Staff Time and Resource Intensity Verification Project Phase II

09/01/2011

Prepared for:	Sheila Lambowitz, Director, Division of Institutional Post Acute Care Chronic Care Policy Group Center for Medicare Management Centers for Medicare & Medicaid Services 7500 Security Blvd., Bldg S3-020-01 Baltimore, MD 21244
Contract:	Skilled Nursing Facility Time Study
Contract Number:	Contract Number: 500-02-0030\0002
Prepared by:	The Iowa Foundation for Medical Care University of Michigan Stepwise Systems CareTrack Systems

EXECUTIVE SUMMARY

Study Purpose and Scope

The Staff Time and Resource Intensity Verification (STRIVE) study was initiated by The Centers for Medicare & Medicaid Services (CMS) primarily to collect and analyze the time that nursing home staff spend caring for residents, based upon current care practices. The study was performed in two phases: I) Data Collection and II) Data Analysis.

The STRIVE study was the first national nursing home time study undertaken in the U.S. since 1997. Since that time, industry utilization and practice patterns changed considerably, impacting the resources necessary to care for residents. However, since that time, CMS had neither validated the Resource Utilization Group, Version 3 (RUG-III) system nor recalibrated the case-mix weights that support the Medicare Skilled Nursing Facility Prospective Payment System (SNF PPS).

The goals of the STRIVE study were to 1) develop a case-mix classification system reflecting current care protocols and resource needs that is usable for adjusting payment to nursing facilities under the federal Medicare SNF PPS, and, secondarily, to provide the basis for nursing facility payments either by state Medicaid systems or by other payers of nursing facility care; and 2) to consider the effectiveness of new assessment items and new scales, specifically combining assessment items to describe an assessment domain that explains resource utilization.

Phase II of the STRIVE project included an analysis of the data collected in Phase I and recommendations for adjustments to the case-mix weights and RUG groups, based on the analytical findings, that eventually led to the derivation of the RUG-IV system.

Study Sample Population

To ensure that the practices and costs of good nursing home care were fairly documented, Phase I of the STRIVE time study collected resource use and resident assessment data from a national sample of nursing homes. The participation of sampled homes was essential to the validity of the study.

The goals of the STRIVE sample design were: (a) to obtain a sample that could be generalized to the national population without bias and with sufficient precision, and (b) to obtain enough cases from certain important special populations of residents to yield sufficient statistical power to support case-mix analyses.

The sample design used for the STRIVE study was complex, involving clustering, stratification, and sampling with probability proportional to stratum size. The three types of clusters selected were: participating states, participating facilities, and the nursing units within each participating facility. Stratification of facilities was applied within

states. Certain strata were deliberately over-sampled to target certain special populations deemed to be important for case-mix analysis that are relatively rare in the population. In addition, the proportion of available facilities that were selected for the study varied greatly from state to state. As a result, facilities and residents did not have an equal probability of inclusion in the sample. Because of this, sampling weights were developed to allow the calculation of unbiased population estimates from the sample data.

Resident Assessment Items

The STRIVE Assessment Addendum form is provided in Appendix B. It includes items representing new concepts from the nursing home MDS 3.0 assessment that was under development, the OASIS assessment used in home health care, and interRAI assessments used in a variety of settings, in addition to items that would allow us to address the explicit goals of STRIVE such as differentiating between pre- and post-admission services. The STRIVE Phase I Report described the construction of the STRIVE Addendum which contained these new items.

As part of the STRIVE project, tasks to check the feasibility and inter-rater reliability of Addendum items, as potential MDS 3.0 items, were performed. Many of these Addendum items were only included during part of the STRIVE data collection and several went through one or more versions. For many items, it was not intended that they would be part of RUG-IV, although data on limited subsets did permit feasibility and reliability testing.

Over the course of the data collection, as many as 151 potential MDS 3.0 items (including revisions of items) were used in the Addendum. Specifically:

- 71 were items describing post-discharge plans, the size and depth of pressure ulcers, and different versions of the pain items included in the data collection for the sole purpose of MDS 3.0 testing; they were not intended for STRIVE RUG-IV use.
- Beyond these, two items were used only in a 57-resident pilot test and were dropped as being redundant with another item (Date of Assessment) or impractical to collect (Estimated Survival).
- An additional 19 items were MDS 3.0 Patient Health Questionnaire (PHQ) depression items that became available late in the STRIVE study. We tested these for use in RUG-IV, but could not fully test for RUG-IV because they were available only for 51% of the sample. Instead, we used the MDS 2.0 measures of depression for RUG-IV.
- The remaining 59 items designed for testing as potential RUG-IV items were included in most versions of the Addendum and collected for over 96% of the sample (these were modified during or soon after the pilot test facility was completed). These items represent dimensions not included in MDS 2.0 or are items needing improvement based on documented feedback from MDS 2.0 users.

- After the reliability testing (see Section 7.2), we eliminated items likely to have poor incentives.. These items could be associated with higher cost and controlled by a facility, e.g. catheter use.

Analysis Approach

The STRIVE RUG-IV derivation analyses were performed, with rare exception, without the use of sample case weights. While the derivation analyses did not require the use of sample weights that account for the sample design, the data used to derive the Case Mix Indexes (CMIs) did require their use. The CMIs are dependent upon the distribution of cases among RUG groups. Because heavier-care residents were over-sampled, it was important to sample-weight the data to produce group means and CMIs derived from the group means that were nationally representative. Therefore, sample weights were used to produce the CMIs.

Most of the analyses were interdependent, i.e., the results of one analysis could affect another. Since RUG-IV is a hierarchical (top to bottom) classification system, interaction among analyses was appropriately controlled by working down the major categories, beginning with the Rehabilitation Plus Extensive category, and then moving sequentially through the Rehabilitation categories, the Extensive Services category, etc. As these categories in RUG-IV form a hierarchy, with the resident assigned to the first (that is, highest) category to which he or she qualifies, determination of each category refines those who will be involved in the analysis of the next "lower" hierarchy category. Once the major categories were determined, secondary and eventually tertiary splits were considered.

The following analyses were performed:

- Summarization of Staff Time Measurement Results
  - The staff time data were analyzed to provide basic descriptive statistics regarding the daily facility staff time spent caring for residents in the sampled facilities.
  - Resource times were based only upon resident-specific time.
  - Concurrent therapy times were evaluated and a Time-Slice method was selected to allocate all group and concurrent time for therapists to individual residents.
  - Two measures of resource intensity were calculated for residents: per-diem minutes (raw minutes) and per-diem wage-weighted minutes (wage-weighted staff time, or WWST).
- Evaluation of current RUG-III 53-group system as a baseline
  - Evaluated the validity of the RUG-III system currently in use.
  - Compared its performance with that in previous studies.

    - Provided the basis to which improvements to the RUG system with RUG-IV can be compared.
    - Identified structures of the RUG-III system that do not adequately explain resource use or differ from previous studies.
- Evaluation of potential new items/scales
    - Considered the effectiveness of new assessment items and new scales, which combine assessment items to describe an assessment domain that explains resource utilization.
    - Tested the feasibility and inter-rater reliability of potential MDS 3.0 items.
    - Singled out specific assessment domains for specific analyses: pain, depression, frailty, and pre-/post-admission services.
- Consideration of specific areas for potential RUG-IV improvement
    - Pre-admission services
    - Concurrent and other therapy
    - Special units
- Update of the ADL Index
    - Investigated changes that increase variance explanation and improved fit across the whole scale range.
- Revision of RUG hierarchy categories, to determine if the criteria for each category should be changed
    - Use of service-based criteria retained from RUG-III
    - ADL thresholds for categories
    - Use of measured therapy staff time for rehabilitation classification
    - Considered the inclusion of procedural memory, violent behavior, and developmental disability to improve the classification system
    - Addressed groups in a hierarchical sequence, starting with the highest resource use groups
- Evaluation of appropriateness of secondary and tertiary category splits
    - Increased the similarity, or homogeneity, in resource use within the resulting RUG groups.
- Testing for a category representing “frailty”
    - Potential utility in explaining resource use in the lower categories (Impaired Cognition, Behavior Problems, and Reduced Physical Function)
- Examination of special populations

- Determined if emerging “Special Populations” of nursing facility residents would be appropriately classified by the RUG system (e.g., younger residents, resident with traumatic brain injury, etc.).
- Investigated whether these subpopulations themselves or characteristics of these subpopulations would be more generally useful case-mix predicators.

Results

Overall, the RUG-III 53-group system had good baseline predictive ability, with variance explanations comparable to the original study and with derived relative cost measures of similar magnitudes. An implication is that while nursing facility care patterns may have changed in the decade since the derivation of the RUG-III system, the relative cost of different types of residents has not changed substantially. On the basis of these results, it was decided that project efforts should focus on refinement of RUG-III to a RUG-IV system, rather than the derivation of a fully new system paradigm.

At the completion of RUG-IV development and before the Final Rule for RUG-IV was published in August 2009, a documented STRIVE public data set was made available by CMS for public review. Included in the release were all of the variables required to recreate RUG-III and RUG-IV results in the Notice of Proposed Rulemaking for RUG-IV in May 2009.

After the Final Rule was published, a revised STRIVE public data set with documentation was created for public release which included the all of the variables needed to create RUG-III and the Final Rule version of RUG-IV results.

While RUG-IV retains the overall structure of RUG-III, it also incorporates several major changes, including the following:

- Revised major categories and individual RUG groups
- Concurrent therapy time allocated when used to identify residents for the rehabilitation and rehabilitation plus extensive categories
- Selected services (e.g., ventilator/respirator, tracheostomy care, IV medications) only considered if provided in the skilled nursing facility since admission
- Additional category: Special Care Low
- Impaired Cognition and Behavior categories merged
- Multiple changes in specific category qualifiers
  - New items added: infection isolation, shortness of breath, Parkinson’s disease, oxygen with respiratory failure
- New ADL Index scoring and consistent ADL splits used across categories
- Category and group labels changed for the Extensive Services category from “SE” to “ES”.

Restorative Nursing (formerly in RUG-III "Nursing Rehabilitation") was retained as a tertiary split for the lowest two RUG-IV categories; but these splits could be eliminated in the future if the incentives provided are not desired.

The changes in RUG-IV require items to be added or modified in the MDS 2.0 to complete the algorithm. The form in which these items are provided in the instrument, including time frame, delimiters, exclusions, and examples, must be developed. These changes include the following:

- Services coded only in the last 7 days, but, if admission is within 7 days, the service must have been performed in the facility:
  - Tracheostomy care
  - Ventilator/respirator
  - Isolation for active infectious disease
  - Parenteral or IV feeding
  - IV medications
  - Transfusions
  - Oxygen therapy
  - Chemotherapy
  - Dialysis
  - Radiation therapy
- Concurrent therapy time provided both as unallocated (i.e., total therapy time provided) and allocated (i.e., time allocated to all individuals in group by the time slice method)
- Diabetes with daily injections
- Chronic obstructive pulmonary disease (COPD)
- Shortness of breath when lying flat
- Parkinson's disease
- Diabetic foot ulcer.

The full RUG-IV system was substantially superior to the RUG-III system in explaining our wage weighted staff time (WWST) measures of resource use, achieving a 41.5% variance explanation of nursing WWST (compared to 30.0% for RUG-III) and 62.0% (compared to 53.0% for RUG-III) for nursing plus therapy WWST. These results held as well in the independent validation sample. Finally, the ratio of the means of the most resource intense group and that of the least resource intense group was 10.0 to 1 for RUG-IV, exceeding the 9.1-to-1 ratio seen in RUG-III, and demonstrating the ability of RUG-IV to identify rare but costly residents.

Case Mix Indices

A critical step in determining appropriate case-mix adjusted PPS rates was to use the STRIVE data to develop Case Mix Indices (CMIs) for the each of the 66 RUG-IV groups, and then to standardize these CMIs across the population to which they will be applied. The STRIVE CMIs represent relative indices of the staff costs for the groups and are based upon the WWST staff cost means for the groups. Separate STRIVE CMI sets were calculated for nursing staff cost and for rehabilitation therapy staff cost, allowing separate rate components to be established for nursing and rehabilitation therapy staff costs.

