bliesmer et al. (1998)</i>	2: TDS, 1 year following admission (excluding deaths and discharges) (O) a: Cohort 1988 b: Cohort 1989 c: Cohort 1990						2a: n.s. 2b: n.s. 2c: n.s.
	3: Predictors of TDS in 1991 based on 1988 admissions (including deaths and discharges) (O)						3: + (0.001)
	4: Predictors of TDS in 1991 based on 1988 admissions (excluding deaths and discharges) (O)						4: n.s.
<i>Harrington et al. (2000)</i>	Total care deficiencies (C)	FTE RN HPRD	+ (0.05)	FTE LVN/LPN HPRD	n.s.	FTE NA HPRD	+ (0.01)
	Quality of care deficiencies (C)		+ (0.01)		n.s.		+ (0.05)
	Quality of life deficiencies (C)		n.s.		n.s.		+ (0.01)
	Other deficiencies (C)		n.s.		n.s.		n.s.
<i>Castle (2000)</i>	Use of physical restraint (P)	FTE RN/100 beds Increase in FTE RN/100 beds	+ (0.01) n.s.	FTE LPN/100 beds Increase in FTE LPN/100 beds	n.s. + (0.05)	FTE NA/100 beds Increase in FTE NA/100 beds	- (0.001) - (0.01)

Study	Quality indicators	Registered nurses (RNs)		Licensed practical nurses (LPNs)		Nurses' aides (NAs)	
		Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)
Castle (2002)	1: Deficiency citation for physical restraint use (P)	FTE RN/100 beds	+ (0.05)	FTE LPN/100 beds	n.s.	FTE NA/100 beds	+ (0.01)
	2: Deficiency citation for physical restraint use (P)		+ (0.01)		n.s.		+ (0.05)
	3: Consecutive deficiency citations for physical restraint use (P)		+ (0.001)		+ (0.01)		- (0.01)
Cartier/Porell (2003)	Hospitalization (coded binary) (O)	RN staffing ex-penses (proportion of total annual nursing ex-penses)	n.s.	LPN staffing ex-penses (proportion of total annual staffing ex-penses)	- (0.05)		
		RN/(RN+LPN)	+ (0.05)			CNA/(RN+LPN)	- (0.01)
Intrator et al. (2004)	Hospitalization with ACS diagnosis (O)		+ (0.05)				- (0.05)
	Hospitalization without ACS diagnosis (O)		+ (0.05)				- (0.05)
Zhang, X./Grabowski (2004)	1: All NHs:	FTE RN HPRD	1a: - (0.001)	FTE LPN HPRD	1a: - (0.05)	FTE NA HPRD	1a: + (0.05)
	a: DUs (%) (O)		1b: - (0.05)		1b: n.s.		1b: + (0.05)
	b: Physical restraints (%) (P)		1c: - (0.01)		1c: - (0.01)		1c: n.s.
	c: Catheter use (%) (P)		2a: n.s./ (0.05)		2a: n.s./ n.s.		2a: n.s./n.s.
	2: NH with lowest/highest staffing quartile of RNs		2b: + (0.01)/ n.s.		2b: n.s./n.s.		2b: n.s./ + (0.001)
			2c: n.s./n.s.		2c: n.s./ (0.05)		2c: n.s./n.s.
3: Nursing homes with lowest/highest staffing quartile of LPNs			3a: + (0.01)/ - (0.05)		3a: n.s./ (0.05)		3a: n.s./ + (0.01)
			3b: n.s./n.s.		3b: n.s./n.s.		3b: n.s./ + (0.01)
			3c: n.s./n.s.		3c: - (0.05)/ n.s.		3c: n.s./n.s.

Study	Quality indicators	Registered nurses (RNs)		Licensed practical nurses (LPNs)		Nurses' aides (NAs)	
		Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)
	4: <i>Nursing homes with lowest/highest staffing quartile of NAs</i>	4a: n.s./- (0.05) 4b: n.s./n.s. 4c: n.s./n.s.	4a: n.s./n.s. 4b: n.s./n.s. 4c: - (0.01)/n.s.	4a: n.s./n.s. 4b: n.s./n.s. 4c: n.s./n.s.	4a: n.s./n.s. 4b: + (0.001)/+ (0.05) 4c: n.s./n.s.		
Akincil Krolukowski (2005)	Total number of deficiencies (C)	RN HPRD	+ (0.05)	LPN/LVN HPRD	n.s.	CNA HPRD	+ (0.05)
Zhang, N. J. et al. (2006)	Quality of resident care (C)	FTE RN HPRD	+ (0.01)	FTE LPN HPRD	- (0.01)	FTE NA HPRD	n.s.
Dellefield (2006)	1: OLS: risk-adjusted DU prevalence (O) 2: Logistic regression: deficiency citation (O)	RN/LVN HPRD	1: + 2: n.s.	LVN/CNA HPRD	1: n.s. 2: n.s.	CNA wages	1: n.s.
Decker (2006)	1: All observations: a: Death (O) b: Hospitalization (O) 2: Length of stay < 60 days: 3: Length of stay ≥ 60 days:	FTE RN-hours per bed per day	1a: + (0.01) 1b: + (0.01) 2a: + (0.01) 2b: + (0.01) 3a: n.s. 3b: + (0.05)	FTE LPN-hours per bed per day	1a: + (0.01) 1b: n.s. 2a: n.s. 2b: n.s. 3a: n.s. 3b: n.s.	FTE NA-hours per bed per day	1a: n.s. 1b: n.s. 2a: n.s. 2b: n.s. 3a: n.s. 3b: n.s.
Decker (2008)	Probability of hospitalization: (O) 1: Residents admitted from hospital a: Short-stay residents b: Long-stay residents	FTE RN daily hours per bed	1a: + (0.05) 1b: + (0.05) 2a: + (0.05) 2b: n.s.	FTE LPN daily hours per bed	1a: n.s. 1b: n.s. 2a: n.s. 2b: n.s.	FTE NA daily hours per bed	1a: n.s. 1b: n.s. 2a: + (0.05) 2b: n.s.

