

An Overview of International Staff Time Measurement Validation Studies of the RUG-III Case-mix System

Luke A Turcotte¹, Jeff Poss¹, Brant Fries²
and John P Hirdes¹

¹School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada.

²Geriatrics Center, Department of Internal Medicine and School of Public Health, University of Michigan, Ann Arbor, MI, USA.

Health Services Insights
Volume 12: 1–11
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1178632919827926



ABSTRACT: The RUG-III case-mix system is a method of grouping patients in long-term and post-acute care settings. RUG-III groups patients by relative *per diem* resource consumption and may be used as the basis for prospective payment systems to ensure that facility reimbursement is commensurate with patient acuity. Since RUG-III's development in 1994, more than a dozen international staff time measurement studies have been published to evaluate the utility of the case-mix system in a variety of diverse health care environments around the world. This overview of the literature summarizes the results of these RUG-III validation studies and compares the performance of the algorithm across countries, patient populations, and health care environments. Limitations of the RUG-III validation literature are discussed for the benefit of health system administrators who are considering implementing RUG-III and next-generation resource utilization group case-mix systems.

KEYWORDS: Resource utilization groups, RUG-III, case-mix, long-term care, post-acute care, costs

RECEIVED: November 30, 2018. **ACCEPTED:** December 13, 2019.

TYPE: Review

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Luke A Turcotte, School of Public Health and Health Systems, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada. Email: luke.turcotte@uwaterloo.ca

Introduction

Case-mix classification systems group patients into clinically related groups with homogeneous resource consumption,¹ facilitating the implementation of prospective payment systems wherein a portion of reimbursement is tied directly to patient characteristics. This permits facility reimbursement that is commensurate with the facility's case-mix distribution, allowing for systematic and equitable allocation of financial resources.² Payment systems that fail to account for patient characteristics may incentivize facilities to preferentially admit the least disabled or medically complex patients as they are typically less costly to care for.²⁻⁴ Additional applications of case-mix systems include comparisons among patient populations across health care settings, regions, and time periods⁵ and facility administration such as staffing levels.^{6,7}

The Resource Utilization Groups (RUGs) Version III case-mix classification system was developed in 1994 and was designed to group nursing home residents and skilled nursing facility patients by *per diem* resource use.⁸ Following the structure established by the preceding RUG algorithms (RUG, RUG-II, and RUG-T18),⁹⁻¹¹ RUG-III classifies patients into 44 mutually exclusive groups according to a hierarchy of patient categories (ie, Special Rehabilitation, Extensive Services, Special Care, Clinically Complex, Impaired Cognition, Behavior Problems, and Reduced Physical Functions). Secondary splits are based on an Activities of Daily Living (ADL) Index that is calculated by measuring level of dependence in bed mobility, toilet use, transferring, and eating. Finally, within some hierarchy categories, a tertiary split, such as provision of nursing rehabilitation or depression status, is used to

further classify patients. Patients generally qualify for one of the 7 categories based on clinical characteristics; however, qualification for the Special Rehabilitation category is based primarily on intensity of physical, occupational, and speech-language pathology therapy that patients receive.⁸

RUG-III was developed through a large-scale staff time measurement study of 7658 patients from 176 nursing homes in 6 US states. This study collected resident-specific nursing time using time sheets completed by nursing staff (ie, registered nurses, licensed practical nurses, aides, and orderlies) over the course of their shift. Nursing time was collected over a 24-hour period. Therapists and other auxiliary staff (ie, social workers and transportation aides) reported their resident-specific time using 7-day logs. The RUG-III algorithm explained 56% of the variation in total cost in the derivation data set.⁸

Since the original derivation study, the RUG-III case-mix system has undergone some modifications. In 1998, the United States Health Care Financing Administration (HCFA), now known as the Centers for Medicare and Medicaid Services (CMS), implemented a version of the RUG-III algorithm that reduced the number of groups in the Clinically Complex category in favor of additional Special Rehabilitation groups for patients receiving at least 720 minutes per week of rehabilitation therapies.¹² This version of the algorithm also introduced a concept called "index maximization," wherein patients who qualify for more than 1 RUG-III group are assigned to the highest-weighted group by payment index rather than to the highest group in the hierarchy.¹² In addition to the 44-group RUG-III algorithm used for Medicare programs, a 34-group variant has been created for use by Medicaid programs. In this



variant, patients that qualify for the Special Rehabilitation category are not grouped based on the intensity of rehabilitation therapy they receive.¹³ In 2006, CMS introduced a second version of the RUG-III algorithm with 9 additional groups for patients that qualify for both Special Rehabilitation and Extensive Services categories.¹⁴ Finally, in 2018, a variant of the RUG-III algorithm was developed with the Canadian Institute for Health Information to redefine the qualification requirements for the Extensive Services category (eg, patients with infection requiring isolation) and to reorder the hierarchy for patients in the Behavior Problems and Impaired Cognition categories.¹⁵

A recent advancement to the RUG classification system was the development of the fourth-generation algorithm, RUG-IV, that was implemented by CMS on October 1, 2010.¹⁶ The RUG-IV case-mix system was derived using a sample of 9766 residents from 205 nursing homes and skilled nursing facilities in 15 US states from the Staff Time and Resource Intensity Verification (STRIVE) Project.¹⁷ Special populations including residents with severe mental illnesses, young age, AIDS/HIV, bariatric, and traumatic brain injury were oversampled as they are relatively rare but important groups. The RUG-IV algorithm expands on the 53-group variant of RUG-III with the addition of a Special Care Low category, the joining of the Impaired Cognition and Behavior Problems categories, modifications to category qualifiers and ADL Index breaks, the requirement that select services (eg, intravenous [IV] medications and dialysis) be provided while the resident is in the facility, and the division of therapy time when more than 1 resident is treated concurrently by a single therapist. In the derivation study, the 66-group RUG-IV algorithm explained 42% of the variance in wage-weighted nursing staff time and 62% of the variance in wage-weighted nursing and therapy time.

Despite the widespread use of the RUG-IV case-mix system for Medicare skilled nursing facility payment in the United States, because it relies on items that are only available on the Minimum Data Set (MDS) 3.0 assessment, it has not been adopted internationally. To date, the predictive validity of the RUG-IV algorithm when applied outside of the United States has only been evaluated once. In Ontario nursing homes and skilled nursing facilities (ie, Complex Continuing Care hospitals), RUG-IV explained 42% of the variance in total costs; however, some groups in the algorithm could not be reproduced using items from the MDS 2.0 assessment used in these settings.¹⁸ Currently, RUG-III is used as the basis for prospective payment systems in nursing homes and post-acute care facilities in several Canadian provinces including Ontario and Alberta.