Other Studies

Drugs

One of the initial goals of STRIVE was to collect, simultaneously with information about the staffing cost of care, information that would permit estimating the daily cost of prescription drugs. This would permit investigation into predictors of the cost of drugs and possible incorporation of drug cost predictors into RUG-IV. Difficulty obtaining drug information seriously undermined the usefulness of this analysis.

Inter-Rater Reliability

Several assessment items used in the STRIVE Addendum (see Appendix B) were new and untested. As described in the Phase I report, Addendum items were from the MDS 3.0 development effort or interRAI instruments with additional items specifically created to address RUG-III problems (such as post-admission use of services). To assure that any items potentially usable for RUG-IV had appropriate reliability, we tested any items for which there had been no prior reliability test. We were also asked by CMS to test some items being considered for MDS 3.0, even though they would not be appropriate RUG-IV items.

1 Overview

1.1 Overview

The Centers for Medicare & Medicaid Services (CMS) funded a national nursing home staff time measurement study to update the Resource Utilization Groups (RUG-III) case-mix weights, which support the Skilled Nursing Facility Prospective Payment System (SNF PPS) and several state Medicaid payment systems. This study, the Staff Time and Resource Intensity Verification (STRIVE), collected staff resource time, resident assessment data, and resident drug data to be analyzed by the STRIVE team. These data are to be used to update the current case-mix weights and resulting payment rates so that they better reflect current care practices and procedures.

This study consists of two logical phases: Phase I Data Collection and Phase II Data Analysis. The Phase I Data Collection Report previously provided to CMS summarized the following methodologies and results:

- State and facility samples
- Special populations
- State and Facility study coordination
- Time data collection
- Assessment data collection

This report for Phase II of the project summarizes the STRIVE team's analysis of the data collected in Phase I and provides recommendations for adjustments to the case-mix weights and RUG groups based on the analytical findings.

1.2 Project Team

The STRIVE Project team has extensive experience with design, training, and analysis of staff time measurement studies, including SNF time studies and RUG-III development and analysis. Members of the team, led by the Iowa Foundation for Medical Care, are identified in Table 1-1.

Table 1-1. STRIVE Project Team

Team Member	Organization	Role
Jean Eby	Iowa Foundation for Medical Care	Project Director
Dane Pelfrey	Iowa Foundation for Medical Care	Project Manager
Kathy Langenberg, R.N.	Iowa Foundation for Medical Care	Operations Manager
Brant Fries, Ph.D.	University of Michigan	Analytic Task Leader/Research Design Specialist

Robert Godbout, Ph.D.	Stepwise Systems, Inc.	Survey Design Consultant
David Maltiz, Ph.D.	Stepwise Systems, Inc.	Survey Design Consultant
David Oatway, R.N., M.P.H.	CareTrack Systems, LLC	Database Manager

1.3 STRIVE Technical Expert Panel

The STRIVE project team established a Technical Expert Panel (TEP) to review and make specific recommendations to CMS and the project team regarding implementing and conducting the time study. The team identified individuals who would bring different perspectives and expertise in Nursing Home care to the meetings. The individuals selected as participants and observers of the STRIVE TEP are noted in Tables 1-2 and 1-3.

Table 1-2. STRIVE Technical Expert Panel Participants

Participant	**Organization**
Buchanan, Joan	Harvard University
Carter, Carol	Medicare Payment Advisory Committee
Ciolek, Cathy	American Physical Therapy Association (APTA) - Trialliance
Dobson, Al	Lewin Group (Dobson Davanzo)
Greene-Burger, Sarah	National Citizens' Coalition for Nursing Home Reform (NCCNHR)
Hines, Lisa	Private citizen, formerly with CMS
Hirdes, John	University of Waterloo - Ontario, Canada
Hojlo, Christa	U.S. Department of Veterans Affairs
Job, Carol	Myers & Stauffer
Karuza, Jurgis	American Medical Directors Association (AMDA), University of Rochester Medical Center (URMC)
Kramer, Andy	University of Colorado Health Sciences Center
Lazarus, Barry	The Alliance for Quality Nursing Home Care (AQNHC)
Manard, Barbara	American Association of Homes and Services for the Aging (AAHSA)
Moore, Terry	Abt Associates
Ousley, Mary	American Health Care Association (AHCA)
Robinson, Alverta	American Hospital Association (AHA)
Scott-Cawiezell, Jill	University of Missouri
Speil, Steve	Federation of American Hospitals (FAH)
Stein-Lloyd, Leslie	American Occupational Therapy Association (AOTA) - Trialliance

Table 1-3. STRIVE Technical Expert Panel Observers

Observer	Organization
Archuleta, Rochelle	American Hospital Association (AHA)
Carter, Diane	American Association of Nurse Assessment Coordinators (AANAC)
Cholakian, Marianne	Department of Health and Human Services, Office of the Inspector General (OIG)
Cornelius, Betty	Private citizen, formerly with CMS
Edelman, Toby	Center for Medicare Advocacy
Fitzler, Sandy	American Health Care Association (AHCA)
Gruhn, Peter	American Health Care Association (AHCA)
Maher, Carol	American Association of Nurse Assessment Coordinators (AANAC)
Munley-Gallagher, Rita	American Nurses Association (ANA)
Nashimi, Robin	National Quality Forum (NQF)
Polniaszek, Susan	Department of Health and Human Services, Assistant Secretary of Planning and Evaluation (ASPE)
Saliba, Deb	RAND Corporation
Stevens, Lynne	American Speech-Language-Hearing Association (ASHA) - Trialliance
Wade, Kathy	Myers & Stauffer
Wern, Maureen	National Association of Subacute and Post Acute Care (NASPAC)
White, Steve	American Speech-Language-Hearing Association (ASHA) – Trialliance
Woody, Iara	American Association of Homes and Services for the Aging (AAHSA)

2 Background to Project

2.1 History of Resource Utilization Groups

Resource Utilization Groups (RUGs) were initially developed circa 1980 as a case-mix system for nursing homes, i.e., a system to identify those patient characteristics associated with measured nursing facility resource use. Since that time, RUGs have been used primarily as the basis for nursing facility payment systems, including the Medicare SNF PPS and the Medicaid payment systems in over half of US states. RUGs have also been used in program management, as the basis for risk adjustment in quality indicators, adjusters for staffing level analyses, policy-making, and so forth.

2.2 Description of RUG-III

The current case-mix system used in the Medicare SNF PPS is the refined RUG-III, a 53-group system, often known as "RUG-III V5.20" and "RUG-53". It is described in Table 2-1 and displayed in Figure 2-1.

Residents are classified in three steps. The first step places residents into one of eight major categories:

- Rehabilitation Plus Extensive Services
- Rehabilitation
- Extensive Services
- Special Care
- Clinically Complex
- Impaired Cognition
- Behavior Problems
- Reduced Physical Function

The criteria for each category are identified in Table 2.1. Since the categories are ordered by decreasing resource use (i.e., weighted staff minutes), the resident is classified into the first category for which he or she qualifies. Once a major category is determined, additional information about the resident is used to determine the specific RUG-III group in the category. In the case of the "Rehabilitation Plus Extensive" and "Rehabilitation" categories, the resident is classified into a rehabilitation level category according to the intensity of rehabilitation and then split into groups based on a summary measure of four Activities of Daily Living (ADLs). For the Extensive category, the group splits are based upon the number of the three categories the resident would qualify for (Special Care, Clinically Complex, and Impaired Cognition) and whether the resident is receiving IV medication or IV feeding. For all lower categories these "secondary splits" are based on the four ADLs. For the Clinically Complex category a

"tertiary split" of the ADL subgroups is based on depression. For the lowest three categories (Impaired Cognition, Behavior Problems, and Reduced Physical Function) tertiary splits of the ADL subgroups are determined by the presence or absence of nursing rehabilitative services.

Table 2-1. RUG-III Groups

CATEGORY	RUG-III ADL INDEX	END SPLITS	RUG-III GROUP CODE
ULTRA HIGH REHABILITATION PLUS EXTENSIVE SERVICES			
Rehabilitation Rx 720 minutes/week minimum	16-18	Not Used	RUX
AND			
At least 1 rehabilitation discipline 5 days/week	7-15	Not Used	RUL
AND			
A second rehabilitation discipline 3 days/week			
AND			
IV Feeding in last 7 days			
OR			
IV medications, suctioning, tracheostomy care, or ventilator/respirator in the last 14 days			
AND			
ADL score of 7 or more			
VERY HIGH REHABILITATION PLUS EXTENSIVE SERVICES:			
Rehabilitation Rx 500 minutes/week minimum	16-18	Not Used	RVX
AND			
At least 1 rehabilitation discipline 5 days/week	7-15	Not Used	RVL
AND			
IV Feeding in last 7 days			
OR			
IV medications, suctioning, tracheostomy care, or ventilator/respirator in the last 14 days			
AND			
ADL score of 7 or more			
HIGH REHABILITATION PLUS EXTENSIVE SERVICES			
Rehabilitation Rx 325 minutes/week minimum	13-18	Not Used	RHX
AND			
At least 1 rehabilitation discipline 5 days/week;	7-12	Not Used	RHL
AND			
IV Feeding in last 7 days			
OR			
IV medications, suctioning, tracheostomy care, or ventilator/respirator in the last 14 days			
AND			
ADL score of 7 or more			