Study	Quality indicators	Registered nurses (RNs)		Licensed practical nurses (LPNs)		Nurses' aides (NAs)	
		Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)
Konetzka et al. (2008)	2: Residents not admitted from hospital	Ratio RN to total nursing staff	1a: + (0.05) 1b: + (0.01) 2a: n.s. 2b: n.s.	FTE LPN daily hours per bed (alternative specification)	1a: + (0.01) 1b: + (0.01) 2a: n.s. 2b: n.s.	FTE NA daily hours per bed (alternative specification)	1a: n.s. 1b: n.s. 2a: + (0.05) 2b: n.s.
	a: Short-stay residents b: Long-stay residents						
	1: Facility fixed effects model: a: DUs (O) b: UTIs (O)	RN HPRD	1a: + (0.01) 1b: - (0.01) 2a: + (0.01) 2b: + (0.01) 3a: + (0.01) 3b: + (0.01)				
Kim et al. (2009a)	2: Facility fixed effects with 2SLS	Skill mix (% of total hours provided by RNs)	1a: n.s. 1b: n.s. 2a: n.s. 2b: + (0.01) 3a: n.s. 3b: + (0.01)				
	3: Facility fixed effects with 2SLS residual inclusion						
	1: Nursing homes meeting the state standard: a: Total deficiencies (C) b: Serious deficiencies (C)	RN to total staff ratio	1a: n.s. 1b: + (0.05) 2a: + (0.001) 2b: n.s.				

Study	Quality indicators	Registered nurses (RNs)		Licensed practical nurses (LPNs)		Nurses' aides (NAs)	
		Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)
	2: <i>Nursing homes not meeting the state standard</i>	RN/LVN	1a: + (0.05) 1b: + (0.05) 2a: + (0.001) 2b: + (0.01)				
Kim et al. (2009b)	Total deficiencies (C) Quality of care deficiencies (C) Serious deficiencies (C)	RN HPRD	+ (0.001) + (0.01) n.s.	LPN HPRD	- (0.001) - (0.001) n.s.	NA HPRD	+ (0.001) + (0.001) + (0.05)
Wagner et al. (2013)	Deficiency citation for physical restraint use (including restrictive side rails) (P)	FTE RN/100 beds	+ (0.001)	FTE LPN/100 beds	+ (0.001)	FTE NA/100 beds	- (0.01)
Xing et al. (2013)	Poor PAH-Performance (O)	RN/(LPN+CNA)	+ (0.001)				
Lim (2014)	1: <i>OLS model:</i> a: Care (negative logarithm of count of deficiencies related to care) (C) 2: <i>IV model:</i> a: Care (C) b: Count of total deficiencies (C) c: Count of deficiencies related to quality of care (C) d: Count of deficiencies related to quality of life (C) e: Count of other deficiencies (C) f: DUs (%) (O) g: Contractures (%) (O)	FTE RN HPRD	1a: n.s.			FTE NA HPRD	1a: n.s. 2b: n.s. 2c: n.s. 2d: n.s. 2e: n.s. 2f: n.s. 2g: n.s. 2a: n.s.
		RN (difference between initial staffing level and RN mandate)	2a: +			NA (difference between initial staffing level and NA mandate)	

Study	Quality indicators	Registered nurses (RNs)		Licensed practical nurses (LPNs)		Nurses' aides (NAs)	
		Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)	Specific measure	Relationship (+, -, n.s.)
Lee et al. (2014)	1: 2SLS regression: a: DUs (%) (O) b: UTIs (%) (O) c: Weight loss (%) (O) 2: OLS regression: a: Antipsychotic drug use (%) (P) b: Catheter use (%) (P)	Estimated RN HPRD RN HPRD	1a: + (0.05) 1b: n.s. 1c: n.s. 2a: n.s. 2b: n.s.				
Σ positive results			49 (53.3%)		6 (10.9%)		22 (31.0%)
Σ negative results			7 (7.6%)		10 (18.2%)		6 (8.5%)
Σ n.s. results			36 (39.1%)		39 (70.9%)		43 (60.6%)
Σ (total = 218)			92		55		71

ACS: ambulatory care sensitive; HPRD: hours per resident day; C: composite measure; DU: decubitus ulcers; FTE: full-time equivalent; LPN/LVN: licensed practical/vocational nurse; NA: nurses' aide; n.s.: no significant results; O: outcome measure; P: process measure; PAH: potentially avoidable hospitalizations; RN: registered nurse; TDS: total dependency score; UTI: urinary tract infections; "+": significant positive result (significance level); "-": significant negative result (significance level); 2SLS: two-stage least squares;

For the data analysis, most studies used solely rather simple models such as logistic regression (10 studies) or OLS regression (6 studies). Twelve studies analyzed cross-sectional data, while nine studies examined longitudinal panel data. Only six studies reported some type of sensitivity analysis. Four of the six studies demonstrated robustness, while two of the studies did not interpret the results of the sensitivity analysis.