Given that RUG-III was developed in the United States, several international studies have sought to evaluate the utility of the classification system when applied in jurisdictions that differ with respect to contextual factors such as staffing patterns, care processes, and financial arrangements. For the

purpose of implementing prospective payment systems, the utility of RUG-III when applied in a new jurisdiction is determined primarily by assessing the predictive validity of the algorithm to explain variance in resource use. This overview describes the international RUG-III case-mix system validation studies that have been published to date. We describe the samples and staff time measurement methodologies used in each staff time measurement study, then proceed to compare the performance of the RUG-III algorithm across care settings and jurisdictions on the basis of several measures of predictive validity. We conclude with a discussion of the limitations of these validation studies, including considerations for health system administrators who are considering implementing RUG-III case-mix system as the basis for a prospective payment system.

Methods

For this overview of the literature, MEDLINE (PubMed), Scopus, Web of Science, and Google Scholar journal indexes were searched using title/abstract keywords to identify RUG-III validation studies in post-acute, long-term care and hospital care settings. Validation studies were defined as articles describing the application of the RUG-III case-mix system algorithm to explain staff time cost measures. Secondary literature sources were obtained by reviewing citations made by the primary article and using journal indexes to retrieve relevant literature citing the primary article. Relevant articles were screened for inclusion by 2 of the authors (LT and JPH) based on the title and abstract, followed by a review of the article's content. Articles describing previous generation RUG classification systems (ie, RUG, RUG-II, and RUG-T18) and RUG classification systems outside of post-acute and long-term care settings (ie, RUG-III/HC for use in home care settings) were excluded.

Results

Since the initial RUG-III derivation,⁸ the predictive validity of the case-mix system has been assessed in 13 individual studies conducted with long-term and post-acute care populations in Canada,¹⁸ China,¹⁹ the Czech Republic,²⁰ England and Wales,^{6,21} Finland,²² Italy,²³ and Korea.²⁴ Although the terms used describe the care settings where the validation studies were conducted are region-specific, broadly, nursing home residents were represented in 10 studies,^{8,13,18-23,25,26} hospital patients were represented in 3 studies,^{6,24,25} skilled nursing facility patients were represented in 3 studies,^{4,8,18} and patients in rehabilitation facilities were represented in 1 study.²⁷

The 44-group version of the RUG-III algorithm was used in 12 studies.^{4,6,8,18-21,23-27} A 22-group variant of the algorithm that omits tertiary splits for depression status and provision of nursing rehabilitation, and collapses the number of Special Rehabilitation groups to 3 was used in 1 study.²² Two studies used the 34-group variant of the algorithm that is commonly

used by Medicaid programs.^{13,18} One study used a 53-group variant of the algorithm which adds 9 additional groups for high-cost patients that receive both rehabilitation and treatments necessary to qualify for the Extensive Services category.¹⁸ Finally, one study tested a non-hierarchical variant of the 44-group RUG-III algorithm where groups are not mutually exclusive. This same study tested a second “simple” non-hierarchical variant of the RUG-III algorithm where patient classification is based on qualification for a RUG-III category.⁴ Table 1 provides an overview of the care settings, patient populations, and RUG-III algorithm used in each validation study.

Staff Time Measurement Methodology

All the RUG-III validation studies relied on nursing, therapy, and other auxiliary staff to collect their own time measurement data while delivering care. With respect to nursing staff time, 3 studies made use of hand-held computers,^{13,18,26} 1 study used electronic wands,⁴ and 3 studies used paper time sheets.^{6,19,21} The remainder of the studies did not specify how nursing staff time data were collected; however, it is likely that paper time sheets were used. When reported, therapy and auxiliary staff time were collected using paper logs.

Most of the validation studies employed a 24-hour resident-specific nursing staff time measurement methodology which measured both direct (ie, hands-on or bedside) care and indirect care (eg, care planning, family meetings, and physician consultation) provided by nursing staff.^{6,18-25} Resident-specific time refers to direct and indirect time that can be attributed to the care of a single individual. Three studies extended this observation period of resident-specific direct and indirect care time to 48 hours.^{8,13,26} One study allocated a uniform amount of non-resident specific nursing staff time to all patients for activities such as meetings, administration, and breaks.⁴ This study measured resident-specific direct care time over a 48-hour period, but did not measure indirect care time. Another study repeated a 24-hour resident-specific direct care nursing staff time measurement each week, extrapolating this time to all other days until the next staff time measurement was performed. This was repeated throughout the episode of care, up to a maximum of 42 days of stay.²⁷ In one study, care provided by family members and other informal caregivers that replaced formal care time was measured and wage weighted at a rate equivalent to nursing assistants/aides.²²

Similar to nursing staff time measurement, nearly all validation studies relied on a 7-day staff time measurement for rehabilitation staff. Six studies extended this 7-day staff time measurement to auxiliary staff such as social workers and psychologists,^{8,13,18-20,22,25} and 5 studies measured physician time.^{8,19,20,22,25} One study measured rehabilitation staff time over a 24-hour period²³ and another measured daily therapy time for each day the patient was on the unit.²⁷ Three studies did not report or measure rehabilitation or other auxiliary staff time.^{21,24,26} Table 1 provides an overview of the staff time measurement methodology used in each study.