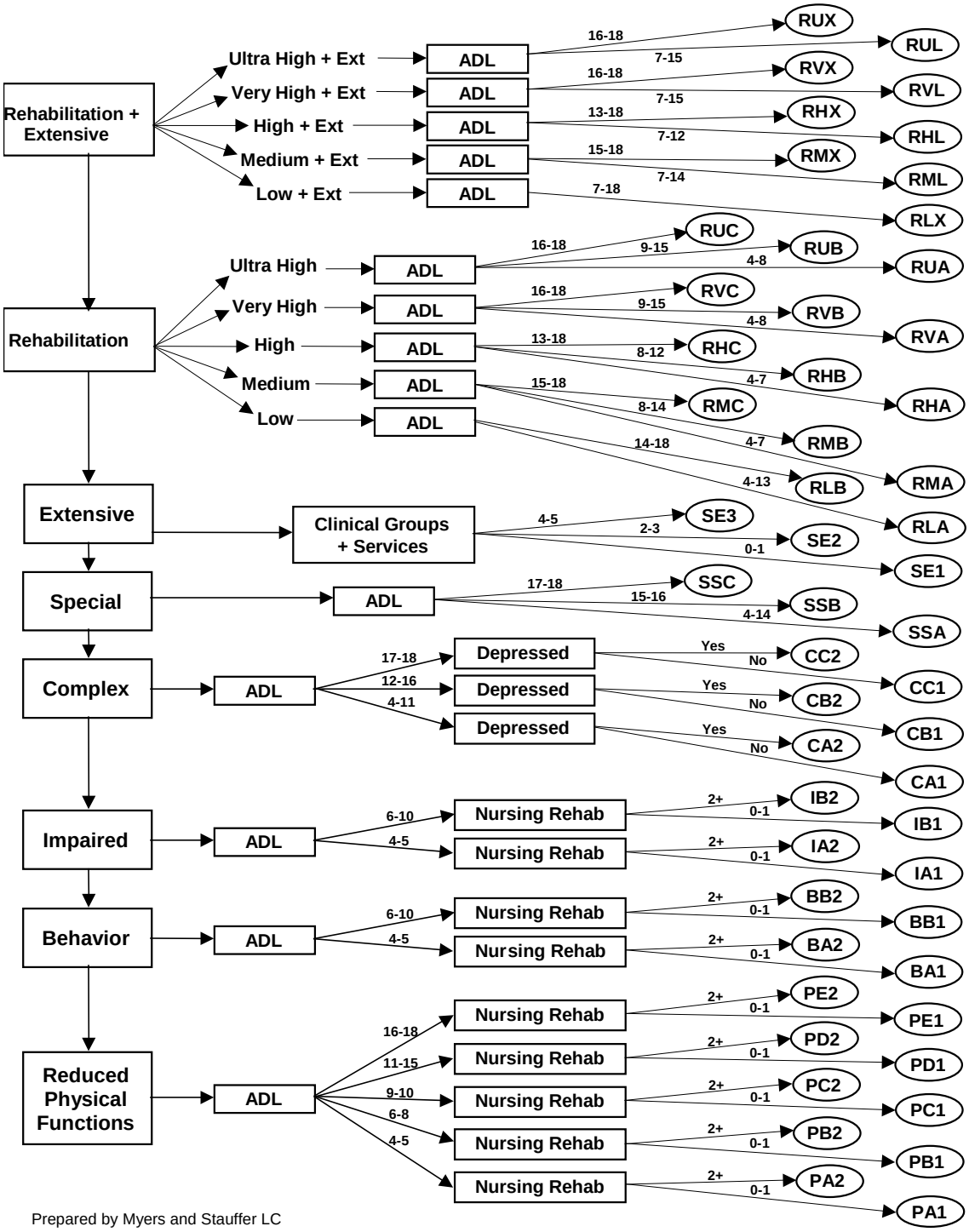
CATEGORY	RUG-III ADL INDEX	END SPLITS	RUG-III GROUP CODE
MEDIUM REHABILITATION PLUS EXTENSIVE SERVICES			
Rehabilitation Rx 150 minutes/week minimum	15-18	Not Used	RMX
AND			
5 days any combination of 3 rehabilitation disciplines;	7-14	Not Used	RML
AND			
IV Feeding in last 7 days			
OR			
IV medications, suctioning, tracheostomy care, or ventilator/respirator in the last 14 days			
AND			
ADL score of 7 or more			
LOW REHABILITATION PLUS EXTENSIVE SERVICES			
Rehabilitation Rx 45 minutes/week minimum	7-18	Not Used	RLX
AND			
3 days any combination of 3 rehabilitation disciplines;			
AND			
Nursing rehabilitation 6 days/week, 2 services (see Reduced Physical Function (below) for nursing rehab services count);			
AND			
IV Feeding in last 7 days			
OR			
IV medications, suctioning, tracheostomy care, or ventilator/respirator in the last 14 days			
AND			
ADL score of 7 or more			
ULTRA HIGH REHABILITATION			
Rehabilitation Rx 720 minutes/week minimum	16-18	Not Used	RUC
AND			
At least 1 rehabilitation discipline 5 days/week	9-15	Not Used	RUB
AND			
A second rehabilitation discipline 3 days/week	4- 8	Not Used	RUA
VERY HIGH REHABILITATION			
Rehabilitation Rx 500 minutes/week minimum	16-18	Not Used	RVC
AND			
At least 1 rehabilitation discipline 5 days/week	9-15	Not Used	RVB
	4- 8	Not Used	RVA
HIGH REHABILITATION			
Rehabilitation Rx 325 minutes/week minimum	13-18	Not Used	RHC
AND			
At least 1 rehabilitation discipline 5 days/week	8-12	Not Used	RHB
	4- 7	Not Used	RHA

CATEGORY	RUG-III ADL INDEX	END SPLITS	RUG-III GROUP CODE
MEDIUM REHABILITATION			
Rehabilitation Rx 150 minutes/week minimum	15-18	Not Used	RMC
AND			
5 days any combination of 3 rehabilitation disciplines	8-14	Not Used	RMB
	4- 7	Not Used	RMA
LOW REHABILITATION			
Rehabilitation Rx 45 minutes/week minimum	14-18	Not Used	RLB
AND			
3 days any combination of 3 rehabilitation disciplines;	4-13	Not Used	RLA
AND			
Nursing rehabilitation 6 days/week, 2 services (see Reduced Physical Function (below) for nursing rehab services count)			
EXTENSIVE SERVICES			
IV Feeding in last 7 days	7-18	Count of other categories	SE3
OR			
IV medications, suctioning, tracheostomy care, or ventilator/respirator in the last 14 days		(special care, clinically complex,	SE2
AND		impaired cognition),	SE1
ADL score of 7 or more		plus IV medications, plus IV feeding	
SPECIAL CARE			
Extensive Services (see above)	17-18	Not Used	SSC
AND			
ADL score of 6 or less	15-16	Not Used	SSB
OR			
Special Care qualifier (any one):	4-14	Not Used	SSA
• CP, MS, or Quad with ADL sum >=10			
• respiratory therapy = 7 days			
• feeding tube (calories >=51%, or calories =26%-50% and fluid >=501 cc) and aphasia			
• radiation tx			
• receiving tx for surgical wounds/open lesions or ulcers (2 sites, any stage; or 1 site stage 3 or 4)			
• fever with dehydration, pneumonia, vomiting, weight loss, or feeding tube (calories >=51%, or calories =26%-50% and fluid >=501 cc)			
AND			
ADL score of 7 or more			

CATEGORY	RUG-III ADL INDEX	END SPLITS	RUG-III GROUP CODE
CLINICALLY COMPLEX			
Special Care qualifier (see above) and ADL score of 6 or less	17-18	Signs of Depression	CC2
OR	17-18	No Signs of Depression	CC1
Clinically complex qualifier(any one):			
• burns	12-16		CB2
• coma		Signs of Depression	
• septicemia	12-16		CB1
• pneumonia		No Signs of Depression	
• receiving treatment for foot lesion/infection	4-11		CA2
• internal bleeding			
• dehydration	4-11	Signs of Depression	CA1
• tube feeding (calories >=51%, or calories =26%-50% and fluid >=501 cc)		No Signs of Depression	
• oxygen			
• transfusions			
• hemiplegia with ADL score >=10			
• chemotherapy			
• dialysis			
• physician visits 1 or more days and order changes 4 or more days (last 14 days)			
• physician visits 2 or more days and order changes 2 or more days (last 14 days)			
• diabetes with injection 7 days/week and order change 2 or more days(last 14 days)			
IMPAIRED COGNITION			
Score on MDS2.0 Cognitive Performance Scale (CPS) >=3	6-10	2 or more nursing rehab services on 6+ days/wk	IB2
AND	6-10	Less nursing rehab	IB1
ADL score of 10 or less	4- 5		IA2
NOTES:		2 or more nursing	
No clinical variables used	4- 5	rehab services on 6+ days/wk	IA1
CPS Score of "6" will be assigned Clinically Complex or PE2-PD1)		Less nursing rehab	
See Reduced Physical Function (below) for nursing rehab services count			
BEHAVIOR PROBLEMS			
Wandering, physical abuse, verbal abuse, inappropriate behavior or resisted care on 4+ days/week	6-10	2 or more nursing rehab services on 6+ days/wk	BB2
OR	6-10	Less nursing rehab	BB1
hallucinations or delusions	4- 5		BA2
AND		2 or more nursing	
ADL score of 10 or less	4- 5	rehab services on 6+ days/wk	BA1
NOTES:		Less nursing rehab	
See Reduced Physical Function (below) for nursing rehab services count			

CATEGORY	RUG-III ADL INDEX	END SPLITS	RUG-III GROUP CODE
REDUCED PHYSICAL FUNCTION Nursing rehab service count: • passive and/or active ROM • amputation/prosthesis care training • splint or brace assistance • dressing or grooming training • eating or swallowing training • transfer training • bed mobility and/or walking training • communication training • scheduled toileting plan and/or bladder retraining program. NOTES: No clinical variables used	16-18 16-18 11-15 11-15 9-10 9-10 6- 8 6- 8 4- 5 4- 5	2 or more nursing rehab services on 6+ days/wk Less nursing rehab 2 or more nursing rehab services on 6+ days/wk Less nursing rehab 2 or more nursing rehab services on 6+ days/wk Less nursing rehab 2 or more nursing rehab services on 6+ days/wk Less nursing rehab	PE2 PE1 PD2 PD1 PC2 PC1 PB2 PB1 PA2 PA1
Default			AAA

Figure 2-1. RUG-III v. 5.20 Case-Mix System



Prepared by Myers and Stauffer LC

Each RUG-53 group was assigned a case-mix index derived from the sample mean wage-weighted staff minutes of that group. We describe later in Section 6 the steps necessary to derive “case-mix indexes” (CMIs) and rates.

In some unusual cases, a resident classified into one group, could qualify for another group lower in the hierarchy yet with a higher CMI (and associated with higher payment). Such “inversions” are described in Section 6.1. To address this problem, CMS established the RUG-III algorithms for “Index Maximization”, i.e., assigning a resident to the RUG-III group with the highest CMI of all groups for which that resident qualified.

With its inception in July 1998, the Medicare Part A SNF PPS used the original RUG-III 44-group system. A few state case-mix Medicaid payment systems have adopted RUG-44, but most state systems have adopted a simplified RUG-III 34-group system. State Medicaid systems rarely pay for rehabilitation care, so the 34-group system collapsed the 14 rehabilitation groups in RUG-44 into 4 rehabilitation groups, differentiated only by ADL function. In January of 2006, the Medicare SNF PPS began using the refined RUG-III 53-group system, while most states continued to use RUG-34 or RUG-44 systems.

3 Overall Approach

In this section, we provide an overview of the entire STRIVE project, including the data collection, database development, and approach to developing the RUG-IV system and its associated case-mix indexes.

3.1 Overview of Data Collection

In our STRIVE Phase I Report[1], we provided details of this project's data collection and database development. These included:

- Sample Design – A complex, probability-based stratified sampling design was developed that involved clustering, stratification, and sampling with probability proportional to stratum size. The goals of this design were twofold: (a) to obtain a probability-based sample that would be unbiased and that could be generalized to the nation, and (b) to collect a sample with enough statistical precision to support the analyses that were anticipated. Analyses presented later in this report (see Section 4.1) were used to evaluate the precision and representativeness of the sample that was obtained.
- Facilities were reviewed for appropriate quality of care, as determined by quality measures and deficiencies reported in the Online Survey and Certification Reporting (OSCAR) system; approximately 11% of facilities were removed from the sampling universe because of low quality of care. Of the remaining facilities, each was classified into strata based on having high Medicare census, being hospital-based, having high concentrations of ventilator or Acquired Immune Deficiency (AIDS)/Human Immunodeficiency Virus (HIV) residents, or "all others." Facilities were chosen to meet strata goals.
- Within a facility, full nursing units were selected in the following order: targeted specialty units (if available), units with high Medicare census, and then all other units. In some cases, often because of geographic proximity, time data were collected as well on a unit "next to" the collection unit. In these instances, when the time data were complete, we included these observations in our study even though the unit was not initially selected for data collection.
- Within selected units, staff times were collected by all nursing, therapy, and ancillary staff providing patient care. Personal Digital Assistants (PDAs) were used by staff to record all time with a resident e.g., hands-on care or non-hands-on care such as a discussion with a physician about the resident, writing notes in the resident's chart, or calling a family member. The staff time measurement (STM) was performed over a 48-hour period for non-

[1] STRIVE PHASE I REPORT, pp. 15, 17, 56

therapy staff. The approach with the PDAs was to collect ***all*** time spent by staff members, then assign the time to the following groups:

- Individual residents;
- Groups of residents;
- General unit tasks (termed "non-resident-specific time." This group included tasks such as generic charting, general distribution of medications, meetings;
- Non-study tasks, such as time spent with residents on other units.

In the analysis, only the resident-specific time was used. When groups of residents were cared for simultaneously, time was apportioned to each person in the group (note that group therapy time was a special case that is discussed more extensively in Section 4.3.1). Residents either admitted or discharged during the two days of the STM were omitted from analysis, as we did not know how much time during these days the resident was in the facility. Therapy staff tracked their time with residents using PDAs when they were available, which usually included the first three or four days in the study week at a facility; on other days therapy staff used paper instruments.