3.2. Staffing Variables

All but one study investigated the effects of RNs, 13 studies analyzed LPNs, and 15 studies included NAs as the variable of interest. In total, we extracted 218 parameter estimates, i.e., marginal effects, describing the relationship between staffing levels and outcomes. The majority analyzed the relationship between RNs and quality (n=92), while the effects of LPNs (n=55) and NAs (n=71) were less researched. Most studies used standardized measures such as hours per resident day or staff per 100 beds as the independent variables, while some studies used measures such as staff expenses as proxies for staffing levels (Carter/Porell 2003).

3.3. Quality measurement

Donabedian's multidimensional definition of structure, process and outcome quality (Donabedian 1966) was predominantly used in the studies to measure the latent construct of health care quality. As quality is a latent construct that is neither objectively nor directly measurable, it must be approximated. Indicators related to process quality include use of physical restraints, catheters and antipsychotic drugs, while outcomes can be approximated, e.g., by the frequency of decubitus ulcers, the proportion of residents with urinary tract infections, hospitalizations, and the occurrence of significant weight loss among residents. An often used proxy that combines process and outcome quality is the number of deficiency citations that are issued if the NH fails to meet federal requirements.

In total, we identified 28 different quality indicators that were used as dependent variables in the models. We classified four as process quality indicators and fifteen as outcome quality indicators. The remaining nine variables were categorized as composite measures of process and outcome indicators (see Table 2).

3.4. Statistical relationships and effects

Across all 218 parameter estimates analyzed, we identified 77 (35.3%) that showed a significant positive relationship, i.e., they showed that higher staffing levels improved quality. However, in 23 (10.6%) cases, a significant negative relationship was found, and in 118 (54.1%) cases, there was no evidence for an investigated relationship ($p > 0.05$). Stratification by nurse type revealed that 53.3% of the 92 parameter estimates measuring the effects of RNs were significantly positive, while this number decreased to 31.0% for the 71 parameter estimates measuring the impact of NAs. This ratio further decreased to 10.9% for the 55 parameter estimates measuring the effect of LPNs. Similarly, the most evidence for a significantly negative relationship, i.e., evidence that higher staffing levels were related to worse outcomes, was found in LPNs (18.2%), followed by NAs (8.5%), and the least evidence for a negative relationship was found for RNs (7.6%) (see Table 2).

Stratification by Donabedian's categories of process and outcome quality variables and composite measures revealed that staffing levels were significantly positively related to 16

(25.8%) of the total 62 dependent process quality variables, while 10 (16.1%) cases of staffing levels were significantly negatively related to process quality. Regarding outcomes, 37 (34.3%) of 108 parameter estimates showed favorable results of increased staffing levels, while the analyzed studies found a significant unfavorable relationship for ten (9.3%) parameter estimates. Studies that evaluated a composite measure of process and outcome quality, e.g., deficiency citations, revealed that higher staffing levels were related to better outcomes in 24 (50.0%) of 48 cases, while only three studies (6.3%) showed evidence of an unfavorable relationship.

Table 1 presents an overview of the models used in the studies. Logistic regression models were employed to produce 67 estimates, while 31 parameter estimates were generated by OLS regression techniques. Studies using these types of models to analyze cross-sectional data found a significant positive relationship between the parameter and outcome in 37 (37.8%) of the cases, while a significant negative relationship was only found in eight (8.2%) of the parameter estimates. Models relying on instrumental variables, such as 2SLS, found a significant positive relationship between staffing levels and outcomes in 13 of 25 cases, while 12 parameter estimates were not significantly related. Studies analyzing panel data, e.g., using panel fixed effects, showed more conflicting results. Of the total 75 reported parameter estimates, 16 were significantly positively and 13 significantly negatively related to the dependent quality variables.

The eight studies that analyzed a single state were more likely to present a significant positive relationship than studies that focused on more than one state. Studies on a single state found that 20 (48.8%) out of 41 parameter estimates were significantly positively related to quality, whereas three (7.3%) of the estimates were significantly negatively related to quality. Studies encompassing more than 10,000 NHs produced more significant results. Seventeen, or half of the total 34 parameter estimates, were significantly positively related to quality, while 12 (35.3%) parameter estimates provided evidence of a significantly negative relationship between staffing levels and quality.

Thirteen of the 21 studies used OSCAR data, and 149 of the total 218 parameter estimates were based on this database. However, these results did not substantially differ from studies relying on other databases. While 78 of the 149 parameter estimates were not significantly related to quality, 49 were significantly positively and 22 significantly negatively related to quality.