RUG-III Category Distribution

The RUG-III category distribution for each validation study is presented in Table 2. Generally, in studies conducted in nursing homes, residents were most frequently classified into Reduced Physical Functions (24.8%–61.4%), the lowest category in the hierarchy. Clinically Complex (10.9%–42.2%) was the next most frequently populated category. Conversely, the Behavior Problems (0.4%–9.7%) and Extensive Services (0.1%–10.9%) categories were the most sparsely populated in the nursing home studies.^{8,13,18-23,25,26} There was substantial variation in the proportion of residents classified into the Special Rehabilitation category across studies. For example, in 2 studies, nearly one-third of residents were classified as Special Rehabilitation,^{20,23} whereas in 3 other studies, fewer than 2% of residents were classified into this category.^{19,21,24} In the validation studies conducted in hospitals, a large proportion of patients (19.0%–46.3%) were classified into the Clinically Complex category. In 2 studies, nearly half of patients were classified into the first 3 RUG-III categories,^{6,25} whereas in the other hospital study, no patients were classified into those categories.²⁴ All patients in the study conducted in inpatient rehabilitation facilities qualified for the Special Rehabilitation category.²⁷

Explained Staff Time Variance

The proportion of the variance in wage-weighted staff time that is explained by the case-mix system is commonly used to assess the predictive validity of the algorithm in a given care setting or jurisdiction. Wage-weighted staff time is calculated by multiplying staff time utilization for each provider type (eg, registered nurse, licensed practice nurse, and physical therapist) by their wage rate. Given that staff time measurement studies only measure actual staff time allocation, wage-weighted staff time measures across different validation studies can be compared directly without the need to adjust for factors such as jurisdiction-specific staffing regulations and practice patterns. For the purposes of implementing a prospective payment system, greater explained variance is desired and provides some indication of the external validity of the algorithm when applied in a different health care environment.

Among the studies conducted in nursing homes, the RUG-III algorithm explained 14% to 65% of the variance in total wage-weighted staff time^{8,18-20,22,23,25} and 27% to 56% of wage-weighted nursing staff time.^{8,18,19,21,23,26} One study reported explained variance for wage-weighted rehabilitation staff time separately, where the RUG-III algorithm explained 65% of rehabilitation staff time.²³ Similarly, another study reported explained variance for the 34-, 44-, and 53-group variants for RUG-III using the nursing home portion of their sample.¹⁸ The 34-group variant explained 16.5% of the variance in wage-weighted rehabilitation staff time, whereas the 44- and 53-group variants improved the explanatory power to 27.0%. One study reported explained variance for wage-weighted staff time separately for licensed and unlicensed health professionals separately.¹³ In this study, the RUG-III algorithm explained

Table 1. Description of RUG-III validation study samples and staff time measurement methodologies.

STUDY	REGION	PATIENT POPULATION	RUG-III VARIANT	STAFF TIME MEASUREMENT METHODOLOGY
Arling et al ¹³	United States—Colorado, Indiana, Minnesota, Mississippi	5314 residents from 156 units in 105 nursing homes	34-group RUG-III	48-h resident-specific direct care staff time and non-resident-specific staff time measurement for nursing staff. 7-d staff time measurement for ancillary staff (eg, physical therapists and social workers)
Björkgren et al ²²	Finland	1964 residents from 67 units/wards across 10 long-term care facilities	22-group RUG-III	24-h resident-specific direct and indirect care staff time measurement for nursing staff. 7-d staff time measurement for therapists, physicians, and other auxiliary staff. Informal care time provided at the facility by family members and friends that replaced formal care time (wage weight for nursing assistant/aide)
Brizioli et al ²³	Italy—Lazio, Marche, Tuscany, Veneto	999 residents from 11 intermediate and long-term care institutions	44-group RUG-III	24-h resident-specific direct and indirect care staff time measurement for nursing, rehabilitation, and auxiliary staff
Carpenter et al ⁶	England and Wales	1675 patients from 26 hospitals in 8 health districts	44-group RUG-III	24-h patient-specific direct and indirect care staff time measurement for nursing staff. 7-d staff time measurement for physiotherapists, occupational and speech therapists
Carpenter et al ²¹	England	193 nursing home residents from 4 nursing homes	44-group RUG-III	24-h resident-specific direct and indirect care time measurement for registered general nurses and care assistants
Chou et al ¹⁹	China—Hong Kong	1127 residents from 7 residential facilities for the elderly	44-group RUG-III	24-h resident-specific direct and indirect care staff time measurement for nursing staff. 7-d staff time measurement for therapists, physicians, and other auxiliary staff
Eilertsen et al ²⁷	United States	183 hip fracture patients and 292 stroke patients from 27 rehabilitation facilities across 17 states	44-group RUG-III	Sum of patient-specific nursing and therapy time for the duration of stay. 24-h direct care staff time measurement for nursing staff, extrapolated to other days until staff time measurement was repeated. Daily therapy time for therapists and other auxiliary staff. Group therapy time was divided by the number of participants
Fries et al ⁸	United States—Kansas, Maine, Mississippi, South Dakota, Nebraska, Texas, New York	6333 residents from 176 nursing homes; 995 residents from 26 skilled nursing facilities	44-group RUG-III	24-h resident-specific direct and indirect care staff time study for nursing staff. 7-d staff time measurement for auxiliary, including therapists, social workers, and physicians
Hirdes et al ¹⁸	Canada—Ontario	2926 residents from 29 long-term care homes and 1510 post-acute “Complex Continuing Care” hospitals	RUG-III—34-, 44-, and 53-group variants	24-h direct and indirect patient care staff time measurement for nursing staff. 7-d staff time measurement for auxiliary staff, including therapists, dietitians, and social workers
Ikegami et al ²⁵	Japan	531 patients from 4 hospitals with a major LTC component; 55 patients from 1 health facility for the elderly, 285 patients from 3 special homes for the aged	44-group RUG-III	24-h resident-specific direct and indirect care staff time measurement for nursing staff. 7-d staff time measurement for auxiliary staff, including therapists, dietitians, and social workers
Kim ²⁴	Korea	382 patients aged 60+ across 5 long-term care hospitals	44-group RUG-III	24-h patient-specific direct care staff time measurement for nursing staff. 7-d staff time measurement for auxiliary staff, including physicians
Martin et al ²⁶	United States	236 nursing home residents with an intellectual disability	44-group RUG-III	24-h patient-specific direct and indirect care staff time measurement for nursing staff
Topinková et al ²⁰	Czech Republic	1162 residents from 18 long-term care facilities	44-group RUG-III	48-h resident-specific direct and indirect care staff time measurement study

Table 1. (Continued)

STUDY	REGION	PATIENT POPULATION	RUG-III VARIANT	STAFF TIME MEASUREMENT METHODOLOGY
White et al ⁴	United States— Kansas, Maine, Mississippi, Ohio, South Dakota, Texas, Washington, California, Florida, Maryland, Colorado, New York	1304 skilled nursing facility residents with Medicare coverage	44-group RUG-III in addition to “44-Variable Non-Hierarchical” and “Simple Non-Hierarchical” modifications	24-h resident-specific direct and indirect care staff time measurement for nursing time. 7-d staff time measurement for auxiliary staff, including physicians

Abbreviation: RUG-III, Resource Utilization Group Version III.