- Facility staff completed an assessment of each resident in the STM with a reference date of Thursday of the study week. The assessment included the MDS 2.0 and a STRIVE Addendum (see Appendix B), the latter including items both for possible use in RUG-IV and for other testing purposes, such as MDS 3.0.
- Other data collection was performed to measure inter-rater reliability (IRR) (dual assessments on a limited number of individuals - see Section 7.2) and obtain Medication Administration Records (MARs) for a subsample of residents (see Section 7.1).

3.2 Analytic Database Development

The analytic databases for the project were assembled from a variety of sources, using several methodologies. Specifically, the analytic sample was derived as follows (see also Figures 3-1 and 3-2):

- **Source: Staff Time Measurement**: The care times collected in the STM (see above). Data records were collected on 10,742 residents. However, these included records describing residents not on study units (12) or on units where the STM process was not completed (149). This resulted in 10,581 records with completed STM information. From this sample, an additional 95 (0.9%) residents with zero nursing staff time were eliminated, resulting in a net 10,486 (99.1%) records.
- **Source: The MDS 2.0 and STRIVE Addendum assessments, and MDS 2.0 Data Archives.** We then matched the STM data with the residents'

assessments. In cases where no STRIVE MDS was performed (e.g., when staff time was collected on "supplemental units" not originally targeted for analysis) or MDS items were missing, we overlaid data using the National MDS database using the CMS Resident Profile Table (RPT) methodology.[2] Of the 10,486 residents with valid, non-zero STM times, 10,136 (96.7%) residents had either a STRIVE MDS, a Resident Profile Table (RPT) assessment, or both.

- **Methodology: Admitted/Discharged Residents.** Of the 10,136 observations with valid, non-zero STM time and MDS/RPT assessments, 415 were admitted or discharged during the two-day STM process. As we did not know the period during the 48 hours that these residents were on the unit to receive nursing care, we dropped them from the analysis, leaving 9,721 (95.9%) observations.
- **Methodology: Missing RUG-III Items and Outlier Nursing Time.** We dropped 14 (0.1%) observations with incomplete data (even after using RPT data to classify into a RUG-III group), resulting in an analytical data set of 9,707 (99.9%) individuals. Late in the RUG-IV derivation process (i.e., RUG-IV Version 7), one outlier with very high nursing staff time was deleted; this reduced the analytical data set to its final size of 9,706 residents.

The analytic database of 9707 observations was divided into two, based on random assignment: a derivation database consisting of two-thirds of the observations (6,454, or 66.5%) and a validation database of the remaining third (3,253 or 33.5%).[3] All derivation was performed using the derivation database except for the analysis of rare subpopulations (see Section 4.12) and the computation of the case-mix indexes (see Section 6). After most of the analyses were completed, the RUG-IV system was applied to the validation database, to assure that the major findings were not spurious. The outlier was found in the reserved validation data set only late in the analysis, and reduced the size of that dataset to 3,252. We describe the validation results in Section 5.4.

The STRIVE Phase I report[4] details the observations that were dropped.

---

[2] STRIVE Phase I Report, pg. 30.
[3] All observations in the STRIVE database with an MDS, a full (two-day) STM measurement, and all the items in the MDS needed to classify into the RUG-III system.
[4] STRIVE Phase I Report, pg. 56-61.

Figure 3-1. STRIVE Analytic Dataset Methodology Resident Counts

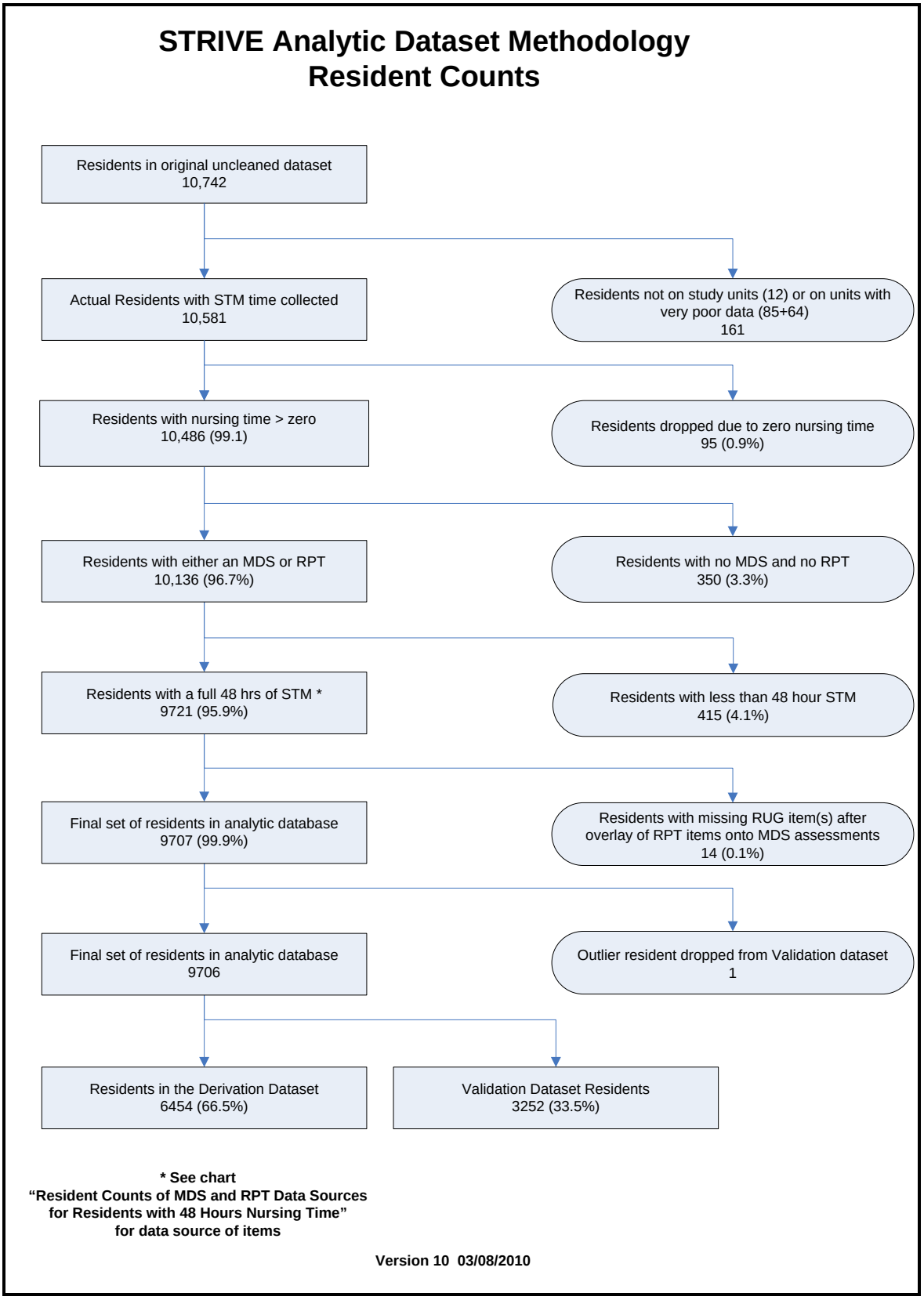
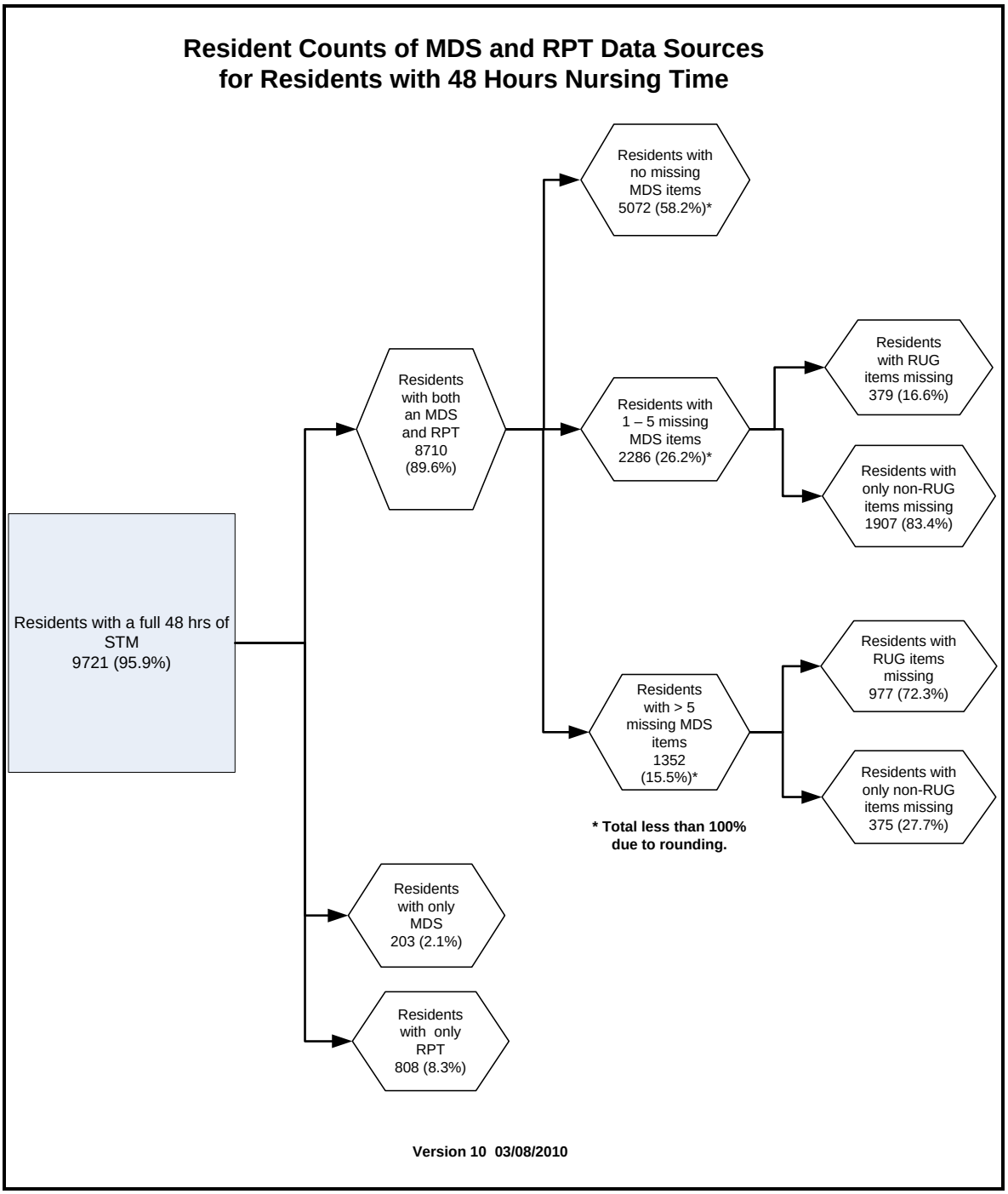


Figure 3-2. Resident Counts of MDS and RPT Data Sources for Residents with 48 Hours Nursing Time



As we indicated earlier, the MDS 2.0 data archives allowed us to use RPT data to fill in missing STRIVE assessment data. Of the 9,721 observations with valid, non-zero STM time and not admitted or discharged during the STM days, the majority (8,710 or 89.6%) had both the STRIVE-generated MDS and the RPT from the archive. The RPT data thus allowed us to fill in missing data items in the STRIVE MDS. For 5,072

(58.23%) of the records, no items were missing; for 2,286 (26.25%), 1-5 items were missing; and for the remaining 1,352 (15.52%) more than 5 items were missing. For the remaining 1,011 of the 9,721 observations, we either had only the STRIVE MDS (203, or 2.1%) or only the RPT assessment values (808, or 8.3%). Those with only the RPT were primarily cases where, due to the physical structure of the nursing unit, the STM was performed on more than just the chosen study unit – such "supplemental units" are described in the Phase I Report.[5]

The resulting analytic database was then supplemented with the following information:

- **Source: Medicare status**. We determined the Medicare status for each resident from both the facility's business office and from Medicare claims.
- **Statistical Methodology: Adjusted Therapy Time.** In preliminary analysis of the STM data, we recognized that therapy time was under-reported on days when therapy staff used paper instruments and were unsupervised. We discuss in the STRIVE Phase I Report[6] and in Section 4.3.4 here the methodology used to adjust therapy times.
- **Statistical Methodology: Weighting Staff Time as a Relative Cost Measure.** We developed a relative cost measure by weighting staff times for each role by their relative wage rate. We discuss in Section 4.3.6 the derivation of nursing and therapy wage-weighted staff times (WWSTs), used as the dependent variable to be explained in the analysis.
- **Statistical Methodology: Truncating Staff Time.** We detected several extremely high and low nursing staff time values in the sample. To keep these values from skewing any analyses, we truncated the staff times for each staff role at the 99th percentile, i.e., any values above the 99th percentile were assigned the value of the 99th percentile. For example, for registered nurses, the 99th percentile of Nursing WWST) was 369, so the 1% of WWST values above this value were reset to 369 (corresponding to 152.5 minutes per day). As we deemed that every resident should have a minimum amount of staff time, we also performed a lower truncation: we assumed that every resident was provided at least 10 minutes (see Section 4.3.3 – Nursing Time).
- **Statistical Methodology: Compute Scales.** We developed a variety of scales to be available for regular testing. These included measures of pain, depression, cognitive function, etc. (see Section 4.5).
- **Statistical Methodology: Code Diagnoses.** We coded ICD-9-CM diagnoses into major diagnostic groups as identified in Table 3-1.