3.5. Study quality assessment

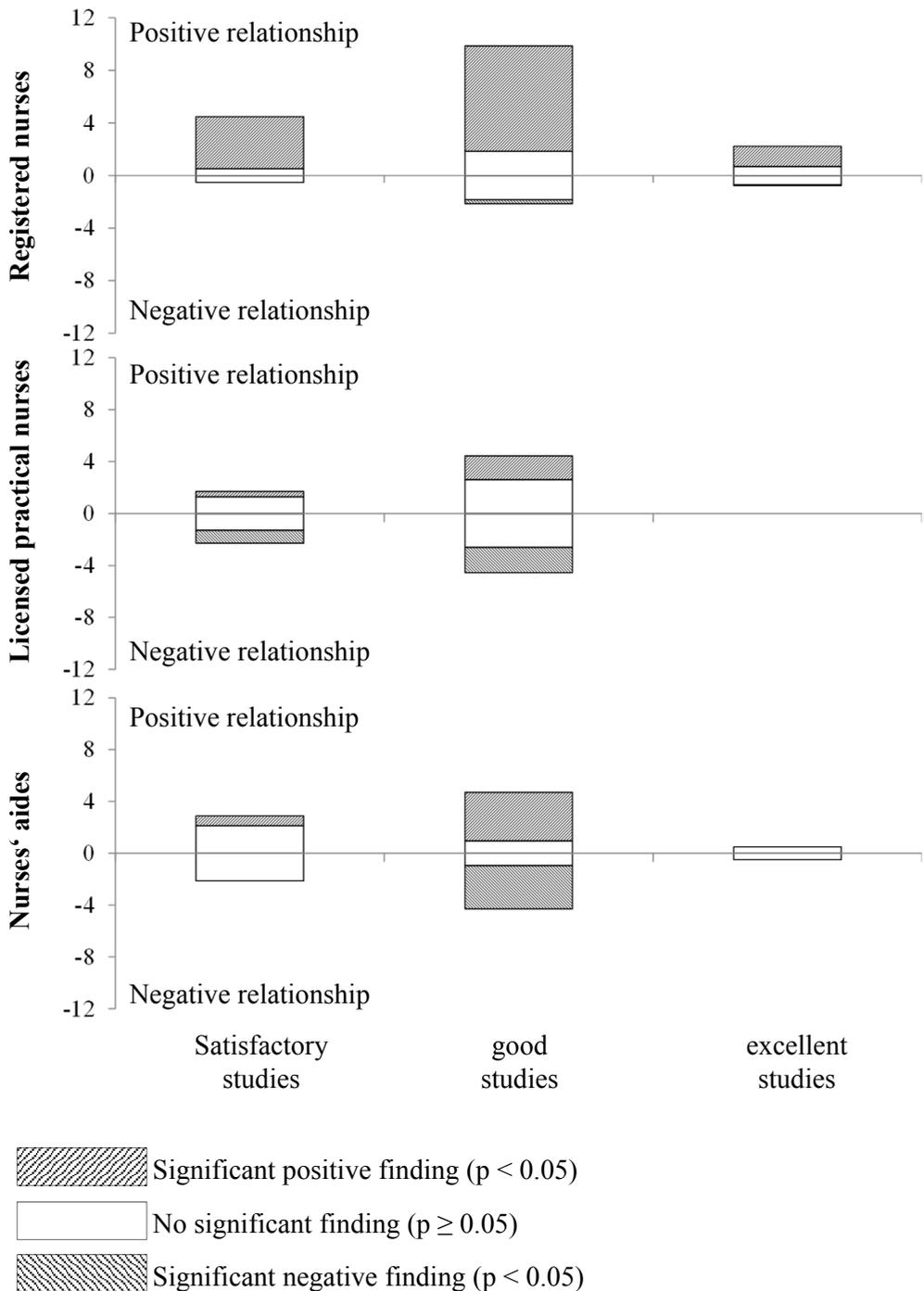
Study quality was heterogeneous, and an average 28.6 out of a maximum 38 points was observed across the 21 studies (see *Table 3*). According to our quality assessment tool, only three studies (*Konetzka et al. 2008; Lee et al. 2014; Lin 2014*) were rated “excellent”, achieving a global score of more than 30 points. Twelve studies were rated “good”, scoring at least 25 points (*Cohen/Spector 1996; Castle 2000; Harrington et al. 2000; Castle 2002; Carter/Porell 2003; Intrator et al. 2004; Zhang, X./Grabowski 2004; Akinci/Krolikowski 2005; Kim et al. 2009a; Kim et al. 2009b; Wagner et al. 2013; Xing et al. 2013*). Six studies were considered “satisfactory”, achieving more than 19 but less than 25 points (*Munroe 1990; Bliesmer et al. 1998; Decker 2006; Dellefield 2006; Zhang, N. J. et al. 2006; Decker 2008*). The satisfactory studies showed the most deficiencies in the “interpretation of the results” category, as generalizability and endogeneity issues were not reported. The reported effect directions by study quality and staff type are shown in *Figure 2*.

	Satisfactory (> 50% - < 65%)						Good (> 65% - < 80%)						Excellent (> 80%)					
<i>Bliesmer et al.</i> (1998)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Decker</i> (2006)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Zhang, N. J. et al.</i> (2006)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Munroe</i> (1990)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Decker</i> (2008)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>DeLiefeld</i> (2006)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Intrator et al.</i> (2004)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Carter/Porell</i> (2003)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Kim et al.</i> (2009a)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>King et al.</i> (2013)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Zhang, X./Grabowski</i> (2004)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Akinci/Krohokowski</i> (2005)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Castle</i> (2000)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Harrington et al.</i> (2000)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Wagner et al.</i> (2013)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Castle</i> (2002)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Kim et al.</i> (2009b)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Cohen/Spector</i> (1996)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Lim</i> (2014)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Konetzka et al.</i> (2008)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
<i>Lee et al.</i> (2014)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41
	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30
	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17
	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26

	Satisfactory (> 50% - < 65%)				Good (> 65% - < 80%)				Excellent (> 80%)							
If biases exist, were they eliminated or included in discussion?	0	0	0	0	1	0	2	1	0	0	1	2	2	2	2	2
Are limitations specified and explained?	2	0	2	2	2	2	2	2	2	2	2	0	1	2	2	2
Is the research question finally responded?	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2
Conflicts of interests	x	x	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	✓	✓
Is the funding source stated?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
No conflicts of interest are available.	20	22	22	23	25	26	26	26	26	26	27	27	27	27	28	28
Total score (max. 38 points):	52.6	57.9	57.9	60.5	65.8	68.4	68.4	68.4	71.1	71.1	71.1	71.1	73.7	73.7	73.7	78.9
Percent achieved	81.6	89.5	89.5	94.7	81.6	89.5	89.5	94.7	81.6	89.5	89.5	94.7	81.6	89.5	89.5	94.7

0: no; 1: partially; 2: yes; ✓: stated; x: not stated;

Figure 2: Study outcomes by study quality and staff type



Results of all reported models were taken into account (n=218). Results were normalized if more than one result was reported per staff category.