Table 2. Distribution of RUG-III categories across validation studies.

CARE SETTING	STUDY	SPECIAL REHABILITATION (%)	EXTENSIVE SERVICES (%)	SPECIAL CARE (%)	CLINICALLY COMPLEX (%)	IMPAIRED COGNITION (%)	BEHAVIOR PROBLEMS (%)	REDUCED PHYSICAL FUNCTIONS (%)
Nursing home	Arling et al ¹³	8.5	3.6	7.6	16.6	17.8	1.1	44.8
	Björkgren et al ²²	4.1	1.9	3.6	42.2	3.9	9.7	34.6
	Brizioli et al ²³	28.8	5.7	15.9	11.6	8.6	4.6	24.8
	Carpenter et al ²¹	1.0	1.0	32.6	10.9	17.4	5.2	43.0
	Chou et al ¹⁹	1.8	0.1	7.9	21.4	13.4	0.4	61.4
	Fries et al ^{8,a}	7.2	2.1	10.0	31.4	10.1	1.5	37.6
	Hirdes et al ¹⁸	1.1	0.7	6.9	18.7	15.8	2.8	54.0
	Ikegami et al ²⁵	5.6	2.1	4.1	29.7	16.8	2.1	39.6
	Martin et al ²⁶	14.8	8.5	15.7	16.1	17.4	0.9	26.7
	Topinková et al ²⁰	29.8	2.2	9.0	17.1	8.5	4.7	28.7
Hospital	Carpenter et al ⁶	26.4	10.9	11.5	35.3	23.9	1.0	18.8
	Ikegami et al ²⁵	19.2	14.5	13.6	19.0	10.5	0.9	22.3
	Kim ²⁴	0	0	0	46.3	9.4	17.0	27.2
Skilled nursing facility	Hirdes et al ¹⁸	27.2	17.6	16.0	26.2	1.6	0.1	11.4
Inpatient rehabilitation	Eilertsen et al ²⁷	100.0	0	0	0	0	0	0

Abbreviation: RUG-III, Resource Utilization Group Version III.

^aNursing home and skilled nursing facility patients presented as a combined sample.

20% of the variance in wage-weighted staff time for licensed health professionals and 23% of wage-weighted staff time for unlicensed health professionals.¹³

Among patients in skilled nursing facilities, RUG-III algorithm explained 20.6% to 40.1% of the variance in wage-weighted total staff time and 47.6% to 66.5%

of the variance in wage-weighted rehabilitation staff time (Table 3).⁴

In addition to measuring wage-weighted staff time, this study estimated total *per diem* cost using Medicare claims to estimate non-therapy ancillary costs such as diagnostic services, supplies, and prescription drugs. Overhead costs

Table 3. Proportion of staff time variance explained by RUG-III and CMI range and sub-sample mean.

CARE SETTING	STUDY	STAFF TIME EXPLAINED VARIANCE	CMI RANGE AND SUB-SAMPLE MEAN
Nursing home	Arling et al ¹³	Explained 20% of the variance in licensed/professional resident-specific staff time, and 23% of the variance in unlicensed resident-specific staff time	Mean CMI: Special Extensive = 1.383, Rehabilitation = 1.292, Special Care = 1.190, Clinically Complex = 1.028, Impaired Cognition = 0.809, Behavior Problems = 0.769, Physical = 0.952
	Björkgren et al ²²	Explained 38% of the variance in total wage-weighted patient-specific time	Range = 0.42-2.52
	Brizioli et al ²³	Explained 45% of the variance in wage-weighted nursing time, 61% of the variance in rehabilitation time, and 65% of the variance of total staff time	Range = 0.451-2.535
	Carpenter et al ²¹	Explained 56% of the variance in wage-weighted nursing staff time	Approximate range = 0.40-1.42 (estimated from figure)
	Chou et al ¹⁹	Explained 28.8% of the variance in nursing staff time; explained 27.0% of the variance in wage-weighted nursing staff time. Explained 21.2% of the variance in all staff time; explained 14.1% variance for wage-adjusted all staff time	Range = 0.52-1.91 (among groups with 10+ cases)
	Fries et al ^{8,a}	Explained 55.5% of the variance in wage-weighted total staff time for direct care time, 52.1% of the variance in wage-weighted total staff time for direct and indirect care, and 41.2% of the variance in wage-weighted nursing staff time	Range = 0.39-3.68
	Hirdes et al ¹⁸	Total Sample: 34-group RUG-III explained 38.6% of the variance in wage-weighted staff time, 35.4% of the variance in wage-weighted nursing staff time, and 45.5% of the variance in wage-weighted rehabilitation staff time. 44-group RUG-III explained 39.9% of the variance in wage-weighted staff time, 34.9% of the variance in wage-weighted nursing staff time, and 66.2% of the variance in wage-weighted rehabilitation staff time. 53-group RUG-III explained 42.6% of the variance in wage-weighted staff time, 37.5% of the variance in wage-weighted nursing staff time, and 66.5% of the variance in wage-weighted rehabilitation staff time. Nursing Home Sample: 34-group RUG-III explained 29.7% of the variance in wage-weighted staff time, 33.1% of the variance in wage-weighted nursing staff time, and 16.5% of the variance in wage-weighted rehabilitation staff time. 44-group RUG-III explained 29.9% of the variance in wage-weighted staff time, 33.2% of the variance in wage-weighted nursing staff time, and 27.0% of the variance in wage-weighted rehabilitation staff time. 53-group RUG-III explained 29.9% of the variance in wage-weighted staff time, 33.2% of the variance in wage-weighted nursing staff time, and 27.0% of the variance in wage-weighted rehabilitation staff time	Range = 0.520-3.493 (combined nursing home and skilled nursing facility sample), M = 0.656
	Ikegami et al ²⁵	Explained 42.4% of the variance for un-weighted staff time; explained 43.8% of the variance for wage-weighted staff time. Explained 54.3% of the variance using facility identifiers as covariates; explained 62.7% of the variance with the wards used as covariates	Mean for health facility for the elderly = 0.67, mean for special homes for the aged = 0.73-0.85. Range = 0.5-3.6 (estimated from figure)
	Martin et al ²⁶	Explained 33.3% of the variance in wage-weighted nursing time	Not presented
	Topinková et al ²⁰	Explained 59% of the variance in wage-weighted total staff time	Range = 0.39-2.70
Hospital	Carpenter et al ⁶	Explained 33.4% of the variance in wage-weighted staff time among all patients. explained 49.2% of the variance for patients in acute wards, 45.6% for patients in acute/rehabilitation wards, 39.8% for patients in rehabilitation wards, 33.9% for patients in rehabilitation/long-stay wards, and 29.1% for patients in long-stay wards	Approximate range = 0.5-2.1 (estimated from figure)
	Ikegami et al ²⁵	See row for corresponding nursing home section. Explained variance was not reported separately for hospital sample	Mean for hospitals = 1.10-1.27 (estimated from figure)