[5] STRIVE Phase I Report, pp. 57
[6] STRIVE PHASE I REPORT, pp. 53

Table 3-1. ICD-9-CM Diagnosis Codes Utilized

Diagnosis	ICD-9-CM Codes
Diabetes mellitus	250, 250.01, 250.02, 250.03, 250.1, 250.12, 250.3, 250.4, 250.5, 250.52, 250.6, 250.61, 250.62, 250.7, 250.72, 250.8, 250.81, 250.9
Aphasia	784.3
Cerebral Palsy	343.9
Hemiplegia/Hemiparesis	342.02, 342.8, 342.81, 342.9, 342.91
Multiple Sclerosis	340
Parkinson's	332, 332.1
Quadriplegia	344, 344.01, 344.046, 344.09
Emphysema/COPD	491.2, 491.21, 496
Pneumonia	136.3, 482.83, 483.7, 486, 507
Septicemia	038.9
Dehydrated	276.5, 276.51
Delusions	297, 297.1, 297.11
Fever	780.6
Hallucinations	780.1
Internal Bleeding	459

- **Statistical Methodology: Case Weight.** We assigned a case weight to each observation of the sample with valid, non-zero nursing time and either an MDS or RPT assessment (N=10,136). The weight was derived from the sample design of the STRIVE study, a three-stage cluster sample with stratification. The clusters selected at the three stages were participating states, participating facilities, and the nursing units within each participating facility. Stratification of facilities was applied within states. The five strata that were used were (1) hospital-based facilities, (2) facilities with high concentrations of ventilator/ respirator residents, (3) facilities with high concentrations of AIDS/HIV residents, (4) facilities with high concentrations of residents in Medicare Part-A stays, and (5) all other facilities. Facilities that were in the first four strata were deliberately over-sampled. Facilities within a stratum were selected with probability proportional to size, meaning that larger facilities had a higher probability of selection. In addition, the proportion of available facilities that were selected for the study varied greatly from state to state, again by design. For all of these reasons, facilities and residents in the sample did not have an equal probability of selection. Therefore, case weights had to be developed to allow the calculation of unbiased population estimates from the sample data. Case weights were computed in the standard approach, as the inverse of the probability of selection and scaled so that the sum of the case weights equaled the actual sample size. Although not used for most of the derivation analysis (see discussion in Section 3.3), STRIVE applied these weights whenever we wished to produce national estimates from the STRIVE analytic data set (for

example, in computation of case-mix indices that would be applied nationally – see Section 4.1).[7]

- **Source: Facility Descriptors.** We matched the sample facility identifiers with OSCAR data to obtain facility-level descriptors. The primary information used was the identification of a facility as hospital-based.

Additional details of many of these steps are included in the STRIVE Phase I Report, Appendix A-1 pp. 7.

3.3 Conceptual Approach to Analysis

The goals of this study were to develop a case-mix classification system usable for adjusting payment to US nursing facilities under the federal Medicare Skilled Nursing Facility Prospective Payment System, and, secondarily, to provide the basis for nursing facility payments either by US state-specific Medicaid systems or by other payers of nursing facility care. In the past, case-mix systems have been useful for other purposes than payment, including as a descriptor of a nursing facility, program, or population, as a risk adjustor, quality indicators, etc.

As with earlier RUG systems, a case-mix system should meet three types of criteria: statistical, clinical, and no negative incentives.[8] [9] Specifically:

- Statistical Criteria
  - These include the variance explanation of the cost measure WWST in this study, and the homogeneity of groups, as measured by their coefficient of variation.[10]
  - It is also important that the system identify residents with high resource use (even if rare). Thus, the analysis often focused on the criteria that would bring into a higher-cost RUG-IV group the most expensive residents, as measured by the WWST.
  - Prior case-mix systems based on measures of ADL alone achieved reasonable variance explanation, but were ineffective in identifying rare but very costly types of residents. One measure of this is the range of the cost measure from the highest- to the lowest-cost RUG group.
- Clinical Criteria

---

[7] In SAS, this is normally done by incorporating the following statement in statistical procedures (such as PROC FREQ, PROC MEANS, etc.): weight case_weight;

[8] Schneider et al., 1988, op cit. and Fries et al. 1994, op cit.

[9] Fries BE. Comparing case-mix systems for nursing home payment. *Health Care Financing Review* 1990; 11:103.

[10] Coefficient of variation for a group is its standard deviation divided by the group mean, all multiplied by 100.

- The clinical criteria are based on the face-validity of the system, i.e., does it make sense?
- Prior research has shown that just searching for the resident characteristics, that best explain cost, yields systems inferior, even in statistical properties (such as variance explanation), to those like RUGs, a system that utilizes an understanding of the ***types*** of residents seen. In addition, the inclusion of an item in a case-mix system that is counter-intuitive will be a detriment to its use.

- Negative Incentives
  - The final criteria to be evaluated are negative incentives. Potential uses of a case-mix system for payment, or other uses, demand care in the choice of the resident characteristics used to identify the groups. Inclusion of any characteristic implicitly provides incentive. The following example was given in describing RUG-III:

    > "Take for example the case of indwelling catheters. Those with catheters use more resources, but not primarily for catheter care. Rather, the catheter is an indicator of a sicker resident with a spectrum of care needs. Including catheters as a criterion for higher case mix (and payment) would provide inappropriate incentives to catheterize residents."[11]

    This problem is substantial regarding the use of service variables in a case-mix system. In the past, RUG systems have included indicators of specific, very costly services such as the use of ventilators and IV medications. Including this type of service variable in a case-mix system generally encourages the use of these services. Thus, such qualifiers must be considered carefully to ensure that they will not be provided due to either their high reimbursement level or that providing the service could result in serious consequences to the resident, or both.
  - Classifying therapy based on the ***need*** for rehabilitation would overpay facilities that didn't provide services and could underpay facilities that did.

Finally, it should be noted that the RUG systems are based on classification algorithms, utilizing the minimum number of criteria necessary. Other systems can be developed founded on index approaches (e.g., the result of statistical regression analysis) and using all characteristics to describe the case-mix of a resident. In contrast, RUGs employs categories that are described by a parsimonious set of variables, with different variables describing different types of residents. The relative cost of a group is therefore associated with the mean of the WWST for that group. We describe in Section 6 how WWSTs for RUG-IV groups are transformed into case-mix indexes (CMIs).

---

[11] Fries et al., 1994, op cit. (pg. 670).

The following chapters describe the development of the analytic database and the analyses performed to develop RUG-IV. The database development and all analyses were performed using SAS V9.1.3.[12]

The derivation analyses described were performed without the use of the case weights. The rare exceptions that used case weights have this noted in the analysis description. There were several reasons for this decision:

- Prior research has shown that the relationships between resident characteristics and staff time are relatively invariant to sample characteristics, as demonstrated by the multiple international validations of the RUG-III system (see, for example, Ljunggren *et al.*[13] and Brizioli *et al.*[14]). It is therefore unlikely that the groups derived would differ substantially from those found if sample weights had been used.
- We deliberately over-sampled certain types of residents (e.g., ventilator residents) so we would have a sufficient number of them in our sample to support our analyses. If we used sample weights, these groups would no longer be over-represented in our sample and we would not have enough weighted cases to support our analyses.
- If we had incorporated sample weights and the sample design into our analyses, it would have substantially complicated our analyses, probably requiring the use of specialized techniques and software, and at little likely gain.

While the derivation analyses did not need to use weights or account for the sample design, the data used to derive the CMIs did require their use. The CMIs are dependent upon the distribution of cases among RUG groups. Because we over-sampled heavier care residents, we felt that it was important to weight the data to produce group means that were nationally representative. We therefore used weights to produce CMIs.

It is important for the reader to understand that most of the analyses were interdependent, i.e., the results of one analysis could affect another. For example, any evaluation of the ADL index would be affected by who was classified into the rehabilitation category and the rehabilitation plus extensive category and vice-versa. Given the large number of analyses, in many cases it was impossible to redo multiple analyses after each decision was made; we did attempt to reanalyze when there was strong reason to believe that results would change. One implication, however, is that results are reported for the variant of RUG-IV that was current at that point in time. For each analysis, we note which variant of RUG-IV was used. The validation database confirmed our revised logic for the RUG IV.

---

[12] SAS Institute, Cary North Carolina, 2005.
[13] Ljunggren G, Fries BE, Winblad U. "International Validation and Reliability Testing of a Patient Classification System for Long-Term Care." *European J Geront. 1:372-383 (1992).*
[14] Brizioli E, et al. "Nursing Home Case-Mix Instruments: Validation of the RUG-III System in Italy." Aging Clin Exp Res 15(3):243-253 (2002)

One way to minimize the interaction among analyses was to begin the derivation of RUG-IV by working down the major categories, beginning with the Rehabilitation Plus Extensive and Rehabilitation categories, and then moving sequentially through the Extensive Services category, etc. As these categories in RUG-IV form a hierarchy, with the resident assigned to the first (that is, highest) category to which he or she qualifies, assignment of that category refines the list of those who will be analyzed for the next "lower" hierarchy category. Once the major categories were determined, we considered secondary and eventually tertiary splits within categories.

We also noted in the initial analyses that the RUG-III system performed quite well "out of the box." When analyzing which criteria should be used for assigning a RUG-IV category, we evaluated the changes caused by including, versus dropping, a criterion – an "in/out analysis." We evaluated potential new criteria as follows:

- For all individuals newly assigned to a category, we examined:
  - The category from which they came and whether this was beneficial (e.g., if they were originally above the mean WWST as compared to the new category)
  - How they affected the new category to which they were assigned (e.g., raised the mean WWST).
- A similar logic was performed to evaluate residents who move out of a category when we dropped a category criterion.

Finally, we had predetermined, with the assistance of CMS and the Project Technical Expert Panel, a limited number of concerns about the RUG-III system. These included:

- Pre-admission services – during the early part of a resident's stay, the "look-back" period for MDS 2.0 services could include a hospital stay. RUG-III classification used the presence of certain services, whether they were within the hospital stay or in the nursing facility. We thus performed analysis to determine which measure – pre-admission service only, post-admission service only, or pre- or post-admission service (as in RUG-III) was the most appropriate for classification.
- Concurrent therapy – therapy is often provided to two or more residents simultaneously. RUG-III is based on the amount of time a resident is in therapy, whether or not the therapist is providing service to others at the same time. The growth of concurrent therapy since the introduction of the RUG-III PPS made it important to evaluate the effect of using allocated therapy times in determining RUG-IV rehabilitation times.
- Special Units – With the input from the TEP, we decided to examine whether beneficiaries in special units for Alzheimer's Disease required increased resource utilization than patients with comparable medical status without Alzheimer's Disease. As well, an earlier study performed using the analytic data from the RUG-III development found higher staffing, despite being case-

mix adjusted, in special care units such as those for Alzheimer's Disease patients.[15] With several types of special units, not just for Alzheimer's Disease, represented in the current study, it was important to understand if this effect was replicated and how it might affect the analysis.