4. Discussion

The different qualifications and tasks of the RNs, LPNs, and NAs certainly influenced the results to a certain extent. RNs usually need three years to graduate in the form of a Bachelors' degree, LPNs usually complete their studies after two years (postsecondary non-degree award), while NAs are mostly trained on the job to obtain certification. These differences in education directly translate into different fields of duties and differentiated pay. Similar to the definitions used by the OECD (*Simoens et al. 2005*), RNs are highly specialized staff who supervise other healthcare workers, such as LPNs or NAs, in the US. To address medical needs, RNs work closely with physicians, assess the residents' symptoms, and supervise their compliance with the medical treatment (*Konetzka et al. 2008*). LPNs are trained to provide basic bedside care, measure and document residents' symptoms and perform simple medical tasks such as giving injections and performing laboratory tests. NAs work most of their time directly with the residents, helping them with their daily needs such as eating, dressing, and washing while under supervision of a RN or LPN. In accordance with the different fields of duties and qualifications, the median wages started at \$12.36 per hour for NAs, increasing to \$20.76 for LPNs, and reaching up to \$32.45 for RNs (*Bureau of Labor Statistics 2015*).

As the qualifications, scope of responsibilities, and pay of different nursing levels largely vary, a differential effect on quality outcomes is not surprising. However, it was interesting to note that there seemed to be evidence that RNs influence quality more than LPNs and NAs. Similarly, studies that analyzed the RN/LPN-ratio found quality improvement with increasing ratios. Although we found that 53.3% of the RN-parameter estimates were significantly positively related to quality, the evidence for a relationship between LPN staffing levels and quality was less strong. Only 7.8% of the LPN-estimates were significantly positive related to quality (see *Table 2*). One may therefore conclude that RNs have a greater impact on quality than LPNs. Nursing home managers who focus on a wealthy high-margin population that demand high-quality care should therefore be well advised to provide high levels of RN care. However, federal and state regulations often group RNs and LPNs into one category, i.e., licensed nurses. NH management might therefore be tempted to substitute RNs with LPNs, as they incur substantially less labor costs (*Kim et al. 2009b*). This substitution may subsequently lead to unintuitive findings that higher overall staffing levels produce lower quality (*Kim et al. 2009b*). Therefore, studies analyzing the relationship between staffing outcomes have to account for not only the overall direct care staffing levels but also the staffing structure to obtain valid results. To obtain even more consistent results, staff should be differentiated according to their tasks and fields of duty, as the workload of nurses may often be administrative in nature. In accordance with this notion, *Harrington et al. (2000)* state that the available staffing data may not capture the differences in education levels, capabilities, motivation and experience of the staff and that these unmeasured factors may be even more important than the absolute levels of staff types. Therefore, statistical models should be designed to consider this type of unobserved heterogeneity or to estimate their impact using sensitivity analyses.

4.1. Quality measurement

As quality is a latent and normative construct, the quality indicators used varied substantially between the studies. In the 21 studies analyzed, a total of 34 different quality indica-

tors were adopted to approximate quality depending on data availability, suitability, and the authors' subjective decision criteria. Accordingly, this variation in dependent variables explains some of the variation in the results. We therefore evaluated process quality and outcomes separately. Based on our results, we can conclude that there is only little evidence that staffing levels drive process quality. The parameter estimates describing the relationship between staffing levels and process quality were often insignificant, and the effect sizes of the significant positive and negative results were relatively small. The insensitivity of process quality to changes in staffing level was surprising because one could assume that staffing has a direct influence on processes. Endogeneity may be the root cause of these unintuitive results, as many process quality improvements, such as catheter use, also lead to less labor involvement in the long run. Additionally, one may assume that processes are not extensively complex in the NH environment and that the improvement potential is therefore limited. However, staffing levels seem to drive outcome quality more than process quality. This positive finding was reported in most studies analyzed; almost all staffing level variables showed a positive sign, although not all significant.

Interestingly, the dependent variables that measured a mix of process and outcome quality, e.g., the occurrence of deficiency citations, indicated a positive relationship between staffing levels and outcomes. Only three out of 48 staffing level parameters showed significant negative effects on quality. *Harrington et al.* (2000) assumed that deficiency citations were more sensitive to staffing level because deficiency citations also reflect other problems in the NHs that are not explicitly measured.