Table 3. (Continued)

CARE SETTING	STUDY	STAFF TIME EXPLAINED VARIANCE	CMI RANGE AND SUB-SAMPLE MEAN
	Kim ²⁴	Not presented	Range=0.81-1.47
Skilled nursing facility	Hirdes et al ¹⁸	Total Sample: see row for corresponding nursing home section. 34-group RUG-III explained 15.8% of the variance in wage-weighted staff time, 15.9% of the variance in wage-weighted nursing staff time, and 36.8% of the variance in wage-weighted rehabilitation staff time. 44-group RUG-III explained 17.4% of the variance in wage-weighted staff time, 14.4% of the variance in wage-weighted nursing staff time, and 63.3% of the variance in wage-weighted rehabilitation staff time. 53-group RUG-III explained 22.5% of the variance in wage-weighted staff time, 19.1% of the variance in wage-weighted nursing staff time, and 63.6% of the variance in wage-weighted rehabilitation staff time	Range=0.520-3.493 (combined nursing home and skilled nursing facility sample), M= 1.008

Abbreviation: CMI, Case-mix Index; RUG-III, Resource Utilization Group Version III.

^aNursing home and skilled nursing facility patients presented as a combined sample.

estimated from the United States Federal Reimbursement Rate were also added to the cost measure. RUG-III algorithm explained 10.4% of the variance in total costs for Medicare patients, whereas the 44-Variable Non-Hierarchical and Simple Non-Hierarchical variants of the algorithm explained 24.9% and 21.1%, respectively.⁴ For patients in rehabilitation facilities, the RUG-III algorithm explained 11% of the variance in total staff time, 11% of the variance of nursing staff time, and 14% of the variance in rehabilitation staff time.²⁷ Finally, among hospital patients, RUG-III explained 33.4% of wage-weighted total staff time.⁶ The other validation studies conducted in hospitals either did not report explained staff time variance²⁴ or did not report it separately for the hospital portion of their sample²⁵ (Table 3).

Coefficient of Variation

Outside of the RUG-III derivation study, 5 studies reported the coefficient of variation (σ / μ) for wage-weighted staff time to describe the homogeneity of resource use within case-mix groups.^{6,8,19,20,22,23} This statistic is used instead of the standard deviation for the mean wage-weighted staff time for each case-mix group because the standard deviation for a mean generally increases proportionally with the mean. The coefficient of variation standardizes the standard deviation of the mean, which allows the variability of resource within RUG-III groups to be compared directly, regardless of the magnitude of the mean staff time measure for a given group. A low coefficient of variation across all RUG-III terminal groups is desirable as it indicates that there is little resource use variation within groups for a broad range of patient types.

Across RUG-III groups, the coefficient of variation was lowest in the Italian nursing home study, where it was less than 0.5 for all 44 RUG-III groups.²³ The algorithm performed similarly in the Czech Republic nursing homes where the mean coefficient of variation between groups was 0.5.²⁰ In comparison, the mean coefficient of variation among residents in Hong Kong nursing homes was 1.17 and ranged between 0.83 and 4.63 among groups that contained at least 10

residents.¹⁹ The mean coefficient of variation was 0.65 for the 22-group variant of the algorithm applied in Finnish nursing homes, with 9 groups scoring less than 0.5 on the metric.²² Finally, among hospital patients, the mean coefficient of variation was 0.60. Three-quarters of the RUG-III groups in this study had a coefficient of variation less than 0.5.⁶

Case-mix Index

Case-mix Index (CMI) is a numerical representation of the resource intensity of 1 RUG-III group relative to another.¹ For each validation study, a figurative “average case,” based on the mean wage-weighted staff time of the entire study sample, is assigned a CMI value of 1.0. A group that consumes 50% more resources than the average case is assigned a CMI of 1.5, whereas one that consumes 50% fewer resources than the average case is assigned a CMI of 0.5.^{26,28} One study reported that the mean CMI was highest for patients in hospitals (1.10-1.27), followed by homes for the aged (0.73), and health facilities for the elderly (0.73).²⁵ Similarly, another study reported the mean CMI for the nursing home and skilled nursing facility portions of their sample separately.¹⁸ The mean CMIs were 0.651 and 1.008, respectively. Among the validation studies conducted in nursing homes, there was a 3.5- to 7.2-fold difference between the least and most resource intensive groups.^{8,18-20,22,23} Among studies completed in hospitals, there was a 1.8- to 4.2-fold difference between groups.^{6,24} Table 3 presents an overview of the CMI range and mean for each validation study.

In addition to the derivation study,⁸ 4 validation studies published CMI values for each of the groups in the 44-group variant of the RUG-III system.^{18-20,23} The CMI values across all 4 validation studies were strongly correlated (Pearson $|R^2| = 0.85-0.92$ with the derivation study and followed a similar “sawtooth” pattern. However, compared with the derivation study, cost weights for common RUG-III groups, especially in the Special Rehabilitation and Extensive Services categories, were generally lower in Italian nursing homes,²³ Czech nursing homes,²⁰ and