- Rehabilitation Plus Extensive category – the original RUG-III system was modified in 2006 to include an extra, very high reimbursement category, for those meeting the qualification for both the Rehabilitation and the Extensive categories. Since implementation, the prevalence of this category has increased. It was, thus, a goal of the current study to understand if this category remained justifiable.
- Tertiary Splits – all RUG-III categories, other than the Rehabilitation Plus Extensive, the Rehabilitation, and the Extensive categories, include tertiary splits (after splitting based on an ADL index). Given the small numbers of individuals involved in each split, there was concern whether these are appropriate constructs.
- New Assessment Items and Scales – the development of MDS 3.0 and other assessment instruments has led to new items and new scales (algorithms combining multiple assessment items) that could be usefully incorporated into a RUG system.
- Emerging Special Populations – over time the nursing facility population has changed, with new subpopulations emerging and increasingly being recognized. At issue, therefore, are whether the resource use of these subpopulations is adequately explained and whether their characteristics are useful in explaining resource use of the larger nursing facility population.

[15] Mehr DR, Fries BE. "Resource Use on Alzheimer Special Care Units" *Gerontologist* 35(2):179-184 (Apr.) 1995.

4 Results – Derivation of the RUG-IV System

The following sections describe the results of the analysis of the STM data and associated assessments collected in Phase I of the study.

4.1 Sampling Results

4.1.1 Introduction

An earlier paper[16] presented details of the STRIVE sampling methodology. This section of this document presents information about the sample that was obtained by the study. The following topics are covered:

- Development of sample weights.
- Effect of sample weights.
- Comparison of facilities that participated in the study with those that were excluded or refused to participate.
- Comparison of facility and resident characteristics in the sample with the population.
- Estimates of the precision of the sample.
- Counts of the number of observations obtained from various special populations.

4.1.2 Development of Sample Weights

As described in the sampling methodology paper, the sample design used for the STRIVE study was a three-stage cluster sample with stratification. The three types of clusters selected at the three stages were participating states, participating facilities, and the nursing units within each participating facility. Stratification of facilities was applied within states. Certain strata were deliberately over-sampled. In addition, the proportion of available facilities that were selected for the study varied greatly from state to state, again by design. As a result, all of the facilities and residents in the sample did not have an equal probability of selection. Because of this, sampling weights were developed to allow the calculation of unbiased population estimates from the sample data.

The sample weight for each resident in the sample was equal to the inverse of that resident's probability of selection for inclusion in the sample. The steps involved in calculating the weights were as follows:

---

[16] *STRIVE Sampling Methodology: Version 4*. October 5, 2007.

1. Determine each resident's probability of selection. This probability was the joint probability of three events and was determined by computing the product of three component probabilities:

   a. The probability that the resident's facility was included on an initial sampling list. The SAS SURVEYSELECT procedure was used to randomly select facilities, with probability proportional to size (assessment volume) from each facility within each of a state's strata. SAS reports the probability of selection for each selected facility, and these probabilities were used as the first of the three components.
   b. The probability that the resident's facility was selected from the initial sampling list for inclusion in the study. The initial sampling lists were an over-sample. In other words, more facilities were selected than were needed within each state's strata. Some facilities on the list were eliminated by state and regional office staff who reviewed the lists for facilities that were unable to participate in the study or that were known to have very poor quality. Furthermore, because participation in the study was voluntary, only some of the facilities that were invited to participate in the study actually joined it. For both of these reasons, only a subset of the facilities that were on the initial sampling list actually participated in the study. The probability of inclusion was computed by dividing the number of facilities that participated by the number of facilities that were on the initial list. Thus, if 60% of the facilities on the list participated in the study, then the probability of participation was considered to be 0.60.

      It was known that the participating facilities were not actually a random sample of those that were on the initial list and that the method of computing this probability was, therefore, not entirely accurate from a statistical perspective. However, practical considerations made it impossible to enforce random selection in this phase of the sampling process. We therefore assumed that the process approximated random selection and later tested this assumption by comparing sample characteristics with population characteristics, as is explained below.

   c. The probability that an individual resident was included in his or her facility's data collection. Once a facility was selected and agreed to participate, nursing units within the facility were selected for inclusion in the data collection process. In smaller facilities, all residents in the facility were included in the study. However, because of equipment and training considerations, we were not able to include all nursing units in the study within larger facilities. When only a subset of nursing units could be included, a standardized protocol was applied to select the units. While this process was standardized, it was not random. Like the previous step, we were unable to use random selection in this step. We therefore assumed that the process approximated random selection, even though this was not accurate from a statistical point of view. Using this approach, the probability of resident selection within a facility was equal to the proportion of facility residents that were included in the study. For example, if 75% of

a facility's residents were located on study units and were included in the data collection, then the probability of selection for each of those residents was equal to 0.75. Again, analysis comparing sample characteristics with the population were performed to test this assumption and are presented below.

2. The joint probability of selection was simply the product of the three components described above. If these three probabilities are designated a, b, and c, then the joint probability (j) is equal to:

   $j = a * b * c$

3. The raw sampling weight, r, is equal to the inverse of the joint probability:

   $r = 1 / j$

4. The sum of the raw sampling weights is equal to the number of residents in the population (the total number of residents in the 15 states that participated in the study). For most purposes, it is desirable for the weighted totals to equal the number of residents in the sample, rather than the population. To achieve this, the raw sample weights were multiplied by an appropriate scaling factor so that they would sum to the appropriate sample count. This scaling factor, s, is equal to the total number of residents in the sample (n) divided by the total number of residents in the population (N):

   $s = n / N$

5. Thus, the final sampling weight for each case in the sample, w, is equal to:

   $w = s * r$

The following example illustrates the steps involved in computing the sampling weights. This example uses actual data for a facility that was in the STRIVE sample, and references the notation defined above. Note that the calculations below may not match one another exactly due to rounding.

- **Step 1: Compute probability #1.** This value comes directly from the SAS SURVEYSELECT procedure and represents that probability of including the facility in the initial sampling list. SAS reported the probability for the example facility as 0.279289.

  $a = 0.279829$

- **Step 2: Compute probability #2.** This value represents the probability that the example facility was selected from the initial sampling list. There were 30 facilities in the facility's state that were in the facility's stratum, and three of these facilities were selected for inclusion in the study. Probability #2 is therefore 3/30, which equals 0.1000000.

  $b = 0.100000$

- **Step 3: Compute probability #3:** This value represents that probability that an individual resident was selected for inclusion in the study from within the example

facility. The example facility had 94 residents, and 55 of these residents were included in the sample. Probability #3 is therefore 55/94, which equals 0.585106.

c = 0.585106

- **Step 4: Compute the joint probability.** The joint probability is the product of the three probabilities described above: 0.279829 * 0.100000 * 0.585106 is equal to 0.016341.

  j = 0.016341

- **Step 5: Compute the raw sampling weight.** The raw sampling weight is the inverse of the joint probability: 1 / 0.016341 is equal to 61.19442.

  r = 61.19442

- **Step 6: Compute the scaling factor.** The scaling factor is equal to the number of residents in the entire sample divided by the number of residents in 15 states that were included in the study. This value is 10,136 / 495,384 = 0.020461.

  s = 0.020461

- **Step 7: Compute the final sampling weight.** The final sampling weight is the product of the raw sampling weight and the scaling factor: 61.19442 * 0.020461 is equal to 1.252093.

  w = 1.252093

This value, 1.252093, was used as the sampling weight for each case that was sampled from within the example facility.

4.1.3 Effect of Sample Weights

Table 4-1 shows the number of facilities and residents in the study sample in each of the 15 participating states. There were a total of 205 facilities and 10,136 residents in the sample.

Table 4-1. Number of Sample Facilities and Residents by State

State	Facilities	Residents
Dist of Columbia	9	395
Florida	4	206
Iowa	21	1194
Illinois	15	929
Kentucky	12	627
Louisiana	10	583
Michigan	5	282
Montana	9	395
Nevada	15	607
New York	21	1063
Ohio	20	860
South Dakota	18	715
Texas	14	744
Virginia	17	918
Washington	15	618
All States	205	10,136

Every facility in the population and in the sample was assigned to one of the following five strata:

- Hospital-based (HB) facilities. Residents in HB facilities typically have stays that are considerably shorter than residents in non-HB facilities. Furthermore, HB facilities typically have different staffing patterns and cost structures than non-HB facilities. For this reason, HB facilities were included as a stratum.
- Facilities with a high concentration of residents on ventilators/respirators (Hi-Vent). Residents who are on ventilators/respirators are known to be costly to care for and to require more intensive staff resources. Using MDS data, facilities were identified in which 12% or more of their residents were on ventilators/respirators. These facilities fell into the Hi-Vent stratum.
- Facilities with a high concentration of residents with HIV (Hi-HIV). Residents with HIV are known to be costly to care for and to require more intensive staff resources. Using MDS data, facilities were identified in which 10% or more of their residents had HIV. These facilities fell into the Hi-HIV stratum.
- Facilities with a high concentration of Medicare Part A residents (Hi-PartA). The STRIVE results and the resulting CMIs are relevant to both the Medicare

and Medicaid programs. However, because only a relatively small minority of residents (16.4%) nationally is served in SNFs under Medicare's Part A program, it was necessary to over-sample facilities that served such residents in order to obtain a sufficient number of residents for analysis. Using MDS and SNF Medicare claims data, facilities were identified in which 20% or more of their residents were in stays that were paid for under Medicare Part A. These facilities fell into the Hi-PartA stratum.

- All remaining facilities ("Other"). Facilities that did not qualify for any of the four strata described above fell into the "Other" stratum.

The strata were defined hierarchically in the order listed above so that they were mutually exclusive. Thus, if an eligible facility qualified for more than one stratum, it was classified into the first one in the list that it qualified for. Using this approach, every eligible facility in each state was classified into one of the five strata listed. Note that some strata were not represented in some states.

Table 4-2 shows the number of residents in the sample and in the population by stratum. The first pair of data columns in Table 4-2 show unweighted resident counts and percentages for the sample. The next pair of columns shows the sample counts after the case weights described above were applied. The remaining columns show the number and percent of residents that fell into each stratum for the 15 STRIVE states and for the entire nation. These counts are based upon all residents who were active (in a facility) on a given date (March 1, 2006).

Table 4-2. Number of Residents in the Sample and Population by Stratum

Stratum	Unweighted Sample		Weighted Sample		STRIVE States		National	
	Residents	Percent	Residents	Percent	Residents	Percent	Residents	Percent
Hosp-based	863	8.5%	418	4.1%	25,120	4.1%	66,639	4.7%
Hi-Vent	440	4.3%	205	2.0%	13,159	2.1%	19,785	1.4%
Hi-HIV	411	4.1%	39	0.4%	4,439	0.7%	6,010	0.4%
Hi-Part A	4,298	42.4%	1,648	16.3%	111,540	18.1%	233,875	16.4%
Other	4,124	40.7%	7,825	77.2%	460,900	74.9%	1,098,077	77.1%
Total	**10,136**	**100.0%**	**10,136**	**100.0%**	**615,158**	**100.0%**	**1,424,386**	**100.0%**

It can be seen that the sample weights dramatically affected the distribution of cases by stratum. For example, in the unweighted sample, 42.4% of the residents fell into the Hi-PartA stratum. However, after applying the sample weights, only 16.3% of the cases fell into this stratum. This shift in the distribution reflects the fact that some of the strata were heavily oversampled. When compared with the two population distributions, it can be seen that the sample weights were successful in producing a weighted distribution that resembled the population.