4.2. Data sources and statistical approaches

In addition to the included variables, the underlying data as well as the statistical analyses conducted are at least equally important for generating valid estimates. Most studies in this sample used OSCAR data. However, OSCAR data have several limitations, which calls into question the reliability, accuracy and validity of the data. OSCAR only provides a two-week sample of staffing data during a yearly official survey. Furthermore, the staffing data depend on self-reported information on NHs that are not audited by the CMS (*Bostick et al.* 2006). It is therefore possible that the NHs enhance their staffing levels in the short-term for the time of the survey. If this is the case, the OSCAR-reported staffing levels might be superficially inflated. This could result in biased staffing-quality measures. Seven studies addressed this issue by excluding NHs with unrealistic staffing levels. *Dellefield* (2006), *Lin* (2014), and *Zhang, X./Grabowski* (2004) excluded facilities that stated their staffing levels were three standard deviations below or above the average. *Harrington et al.* (2000) and *Zhang, N. J. et al.* (2006) used two standard deviations as the marginal value. *Lee et al.* (2014) fully abstained from the use of OSCAR data and relied on data from the "Colorado State Inspections", and *Kim et al.* (2009b) used the "California Office of Statewide Health Planning and Development" data to identify staffing levels. An additional six studies used OSCAR data without excluding outliers, and thus biased effects are likely if the staff inflation was not random (*Castle* 2000; *Castle* 2002; *Intrator et al.* 2004; *Akinci/Krolukowski* 2005; *Konetzka et al.* 2008; *Wagner et al.* 2013). In addition, OSCAR data do not capture differences in the education, abilities, motivation and experience of the staff. Furthermore, quality indicators in the OSCAR database are reported on the facility level, and the use of MDS data might therefore be more appropriate, as they report quality indicators on the resident level (*Bostick et al.* 2006).

4.3. Sample selection and size

The results and conclusions drawn regarding the estimated effects depend on the sample selection and size. The results of restricted samples, i.e., from one state, should be interpreted with caution, as the results may not be generalized to the whole country or even internationally. A small sample size increases the chances of Type II error (false negative) and leads to an increased chance that the null hypothesis is not rejected although the alternative hypothesis is true. Similarly, *Castle (2008)* notes in his review that the power of national studies that include all NHs is large, but the effect sizes may be small. In this context, he stresses the need to examine the effect sizes. In our review, we found similar results: 43.9% of the effects analyzed by studies that focused on one state only were insignificant, while the effects of studies examining more than 10,000 NHs were insignificant in only 14.7% of the cases.

4.4. Statistical models

The choice of statistical model might also have an impact on the results. The statistical models used are presented in *Table 1*, which shows that linear OLS and non-linear logistic models were the most preferred. These models are widely used and are easy to interpret, but there are doubts regarding whether these approaches are the best choice. *Zhang, X./Grabowski (2004)* showed that the relationship between staffing and quality was possibly non-linear in nature and found support for diminishing marginal returns. In a later study, *Zhang, N. J. et al. (2006)* assessed the strengths and weaknesses of linear and non-linear models by goodness-of-fit indices and concluded that a logistic model fit better than an OLS regression model. However, if the assumption of a non-linear relationship between staffing and quality is true, then the use of an OLS model is inappropriate. An indication of a non-linear relationship can be found in rather low reported adjusted R^2 values. Adjusted R^2 , i.e., the coefficient of determination, describes the proportion of the variance in the dependent variable that is explained by the independent variables. With the exception of *Bliesmer et al. (1998)*, who reported a maximal adjusted R^2 of 0.4059, the dependent variables always explained less than 21.1% of the variance, and most studies reported an adjusted R^2 lower than 0.1. Although there is no common agreed upon threshold for R^2 , interpretations and conclusions should be conducted with care in cases of low R^2 . However, based on our study results, we cannot conclude that the goodness-of-fit of non-linear logistic models outperformed that of linear models. First, seven out of ten studies that employed logistic models did not present any goodness-of-fit statistics, while three studies used different metrics. *Carter/Porell (2003)* reported a pseudo- R^2 value of 0.09, *Zhang, N. J. et al. (2006)* reported deviance values between 1.3372 and 1.3443 (perfect fit = 1), and *Dellefield (2006)* reported the values of likelihood ratio-, Wald-, and score tests, all arguing that the model fit was quite good.

The facilities' number of deficiencies was used in some studies as the dependent variable. The deficiencies were operationalized as either the total number of deficiencies or the number of quality of life or quality of care deficiencies. To analyze integer, non-negative, and likely Poisson-distributed counts, Poisson or more generalized models of the negative binomial family would be the first choice, although one could argue that a linear model might be acceptable if the events were occurring with enough frequency. With an average occurrence of 7.366 deficiency citations per facility per year (*Lin 2014*), this line of rea-

soning can be followed. Earlier studies of *Harrington et al.* (2000), *Akinci/Krolikowski* (2005), and *Dellefield* (2006) used OLS regression to analyze the relationship between staffing and the number of deficiencies. However, more recent studies employed Poisson models in their studies, as they are more appropriate in terms of the discrete nature of the data (*Kim et al.* 2009a; *Kim et al.* 2009b; *Xing et al.* 2013).

Interestingly, a large number of studies did not exploit the richness of information potentially contained in longitudinal data, even if such data were available. Analyses of longitudinal data allow for a greater capacity to capture the complexity of the staff-quality relationship as a result of the ability to control for omitted variable bias, to exploit time-varying effects and to account for outliers. However, only five studies explicitly analyzed longitudinal data using panel regression models (*Zhang, X./Grabowski* 2004; *Konetzka et al.* 2008; *Kim et al.* 2009a; *Kim et al.* 2009b; *Wagner et al.* 2013).

A more recent trend is to account for endogeneity issues between staffing and NH outcomes using an instrumental variable approach. Less care-intensive potential residents will have a better chance of choosing a NH with better staffing ratios and outcomes than a potential resident who needs high levels of care. Additionally, well performing NHs are better able to select residents who are less likely to show negative quality outcomes. In addition, facilities with a high case mix are likely to increase their staffing levels, which leads to an underestimation of the effect of staffing on outcomes. Not accounting for endogeneity would lead to an underestimation of this effect on quality outcomes. Therefore, *Cohen/Spector* (1996) used the ratio of the state's Medicaid spending to general revenue funds for the preceding year, the number of NHs per thousand population in the state, and state and local health spending per capita as instruments to address issues with endogeneity. Two more recent studies from *Lin* (2014) and *Lee et al.* (2014) use the introduction of minimum staffing legislations at the state level and the percent of the population over 65 as well as the percent female in the workforce as instruments to account for endogeneity.