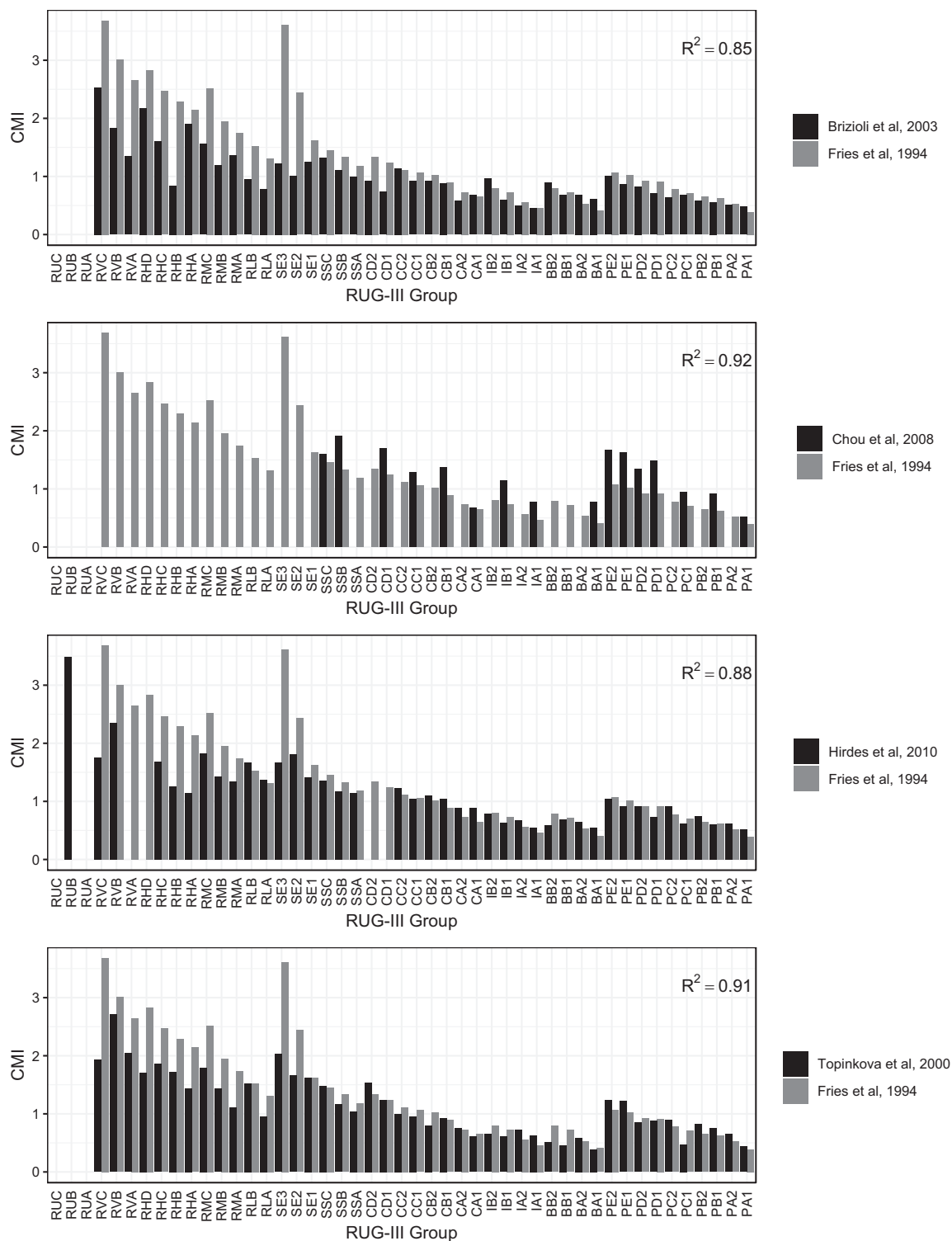


Figure 1. Comparison of RUG-III group and case-mix index (CMI) values by validation study.

Canadian nursing homes and skilled nursing facilities¹⁸ (Figure 1).

Discussion

The RUG-III case-mix classification system is a widely implemented method of grouping individuals in post-acute and long-term care settings. In part, this is because the patient characteristics and provision variables that are used in

the algorithm are collected as part of routine comprehensive clinical assessment using the MDS 2.0 assessment and the interRAI suite of instruments that are applicable to these care settings. In addition to case-mix classification, these instruments are used for patient care planning, outcome measurement, and health system performance measurement.²⁹⁻³²

Since its initial derivation, many studies have evaluated the utility of the RUG-III case-mix classification system

in multiple different patient populations, care settings, and countries around the world. These validation studies employed similar staff time measurement methodologies and reported on the explanatory power of the algorithm using either total, nursing-specific, or rehabilitation-specific staff time cost measures. Although the proportion of the variance in resource use that is explained by the algorithm varied across studies, in most environments, the explanatory power of the algorithm did not differ substantially from the derivation study.⁸ Based on this criterion, RUG-III and other next-generation RUG systems are likely to have utility as the basis of prospective reimbursement of nursing and other auxiliary staff time costs in most post-acute and long-term care settings outside of the United States.

The RUG-III case-mix system has been criticized because the algorithm that groups patients into the Special Rehabilitation category is “self-evident” in its explanation of staff time variance.⁶ This is because patients are grouped based on the amount of therapy that is provided, as opposed to need for rehabilitation.²⁵ In effect, RUG-III acts as a fee-for-service pass-through for rehabilitation reimbursement in prospective payment systems. The merits of this approach have been discussed elsewhere^{8,11,33,34}; however, this aspect of the RUG-III algorithm also has implications when evaluating the predictive validity of the case-mix system for a given health care environment. RUG-III validation studies where a large proportion of the sample is classified into the Special Rehabilitation category are expected to explain a greater proportion of total wage-weighted staff time variance. This is because a large proportion of the costs associated with caring for a patient receiving rehabilitation are built directly into the classification system. For example, excluding Special Rehabilitation patients in 1 validation study resulted in a 29% reduction in explained variance for resource use.⁶

When comparing nursing home RUG-III validation studies with comparable staff time measurement methodologies, a positive relationship between the proportion of the validation sample that is classified as Special Rehabilitation and RUG-III’s explanatory power is observed.^{18,19,20,23,25} Because RUG-III was designed to accommodate a broad range of patients in long-term care settings, when implemented in specific care settings, certain groups, especially those in the Special Rehabilitation category, may be sparsely populated and result in less stable group cost weight estimates. For example, in the sample used by 19, the 44-group version of the RUG-III algorithm only explained 14.1% of the variance in wage-weighted total staff time because 1.8% of the sample was represented in the Special Rehabilitation category. Two studies accounted for this issue with nursing home populations that are unlikely to receive rehabilitation therapy using collapsed 22- and 34-group versions of the RUG-III algorithm which reduce the number of Special Rehabilitation groups.^{13,22} Although Finland is the only country to use the 22-group variant of the algorithm,^{22,35} the

34-group algorithm is more widely used in Canadian nursing home populations.