4.1.4 Comparison of Participating with Non-Participating Facilities

As noted above, several of the steps involved in the sampling process were not random processes. Due to practical constraints, we were forced to treat them from a sampling perspective as if they were random, even though they were not. In order to determine whether this approach was reasonable and whether biases may have been introduced by the non-random selection procedures, two sets of analyses were performed. The first analysis, which is reported in the current section of this report, compared facilities that participated in the study with those that refused to participate or were eliminated from the sample by state and regional office staff. The second set of analyses, which is reported in the next section of this report, compared sample characteristics with population characteristics on selected variables that were deemed to be of importance to the measurement of case mix.

State agencies perform surveys of every certified nursing facility in the nation approximately once per year. In the course of the survey, data are collected regarding the staffing levels of the nursing home. These data are stored in CMS's OSCAR database. CMS has developed procedures to identify and adjust outliers in the staffing data in order to improve its accuracy. The adjusted data are reported on a quarterly basis on CMS's Nursing Home Compare web site. Staffing levels are, of course, highly relevant to STRIVE which is aimed at measuring and predicting the use of staff resources. Since the OSCAR staffing data are available for every certified nursing home in the nation, these data provided an opportunity to determine whether the sampling procedures that were used by STRIVE introduced any biases with regard to staffing levels.

OSCAR staffing data were downloaded from the Nursing Home Compare web site on 12/11/2007. This data set contains staffing data collected on the last available regular survey for every certified facility nationally. Since surveys are performed every 9-18 months, the staffing data roughly coincide with the time period during which STRIVE field work was being performed.

Table 4-3 shows the mean minutes per resident day by staff type for the following groups of STRIVE nursing homes in the first 3 rows: (1) STRIVE nursing homes that were eliminated from consideration by State and Regional staff, (2) STRIVE nursing homes that were invited but declined to participate, and (3) STRIVE nursing homes that participated in the study. We also show three national groups of nursing homes: (4) all nursing homes nationally that passed the QI/QM and survey deficiency quality data screens, (5) all nursing homes nationally that failed the quality data screens, and (6) all nursing homes nationally.

Table 4-3. Mean Minutes Per Resident Day

		Mean Minutes Per Resident Day				
Row	**Group**	**Nursing Homes**	**RNs**	**LVNs**	**Aides**	**Total**
	STRIVE Nursing Homes					
1	Eliminated by states and regions	90	32.2	49.3	144.9	226.4
2	Declined to participate	287	37.4	46.8	136.5*	220.7*
3	Participated	198[1]	34.4	54.7	146.7	235.9
	National Nursing Homes					
4	Passed quality data screens	13,419	38.2	47.4	141.3	226.9
5	Excluded by quality data screens	1,149	38.1	51.8	138.6	228.6
6	All facilities	14,636	38.2	47.8	141.1	227.1

[1]There were 205 nursing homes that participated in the STRIVE study, but only 198 could be matched to OSCAR data.
*Asterisks indicate statistically significant differences ($p<0.05$) between the values in Rows 1, 2, or 3 compared with corresponding values in Row 4.

The proper basis for comparison between the STRIVE sample groups and the nation is Row 4: facilities that passed the quality data screens. As part of the design, we excluded about 8% of all nursing homes nationally from the sampling frame that had very poor QI, QM, or survey deficiency histories (Row 5). Since these nursing homes were not in the sampling frame, we would not necessarily expect the staffing levels of STRIVE nursing homes to match their staffing levels. Therefore, statistical comparisons were made between corresponding values in Rows 1, 2, and 3 and the values in Row 4. 95% confidence intervals were computed for the group means in the first three rows of the table (eliminated, declined, and STRIVE sample facilities). If the confidence interval did not include the population value for the STRIVE states, the group mean was considered to be significantly different ($p<0.05$) from the population value; asterisks indicate values that are significantly different from the values in Row 4.

The three groups of STRIVE nursing homes matched the national statistics in Row 4 fairly well. Nursing homes that declined to participate (Row 2) had significantly lower aide and total time, but the staff times for nursing homes that completed the study were not significantly different from the nation. We therefore conclude that the factors related to self-selection did not create a sample that was biased with regard to staff time. The total staff minutes for STRIVE nursing homes was greater than the national average total staff minutes which contributed to creating accurate CMIs.

Overall, these results suggest that two of the non-random processes that were used during sample selection, elimination of sampled facilities and refusal to participate, did not introduce any substantial bias in the sample with regard to staffing levels as measured by OSCAR data.

4.1.5 Comparison of Sample and Population Statistics

Another set of analyses was performed to check for possible biases in the STRIVE sample. These analyses involved using MDS data to compare the sample with population values on variables that are related in important ways to the measurement of case mix.

Table 4-4 compares STRIVE statistics for the entire sample with national MDS statistics. For these comparisons, a cross-section of MDS data was selected that contained the latest assessment for every resident who was active (i.e., in a nursing home) on a given date. March 1, 2006 was selected for this analysis so that the data would be as contemporaneous as possible with the STRIVE data. Variables important to case mix determination were selected for analysis. Chi-square tests were performed to determine whether the distribution of scores on each variable deviated significantly from the national distribution. The columns in the table below show the MDS variable, the number and percent of cases for each value of the variable for the nation and for STRIVE, and an indicator of whether or not the chi-square test showed the STRIVE distribution to be significantly different from the national distribution.

Table 4-4. Comparison of Sample with Population on Gender and Selected ADLs

		MDS National Snapshot		STRIVE: Sample Weighted		Signif Diff (p<0.05)
MDS Variable	**Value**	**Freq**	**Pcnt**	**Freq**	**Pcnt**	
G1aa (bed mobility self-performance)	0. Independent	393,296	28.4%	2,724	27.9%	Yes
	1. Supervision	86,778	6.3%	612	6.3%	
	2. Limited assist	241,342	17.4%	1,638	16.8%	
	3. Extens assist	438,795	31.7%	2,871	29.4%	
	4. Total depend	224,203	16.2%	1,918	19.6%	
	8. Did not occur	634	0.0%	2	0.0%	
	Total	1,385,048	100.0%	9,766	100.0%	
G1ba (transferring self-performance)	0. Independent	271,891	19.6%	1,600	16.4%	Yes
	1. Supervision	96,985	7.0%	602	6.2%	
	2. Limited assist	258,049	18.6%	1,946	19.9%	
	3. Extens assist	432,545	31.2%	3,115	31.9%	
	4. Total depend	313,808	22.7%	2,410	24.7%	
	8. Did not occur	11,817	0.9%	93	0.9%	
	Total	1,385,095	100.0%	9,766	100.0%	
G1ha (eating self-performance)	0. Independent	599,025	43.2%	3,556	36.4%	Yes
	1. Supervision	327,129	23.6%	2,448	25.1%	
	2. Limited assist	128,760	9.3%	1,046	10.7%	
	3. Extens assist	123,645	8.9%	1,019	10.4%	
	4. Total depend	206,050	14.9%	1,696	17.4%	
	8. Did not occur	478	0.0%	1	0.0%	
	Total	1,385,087	100.0%	9,766	100.0%	

		MDS National Snapshot		STRIVE: Sample Weighted		Signif Diff (p<0.05)
MDS Variable	**Value**	**Freq**	**Pcnt**	**Freq**	**Pcnt**	
G1ia (toileting self-performance)	0. Independent	206,103	14.9%	1,048	10.7%	Yes
	1. Supervision	79,396	5.7%	450	4.6%	
	2. Limited assist	215,647	15.6%	1,548	15.9%	
	3. Extens assist	451,917	32.6%	3,338	34.2%	
	4. Total depend	427,881	30.9%	3,181	32.6%	
	8. Did not occur	4,154	0.3%	200	2.1%	
	Total	1,385,098	100.0%	9,766	100.0%	
Verbal/physical abuse	No	1,373,940	99.2%	9,737	99.3%	No
	Yes	11,173	0.8%	66	0.7%	
	Total	1,385,113	100.0%	9,802	100.0%	
K5a (parenteral/IV)	No	1,343,588	98.3%	9,634	98.3%	No
	Yes	22,972	1.7%	163	1.7%	
	Total	1,366,560	100.0%	9,798	100.0%	
K5b (feeding tube)	No	1,295,170	93.7%	9,036	92.2%	Yes
	Yes	87,738	6.3%	762	7.8%	
	Total	1,382,908	100.0%	9,798	100.0%	
P1ac (IV medication)	No	1,255,886	91.7%	9,138	93.3%	Yes
	Yes	113,052	8.3%	661	6.7%	
	Total	1,368,938	100.0%	9,799	100.0%	
P1ag (oxygen therapy)	No	1,198,577	87.6%	8,656	88.3%	Yes
	Yes	170,392	12.4%	1,143	11.7%	
	Total	1,368,969	100.0%	9,799	100.0%	
P1ai (suctioning)	No	1,354,628	99.0%	9,595	97.9%	Yes
	Yes	14,356	1.0%	203	2.1%	
	Total	1,368,984	100.0%	9,799	100.0%	
P1aj (tracheostomy care)	No	1,355,834	99.0%	9,618	98.2%	Yes
	Yes	13,150	1.0%	181	1.8%	
	Total	1,368,984	100.0%	9,799	100.0%	
I1a (diabetes mellitus)	No	971,074	71.0%	6,824	69.6%	Yes
	Yes	397,044	29.0%	2,975	30.4%	
	Total	1,368,118	100.0%	9,799	100.0%	
I1v (hemiplegia/ hemiparesis)	No	1,231,378	90.0%	8,807	89.9%	No
	Yes	137,410	10.0%	993	10.1%	
	Total	1,368,788	100.0%	9,799	100.0%	
I1z (quadriplegia)	No	1,358,262	99.2%	9,722	99.2%	No
	Yes	10,531	0.8%	77	0.8%	
	Total	1,368,793	100.0%	9,799	100.0%	
M2a (stage 3 or 4 pressure ulcer)	No	1,346,209	97.2%	9,419	96.4%	Yes
	Yes	38,827	2.8%	348	3.6%	
	Total	1,385,036	100.0%	9,767	100.0%	

While several of the analyzed variables showed no significant differences, there were significant differences between the sample and the nation on a number of other variables. On the ADLs (viz. G1aa [bed mobility self-performance], G1ba [transferring self-performance], G1ha [eating self-performance], and G1ia [toileting self-performance]), there was a consistent trend for residents in the sample to show slightly more dependence than residents nationally. On each of these ADLs, the percent of

STRIVE cases in the “total dependence” category exceeded the national percentage by between 1.7 and 3.4 percentage points. Conversely, the percent of residents in the “independent” category was lower for the STRIVE sample by between 0.5 and 6.8 percentage points. The picture was mixed on the services items that displayed significant differences. Among these items, the STRIVE residents were slightly more likely to receive feeding tubes, suctioning, and tracheostomy care, but less likely to receive IV medications or oxygen therapy. Slightly more STRIVE residents had diabetes mellitus and Stage 3 or 4 pressure ulcers than was seen nationally.

The overall picture from these comparisons is that the STRIVE sample has somewhat higher acuity than the nation. This could have been due to the last stage in the sample selection process where nursing units within larger nursing homes were selected for inclusion in the study. In selecting units for inclusion, the protocol used by data monitors tended to favor SNF units and other specialty units that likely had higher acuity. It is possible that a greater proportion of higher acuity residents were included in the sample and that the sample weights did not correct for this.

The impact of this bias should be small, however. First, while the differences displayed above were statistically significant, due to the large sample sizes involved, they were not substantial. Second, RUG classification models are designed specifically to classify residents into groups with similar acuity levels. For example, ADL scores are used explicitly to subdivide residents falling into each of the major hierarchical groups. Thus the small amount of bias that is evident did not affect the development of classification models.