4.5. Quality assessment and outcomes

The existence of a relationship between nurse staffing and quality in NHs cannot be answered in a straightforward manner, as the results are conditional on staff type (see *Figure 2*). In our review, we found that much of the literature supported the evidence of a relationship between RN staffing and better quality. This relationship was consistent regardless of study quality. However, one could question whether there are effects of higher LPN and NA staffing levels on process quality or outcomes. Based on our quality assessment, we could not identify any studies of excellent quality that analyzed the effect of LPNs or NAs on quality. Most studies analyzing these relationships had only satisfactory quality, and most of the findings were not significant. The number of parameter estimates showing a significant positive or a significant negative impact was about equal.

5. Conclusion

According to our review, most research on the staffing-quality relationship focuses on the impact of highly qualified RNs, rather than on the impact of LPNs and NAs. Studies seemed to not provide robust evidence of a relationship of LPN and NA staffing with quality, although many regulatory measures on staffing focused on increasing levels of less educated nursing staff instead of increasing levels of RN staffing. Therefore, policy makers

should ensure that the staffing-quality relationship is fully understood before enacting legislation on long-term care staffing. Additionally, NH managers should be careful when reducing or substituting RN staff to improve short-term financial performance, as deficiencies in quality are likely to negatively impact long-term business opportunities. From a research perspective, further investigation on the relationship between lower qualified staff and outcomes should be conducted. This is especially important as one might assume that education levels, not the number of staff, drive positive quality outcomes. New high quality evidence of the relationship between nurse staffing and quality is not only of great importance for the long-term care sector in the US but also for other countries' considerations. The results can be used to further improve the health workforce policy harmonization efforts implemented by the EU Joint Action on Health Workforce Planning & Forecasting that was commissioned in 2012 (<http://healthworkforce.eu>) or the more recent health workforce qualification draft bill that has been discussed in Switzerland since 2015.

6. References

- Akinci, F./Krolukowski, D. (2005): Nurse staffing levels and quality of care in northeastern pennsylvania nursing homes, in: *Applied Nursing Research*, Jg. 18, Nr. 3, S. 130-137.
- Bliesmer, M. M./Smayling, M./Kane, R. L./Shannon, I. (1998): The relationship between nursing staffing levels and nursing home outcomes, in: *Journal of Aging and Health*, Jg. 10, Nr. 3, S. 351-371.
- Bostick, J. E./Rantz, M. J./Flesner, M. K./Riggs, C. J. (2006): Systematic review of studies of staffing and quality in nursing homes, in: *Journal of the American Medical Directors Association*, Jg. 7, Nr. 6, S. 366-376.
- Bowlis, J. R. (2011): Staffing ratios and quality: An analysis of minimum direct care staffing requirements for nursing homes, in: *Health Services Research*, Jg. 46, Nr. 5, S. 1495-1516.
- Bureau of Labor Statistics. (2015): *Occupational outlook handbook*
- Carter, M. W./Porell, F. W. (2003): Variations in hospitalization rates among nursing home residents: The role of facility and market attributes, in: *Gerontologist*, Jg. 43, Nr. 2, S. 175-191.
- Castle, N. G. (2000): Differences in nursing homes with increasing and decreasing use of physical restraints, in: *Medical Care*, Jg. 38, Nr. 12, S. 1154-1163.
- Castle, N. G. (2002): Nursing homes with persistent deficiency citations for physical restraint use, in: *Medical Care*, Jg. 40, Nr. 10, S. 868-878.
- Castle, N. G. (2008): Nursing home caregiver staffing levels and quality of care: A literature review, in: *Journal of Applied Gerontology*, Jg. 27, Nr. 4, S. 375-405.
- Cohen, J. W./Spector, W. D. (1996): The effect of medicaid reimbursement on quality of care in nursing homes, in: *Journal of Health Economics*, Jg. 15, Nr. 1, S. 23-48.
- Decker, F. H. (2006): Nursing staff and the outcomes of nursing home stays, in: *Medical Care*, Jg. 44, Nr. 9, S. 812-821.
- Decker, F. H. (2008): The relationship of nursing staff to the hospitalization of nursing home residents, in: *Research in Nursing & Health*, Jg. 31, Nr. 3, S. 238-251.
- Dellefield, M. E. (2006): Organizational correlates of the risk-adjusted pressure ulcer prevalence and subsequent survey deficiency citation in california nursing homes, in: *Research in Nursing & Health*, Jg. 29, Nr. 4, S. 345-358.
- Donabedian, A. (1966): Evaluating the quality of medical care, in: *The Milbank Memorial Fund Quarterly*, Jg. 44, Nr. 3, S. 166-206.