Because classification into Special Rehabilitation category is based on rehabilitation inputs, nearly all patients in inpatient rehabilitation²⁷ and skilled nursing facilities⁴ are expected to be classified into first 14 groups in the RUG-III hierarchy. For example, in one sample, 92% of patients were classified into the “Very High” and “Ultra High” Special Rehabilitation levels.²⁷ Homogeneous samples limit the explanatory power of the algorithm in these care settings because correlation estimates between RUG-III groups and staff time costs are attenuated when the sample is distributed among a limited number of case-mix groups. Because of the samples used in these validation studies, inferences about the predictive validity of RUG-III in these rehabilitation-focused care settings should be avoided. Jurisdictions wishing to implement the RUG-III case-mix system as the basis for a prospective payment in inpatient rehabilitation and skilled nursing facilities where most of the patients receive rehabilitation therapy may account for this scenario by deriving cost weights using a broader sample of patients that represent the range of patients typically seen in post-acute and long-term care facilities.

Given that RUG case-mix classification systems are derived using staff time measurement studies, the relative wage-weighted resource intensity of each group is based on observed practice patterns as opposed to “best practice.” In the case of the RUG-III derivation study, potential poor-quality facilities were excluded from the sample in an effort to ensure that group weights (ie, CMIs) were based on reasonable resource allocation patterns. In addition, many facilities across numerous US states were included in the sample to reduce the influence of facility and region-specific practice patterns.⁸ As highlighted by Bowlblis and Brunt,³⁶ there are a lack of clear treatment guidelines in nursing home and post-acute care facilities, allowing providers some discretion when allocating therapy time. Given that classification within the Special Rehabilitation category is partly based on therapy intensity process measures, there is some evidence to suggest that providers may exploit this feature of the algorithm to increase revenue.^{36,37}

The RUG-III derivation study implemented several processes, including structured training, unit pilot testing, daily shift audits, and a 24-hour support telephone line, to ensure that staff time was collected reliably.⁸ Across the 13 validation studies, the reliability of the staff time measure used to assess the predictive validity of the RUG-III algorithm was only reported by 1 study,⁶ where discrepancies between total and recorded time were found in 30% of the nursing time sheets. Some studies sought to reconcile total shift time with patient and non-patient-related time^{6,21}; however, one study²⁷ found that only direct patient time could be recorded reliably. Based on this finding, they excluded indirect patient time from their cost measure. Staff time measurement studies are both costly and burdensome. Therefore, most studies constructed their

staff time cost measure based on only a 24- or 48-hour observation period for nursing staff, which may be easily influenced by performance bias. Unfortunately, none of the validation studies that measured staff time over a 48-hour period reported the extent to which staff time varied between days.^{4,13,26} Some more recent validation studies used electronic devices to measure staff time^{4,18,26}; however, it is unknown whether these tools increase the reliability of the staff time measurement by reducing self-report bias.

Similarly, just as providers may attempt to increase rehabilitation intensity during assessment periods to influence classification into Special Rehabilitation RUG-III groups, it is unclear to what extent the intensity of rehabilitation therapy that was provided during the staff time measurement study differed from the remainder of the episode of care. The RUG-III case-mix system is designed to classify patients by relative *per diem* resource use, and although resource use is expected to vary over a patient's episode of care,²⁸ providers implementing case-mix classification systems for the purposes of prospective reimbursement are required to balance patient reassessment burden with the accuracy of resource use measures over the length of stay.

Although many validation studies evaluated the inter-rater reliability of the MDS 2.0 assessment items used in the RUG-III algorithm,^{6,19,22,23,25} only select studies reported the difference in time between the staff time measurement study period and patient assessment for case-mix classification. When reported, this gap ranged between 7 days and 4 weeks^{6,13,21,23}; however, the rehabilitation intensity portion of the classification is based on rehabilitation provided during the staff time measurement period. This means that patients are classified into the Special Rehabilitation RUG-III category on the basis of information collected during the staff time measurement study, whereas all other patients are classified based on previous or future health status. Future validation studies should seek to measure patient characteristics as close to the staff time measurement period as possible to ensure that cost weights are based on current patient status.

Conclusions

This review of international validation studies of the RUG-III case-mix system indicates that the algorithm explains a similar amount of variance in resource use as the derivation study in most post-acute and long-term care settings is outside of the United States. Based on descriptions of the methods used in these studies, the methodological quality of these staff time measurement studies was reasonable. However, it was difficult to assess the accuracy of the cost measures used to evaluate the predictive validity of the algorithm and some studies used relatively homogeneous samples which attenuated the explanatory power of the algorithm. Health system administrators that are considering implementing RUG-III as the basis for a prospective payment system should be aware of the limitations of these validation study limitations when surveying the case-mix literature.

Author Contributions

LAT and JPH screened relevant articles for inclusion. LAT wrote the manuscript. JPH, JP and BF provided critical feedback. All authors have read and approved the final manuscript.