- Drummond, M. F./Jefferson, T. O. (1996): Guidelines for authors and peer reviewers of economic submissions to the *bmj*, in: *BMJ*, Jg. 313, Nr. 7052, S. 275-283.
- Evers, S./Goossens, M./de Vet, H./van Tulder, M./Ament, A. (2005): Criteria list for assessment of methodological quality of economic evaluations: Consensus on health economic criteria, in: *International Journal of Technology Assessment in Health Care*, Jg. 21, Nr. 2, S. 240-245.
- Gerkens, S./Crott, R./Cleemput, I./Thissen, J.-P./Closon, M.-C./Horsmans, Y./Beguïn, C. (2008): Comparison of three instruments assessing the quality of economic evaluations: A practical exercise on economic evaluations of the surgical treatment of obesity, in: *International Journal of Technology Assessment in Health Care*, Jg. 24, Nr. 03, S. 318-325.
- Grabowski, D. C./Stewart, K. A./Broderick, S. M./Coots, L. A. (2008): Predictors of nursing home hospitalization: A review of the literature, in: *Medical Care Research and Review*, Jg. 65, Nr. 1, S. 3-39.
- Harrington, C./Zimmerman, D./Karon, S. L./Robinson, J./Beutel, P. (2000): Nursing home staffing and its relationship to deficiencies, in: *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Jg. 55, Nr. 5, S. S278-287.
- Intrator, O./Zinn, J./Mor, V. (2004): Nursing home characteristics and potentially preventable hospitalizations of long-stay residents, in: *Journal of the American Geriatrics Society*, Jg. 52, Nr. 10, S. 1730-1736.
- Kim, H./Harrington, C./Greene, W. H. (2009a): Registered nurse staffing mix and quality of care in nursing homes: A longitudinal analysis, in: *Gerontologist*, Jg. 49, Nr. 1, S. 81-90.
- Kim, H./Kovner, C./Harrington, C./Greene, W./Mezey, M. (2009b): A panel data analysis of the relationships of nursing home staffing levels and standards to regulatory deficiencies, in: *Journal of Gerontology: Social Sciences*, Jg. 64, Nr. 2, S. 269-278.
- Konetzka, R. T./Stearns, S. C./Park, J. (2008): The staffing-outcomes relationship in nursing homes, in: *Health Services Research*, Jg. 43, Nr. 3, S. 1025-1042.
- Lee, H. Y./Blegen, M. A./Harrington, C. (2014): The effects of rn staffing hours on nursing home quality: A two-stage model, in: *International Journal of Nursing Studies*, Jg. 51, Nr. 3, S. 409-417.
- Liberati, A./Altman, D. G./Tetzlaff, J./Mulrow, C./Gotzsche, P. C./Ioannidis, J. P./Clarke, M./Devereaux, P. J./Kleijnen, J./Moher, D. (2009): The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration, in: *BMJ*, Jg. 339, S. b2700.
- Lin, H. (2014): Revisiting the relationship between nurse staffing and quality of care in nursing homes: An instrumental variables approach, in: *Journal of Health Economics*, Jg. 37, S. 13-24.
- Munroe, D. J. (1990): The influence of registered nurse staffing on the quality of nursing home care, in: *Research in Nursing & Health*, Jg. 13, Nr. 4, S. 263-270.
- Ofman, J. J./Sullivan, S. D./Neumann, P. J./Chiou, C. F./Henning, J. M./Wade, S. W./Hay, J. W. (2003): Examining the value and quality of health economic analyses: Implications of utilizing the qhes, in: *Journal of Managed Care & Specialty Pharmacy*, Jg. 9, Nr. 1, S. 53-61.
- Simoens, S./Villeneuve, M./Hurst, J. (2005): Tackling nurse shortages in oecd countries. OECD, Paris, France.
- Wagner, L. M./McDonald, S. M./Castle, N. G. (2013): Nursing home deficiency citations for physical restraints and restrictive side rails, in: *Western Journal of Nursing Research*, Jg. 35, Nr. 5, S. 546-565.

- Xing, J./Mukamel, D. B./Temkin-Greener, H. (2013): Hospitalizations of nursing home residents in the last year of life: Nursing home characteristics and variation in potentially avoidable hospitalizations, in: *Journal of the American Geriatrics Society*, Jg. 61, Nr. 11, S. 1900-1908.
- Zhang, N. J./Unruh, L./Liu, R./Wan, T. T. (2006): Minimum nurse staffing ratios for nursing homes, in: *Nursing Economics*, Jg. 24, Nr. 2, S. 78-85, 93, 55.
- Zhang, X./Grabowski, D. C. (2004): Nursing home staffing and quality under the nursing home reform act, in: *Gerontologist*, Jg. 44, Nr. 1, S. 13-23.

Vanessa Schlage, M.Sc. hat einen Abschluss in Betriebswirtschaft von der Universität Hamburg mit den Schwerpunkten Management im Gesundheitswesen und Marketing. Sie arbeitet als Junior Produktmanagerin bei der DMG Dental Material Gesellschaft mbH.

Anschrift: Universität Hamburg, Hamburg Center for Health Economics, Esplanade 36, 20354 Hamburg, Tel.: +49 40 42838-8041, Fax: +49 40 42838-8043, E-Mail: vanessa.schlage@uni-hamburg.de

Carl Rudolf Blankart, Dr. hat die Juniorprofessur für Technologie- und Qualitätsmanagement im Gesundheitswesen am Hamburg Center for Health Economics (HCHE) an der Universität Hamburg inne. Zurzeit ist er als Commonwealth Fund Harkness/B. Braun Stiftung Fellow in Health Care Policy and Practice an der Brown University School of Public Health in Providence, RI, USA tätig.

Anschrift: Universität Hamburg, Hamburg Center for Health Economics, Esplanade 36, 20354 Hamburg, Tel.: +49 40 42838-4677, Fax: +49 40 42838-8043, E-Mail: rudolf.blankart@uni-hamburg.de

Acknowledgement

Carl Rudolf Blankart was supported by the Commonwealth Fund Harkness/B. Braun Stiftung Fellowship in Health Care Policy and Practice for the 2015/16 academic year at Brown University School of Public Health.

We are grateful for *Ricarda Milstein* for her valuable comments and inputs on this review.