REFERENCES

- Costa AP, Poss JW, McKillop I. Contemplating case mix: a primer on case mix classification and management. *Health Manage Forum*. 2015;28:12–15. doi:10.1177/0840470414551866.
- Hirdes JP, Botz CA, Kozak J, Lepp V. Identifying an appropriate case mix measure for chronic care: evidence from an Ontario pilot study. *Health Manage Forum*. 1996;9:40–46.
- Stineman MG, Escarce JJ, Goin JE, Hamilton BB, Granger CV, Williams SV. A case-mix classification system for medical rehabilitation. *Med Care*. 1994;32:366–379.
- White C, Pizer SD, White AJ. Assessing the RUG-III resident classification system for skilled nursing facilities. *Health Care Financ Rev*. 2002;24:7–15.
- Carpenter GI, Ikegami N, Ljunggren G, Carrillo E, Fries BE. RUG-III and resource allocation: comparing the relationship of direct care time with patient characteristics in five countries. *Age Ageing*. 1997;26:61–66.
- Carpenter GI, Main A, Turner GF. Casemix for the elderly inpatient: Resource Utilization Groups (RUGs) validation project. *Age Ageing*. 1995;24:5–13.
- Dellefield ME. Using the Resource Utilization Groups (RUG-III) system as a staffing tool in nursing homes. *Geriatr Nurs*. 2006;27:160–165.
- Fries BE, Schneider DP, Folley WJ, Gavazzi M, Burke R, Cornelius E. Refining a case-mix measure for nursing homes: Resource Utilization Groups (RUG-III). *Med Care*. 1994;32:668–685.
- Fries BE, Cooney LM Jr. Resource Utilization Groups: a patient classification system for long-term care. *Med Care*. 1985;23:110–122.
- Schneider D, Fries B, Foley WJ, Desmond M, Gormley W. Case mix for nursing home payment: Resource Utilization Groups, version II. *Health Care Financ Rev*. 1988;9:39–52.
- Fries BE, Schneider DP, Foley WJ, Dowling M. Case-mix classification of Medicare residents in skilled nursing facilities: Resource Utilization Groups (RUG-T18). *Med Care*. 1989;27:843–858.
- Health Care Financing Administration. Medicare program; prospective payment system and consolidated billing for skilled nursing facilities (63 FR 26252). *Feder Reg*. 1998;63:26252–26316.
- Arling G, Kane RL, Mueller C, Lewis T. Explaining direct care resource use of nursing home residents: findings from time studies in four states. *Health Serv Res*. 2007;42:827–846.
- Centers for Medicare Medicaid Services. Medicare program; prospective payment system and consolidated billing for skilled nursing facilities for FY 2006; final rule (70 FR 45025). *Feder Reg*. 2005;70:45025–45127.
- Canadian Institute for Health Information. Resource Utilization Groups version III Plus. <https://www.cihi.ca/en/resource-utilization-groups-version-iii-plus>. Accessed November 23, 2018.
- Centers for Medicare and Medicaid Services. Medicare program; prospective payment system and consolidated billing for skilled nursing facilities for FY 2010; minimum data set, version 3.0 for skilled nursing facilities and Medicaid nursing facilities (74 FR 40287). *Feder Reg*. 2009;74:40287–40395.
- Eby J, Pelfrey D, Langenberg K, et al. *Staff Time and Resource Intensity Verification Project Phase II*. Baltimore, MD: Iowa Foundation for Medical Care, University of Michigan, Stepwise Systems, CareTrack Systems; 2011.
- Hirdes JP, Poss JW, Fries BE, et al. *Canadian Staff Time and Resource Intensity Verification (CAN-STRIVE) Project: Validation of the Resource Utilization Groups (RUG-III) and Resource Utilization Groups for Home Care (RUG-III/HC) Case-Mix Systems*. Final Report to the Ontario Ministry of Health and Long-term Care. Waterloo, ON: University of Waterloo; 2010.
- Chou K-L, Chi I, Leung JC. Applying Resource Utilization Groups (RUG-III) in Hong Kong nursing homes. *Can J Aging*. 2008;27:233–239.
- Topinková E, Neuwirth J, Mellanova A, Stankova M, Haas T. Case-mix classification in post-acute and long-term care. Validation of Resource Utilization Groups III (RUG-III) in the Czech Republic. *Casopis Lek Cesky*. 2000; 139:42–48.
- Carpenter I, Perry M, Challis D, Hope K. Identification of registered nursing care of residents in English nursing homes using the Minimum Data Set Resident Assessment Instrument (MDS/RAI) and Resource Utilisation Groups version III (RUG-III). *Age Ageing*. 2003;32:279–285.
- Björkgren MA, Häkkinen U, Finne-Soveri UH, Fries BE. Validity and reliability of Resource Utilization Groups (RUG-III) in Finnish long-term care facilities. *Scand J Public Health*. 1999;27:228–234.

23. Brizioli E, Bernabei R, Grechi F, et al. Nursing home case-mix instruments: validation of the RUG-III system in Italy. *Aging Clin Exp Res*. 2003;15:243–253.
24. Kim EK. Resource use of the elderly in long-term care hospital using RUG-III. *J Korean Acad Nurs*. 2003;33:275–283.
25. Ikegami N, Fries BE, Takagi Y, Ikeda S, Ibe T. Applying RUG-III in Japanese long-term care facilities. *Gerontologist*. 1994;34:628–639.
26. Martin L, Fries BE, Hirdes JP, James M. Using the RUG-III classification system for understanding the resource intensity of persons with intellectual disability residing in nursing homes. *J Intellect Disabil*. 2011;15:131–141.
27. Eilertsen TB, Kramer AM, Schlenker RE, Hrincevich CA. Application of functional independence measure-function related groups and Resource Utilization Groups-version III systems across post acute settings. *Med Care*. 1998;36:695–705.
28. Carpenter IG, Turner GF, Fowler RW. Casemix for inpatient care of elderly people: rehabilitation and post-acute care. *Age Ageing*. 1997;26:123–131. doi:10.1093/ageing/26.2.123.
29. Bernabei RR, Landi FF, Onder GG, Liperoti RR, Gambassi GG. Second and third generation assessment instruments: the birth of standardization in geriatric care. *J Gerontol*. 2008;63:308–313. <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&cid=18375880&retmode=ref&cmd=prlinks>.
30. Gray LC, Berg K, Fries BE, et al. Sharing clinical information across care settings: the birth of an integrated assessment system. *BMC Health Serv Res*. 2009;9:71. doi:10.1186/1472-6963-9-71.
31. Hirdes JP, Ljunggren G, Morris JN, et al. Reliability of the interRAI suite of assessment instruments: a 12-country study of an integrated health information system. *BMC Health Serv Res*. 2008;8:277. <http://www.biomedcentral.com/1472-6963/8/277>.
32. Ikegami N, Hirdes JP, Carpenter I. Measuring the quality of long-term care in institutional and community settings. In: *Measuring Up: Improving Health System Performance in OECD Countries*. OECD Publishing; 2002:277–293. http://books.google.ca/books?id=IBwqET3nfhgC&printsec=frontcover&dq=10.1787/9789264195950&hl=&cd=1&source=gbs_api.
33. Stineman MG, Morrison M, Morris JN, Leiter P, Markello SJ. Measuring case-mix, severity, and complexity in geriatric patients undergoing rehabilitation. *Med Care*. 1997;35:JS90–JS105.
34. Wodchis WP. Physical rehabilitation following medicare prospective payment for skilled nursing facilities. *Health Serv Res*. 2004;39:1299–1318. doi:10.1111/j.1475-6773.2004.00291.x.
35. Laine J. RUG-III for exploring the association between staffing levels and cost-efficiency in nursing facility care in Finland. *Health Care Manage Rev*. 2006;31:73–77.
36. Bowblis JR, Brunt CS. Medicare skilled nursing facility reimbursement and upcoding. *Health Econ*. 2013;23:821–840.
37. Bowblis JR, Brunt CS, Grabowski DC. Competitive spillovers and regulatory exploitation by skilled nursing facilities. *Forum Health Econ Policy*. 2016;19(1):45–70.