In theory, bias could affect the development of case mix indices because the CMIs are influenced by the distribution of cases across groups. However, in practice any influence should be eliminated because the CMIs are standardized to match observed national days of service distributions. Thus, while the sample may contain a small amount of bias that might place slightly more residents into heavier care nursing groups, this bias should be corrected when computing CMIs that are standardized to match the national days of service distribution.

4.1.6 Precision of the Sample

As discussed in the sampling methodology paper, the STRIVE sample design had two goals: (a) to obtain a sample that could be generalized to the national population without bias and with sufficient precision, and (b) to obtain enough cases from certain important special populations of residents to yield sufficient statistical power to support case mix analyses. The current section of the paper presents estimates of the precision of the sample, which is relevant to the first goal. The following section of the paper presents data regarding special populations, which is relevant to the second goal.

Table 4-5 presents the margins of error for the STRIVE study as well as the margin of error for the 1995/1997 time study, for comparison.

Table 4-5. Projected Margin of Error for Mean Staff Time

Parameter	STRIVE*	1995/97 Time Study	Percent Improvement
Nursing Time			
Number of cases	9,766	3,933	
Mean wage-weighted time	135.2	228.3	
Standard error of mean	3.1	7.2	
Margin of error (95% confidence interval)	±6.2	±14.3	
Margin of error (percent of mean)	±4.6%	±6.2%	25.8%
Therapy Time			
Number of cases	1,510	1,133	
Mean wage-weighted time	144.0	86.0	
Standard error of mean	5.5	3.5	
Margin of error (95% confidence interval)	±10.9	±6.9	
Margin of error (percent of mean)	±7.6%	±8.0%	5.0%

*Note: The STRIVE results reported in this table were computed before data cleaning was final. The number of cases reported for STRIVE in this table are therefore slightly higher than the number used for later analyses.

For each of these studies, Table 4-5 presents statistics for mean nursing time (based upon all residents in the sample) and for therapy time (based upon all residents who received any therapy time). For each of these datasets, the table presents the number of cases (raw, unweighted counts), the mean of the wage-weighted minutes, the standard error of the mean, the margin of error associated with the mean (at the 95% confidence interval), and the margin of error expressed as a percentage of the mean.

Note that for both nursing and therapy time, the methodology used to wage weight time differed between the two studies. Therefore, the means, standard errors, and margins of error cannot be directly compared between the two studies. We have, therefore, computed the margin of error as a percentage of the mean to allow such comparison.

It can be seen that in the STRIVE sample the margin of error for the nursing time is about ±4.6% of mean nursing time, compared with ±6.2% in the earlier study. This represents a 25.8% improvement in precision over the earlier study. For therapy time, the STRIVE margin of error is ±7.6%, a 5.0% improvement over the earlier study. With regard to therapy time, the improvement is modest because of the relatively large number of cases in the 1995/97 time study that had therapy time. We believe this is due to the fact that the convenience sample that was used for the earlier study was largely aimed at identifying and enlisting nursing homes that had Medicare residents and provided therapy. Thus, the STRIVE sample is larger and has considerably more precision for nursing time than the previous time study.

4.1.7 Special Populations

As described above, the sampling plan was designed to target certain special populations that were deemed to be important for case mix analysis but are relatively rare in the population. Based upon a review of the literature and discussions with CMS and with the project's technical expert panel, a list of important special populations was compiled. MDS and other data were then analyzed to determine which of these groups could be successfully targeted. Our goal was to obtain at least 50-100 residents in each of these groups in order to support planned case mix analyses. Table 4-6 presents counts of the number of observations in each of these groups that were obtained.

Table 4-6. Number of Sample Residents in Special Populations

Special Population	Number of Observations	Special Population	Number of Observations
Ventilators	254	Respiratory therapy	329
HIV*	253	Suctioning	362
Alzheimer's	1,313	Physical/verbal abuse	59
SMI*-All	3,003	Burns	30
SMI*-Schizophrenia	679	Chemotherapy	49
SMI*-Bipolar	279	Traumatic brain injury	112
SMI*-Other	2,045	Surgical wound care	653
Deaf/blind	169	Dialysis	176
Deaf	935	Hospice	282
Blind	870	RUG44: BA/BB* group	73
Under age 18 years	5	Bariatric (weight>=300 lbs)	72
19-40 years old	133	Palliative care	370
41-64 years old	1,324	Pain	3,805

*Note: SMI = Severe Mental Illness; HIV=Human Immunodeficiency Virus; RUG44 BA/BB = RUG-III, 44-group system, behavior problem groups

It can be seen that only three groups fell below the goal of having 50-100 residents. There were only 5 residents under 18 years of age. This is not surprising because pediatric facilities were specifically excluded from the STRIVE sample and no attempt was made to sample enough pediatric residents for special analysis. There were only 30 residents with burns. Preliminary analysis indicated that these residents are exceedingly rare and that, because they do not tend to cluster in particular facilities, they are not amenable to targeted sampling. The chemotherapy group fell just below the goal with 49 residents. Again, they are very rare in the population and tend not to be concentrated in specialty facilities. Based upon these results, we concluded that we met the special population goals that we had set for all groups where this was feasible.

4.2 Overview of RUG-IV Systems

In the remainder of this Section 4, we describe the analyses performed that eventually led to the derivation of the RUG-IV system. As described earlier, many of these analyses were overlapping or sequential. As a result, one analysis would be based on a RUG-IV version and would inform the development of the next version. Table 4-7 describes the major changes made as we developed RUG-IV (starting with RUG-III) and can be useful for any reader wishing to reproduce individual results. The final version of RUG-IV is Version 9. Later sections of this report describe the analyses performed to develop each of the changes listed.

Table 4-7. Versions in the Development of RUG-IV

RUG-IV Version	Rehabilitation Plus Extensive (RE)	Rehabilitation (R)	Extensive Services (E)	Special Care High (H)	Special Care Low (L)	Clinically Complex (C)	Impaired / Behavior (I)	Other
Version 1	-Post-admit extensive services (tracheostomy, ventilator/respirator, suctioning, IV medications)		-Post-admit extensive services (tracheostomy, ventilator/respirator, suctioning, IV medications)					
Version 2	-Drop IV medications		-Drop IV medications -Drop count of extensive services to break out subgroups					
Version 3	-Drop IV/parenteral feeding		-Drop IV/parenteral feeding	-Add septicemia and infection isolation to Special Care				
Version 4	-Changed to allocated therapy time -Drop suctioning -Removed ADL category restriction	-Changed to allocated therapy time -Removed ADL category restriction	-Changed criterion for subgroups -Drop suctioning -Add infection isolation -Removed ADL category restriction	-Split Special Care into two groups - New qualifiers -Removed ADL category restriction	-Split Special Care into two groups - New qualifiers -Add IV/parenteral feeding to Special Care -Removed ADL category restriction	- New qualifiers	-New ADL-IV Index threshold	
Version 5				-New ADL category threshold -Added PHQ depression tertiary splits	-New ADL category threshold -Added PHQ depression tertiary splits	-Automatic qualification for ADL=0,1 from S and L categories -Added PHQ depression tertiary splits	-Combined RUG-III Impaired Cognition and Behavior categories	
Version 6				-Returned to mood depression tertiary splits	-Returned to mood depression tertiary splits			-Consistent ADL Splits using RUG-IV ADL (ADL-IV) Index
Version 7				-Relabeled groups	-Relabeled groups	-Created CA group from combined CA/CB group - Tertiary splits		-Dropped one high outlier observation

RUG-IV Version	Rehabilitation Plus Extensive (RE)	Rehabilitation (R)	Extensive Services (E)	Special Care High (H)	Special Care Low (L)	Clinically Complex (C)	Impaired / Behavior (I)	Other
Version 8	-Changed to time-slice therapy allocation	-Changed to time-slice therapy allocation		-Altered fever criterion by dropping dehydration qualifier	-Altered ulcer criterion to include stasis ulcer -Added oxygen therapy with respiratory failure -Dropped oxygen therapy without respiratory failure	-Added oxygen therapy without respiratory failure		
Version 9 (Final)				-Altered fever criterion by adding qualifier of tube feeding (with intake restriction)				

4.3 Summarizing the Staff Time Measurement (STM)

In this section, we provide a brief description of the staff time data collection (further detail can be found in the STRIVE Phase I report[17]) and how we cleaned these data. We provide basic descriptive statistics on the facility staff daily time spent caring for residents in the sampled facilities.

4.3.1 Allocating Parallel Care Time

During the Staff Time Measurement, there were instances when a staff member was caring for multiple residents simultaneously. While not common practice for nursing staff, it was common for therapists. In the case of therapy, this parallel time was classified either as concurrent therapy or group therapy. When a therapist cares for multiple residents at the same time but is performing different tasks for each, it is denoted "concurrent therapy"; when the therapist does the same task on all persons (e.g., a front- and side-reach balancing activity), it is denoted "group therapy." Concurrent therapy was identified in our data collection using two criteria: 1) time records were part of a "group" (i.e., a therapist was treating more than one resident at the same time), and 2) the therapy was denoted "concurrent" by different HCPCS codes (provided by the therapist and entered into the PDA) for the different residents in a "group". Group therapy was identified in our data collection using two criteria: 1) time records were part of a "group" (i.e., a therapist was treating more than one resident at the same time), and 2) the therapy was denoted "group" by the same HCPCS code (provided by the therapist and entered into the PDA) for the different residents in a "group". When data were collected on paper, HCPCS codes were not entered by therapists, making it impossible to distinguish between concurrent and group time. Thus, for the paper time collection, when more than one resident received care the time was denoted as "Group/Concurrent" in the STRIVE dataset.

In all cases of parallel care time, we determined the amount of staff time attributed to each resident; this was the basis on which to determine the relative cost of care. There were 2 methods for allocating parallel time: Proportional and Time Slice. These methods are illustrated in Table 4-8.

Proportional: This method sums the time across residents and then determines proportion of the total time for all residents in therapy that each resident had. This proportion is then applied to the therapist's time to obtain the allocated time.

Time Slice: This method "slices" the session by when residents entered/left the "group." Thus, in each time slice, the time is allocated to a resident according to how many residents are in that "slice."

The two methods give slightly different results. For instance, consider the scenario displayed in Table 4-8. The therapist begins a half-hour session with Resident

[17] STRIVE Phase I report, pp. 15 and 17

A, who stays the full 30 minutes. Resident B joins 10 minutes later, and also stays until the end of the session (i.e., 20 minutes). Resident C only attends for the last 10 minutes of the session. Overall, the therapist spends 30 minutes with 3 residents, who accumulate a total of 60 minutes of "resident" time. Under the proportional method, we calculate Resident A's time is 30 minutes out of the 60, or 50% of the time. This results in 15 minutes (50% of the total 30 minutes) allocated to Resident A. By a similar calculation, Resident B's and C's share are 33.3% and 16.7%, respectively, and their allocated times are 10 and 5 minutes, respectively.

Under the time-slice methodology, we identify three slices defined by the set of resident in each slice– the first, second, and third 10-minute periods. As shown in Table 4-8, Resident A is alone during the first slice, so that resident gets all 10 minutes.. In the next slice, Residents A and B share the 10 minutes, so each gets half the time, or 5 minutes each. In the final slice, all three residents are present, each getting a third of the 10 minutes, or 3.3 minutes each. In total, Resident A gets 18.3 minutes, Resident B 8.3 minutes, and Resident C 3.3 minutes.

It should be noted that the two methodologies give the same answer when all persons are in the group for the full period of time.

Prior RUG studies used the Proportional method. However, it was determined that the Time-Slice method is more accurate, and it was used in this study to allocate all group and concurrent time for therapists for analysis using RUG-IV Version 8 and 9. It should noted that very few nursing group times were recorded and most of the time all residents were in the group for the full time, thus the change to the time-slice methodology caused very little difference in the nursing time results.

Table 4-8. Hypothetical Scenario: Proportional vs. Time Slice Allocation Methodologies

Resident	Resident Time in Therapy	Proportional: Proportion of Resident Time	Proportional: Allocated Time	Time Slice 1	Time: Slice 2	Time: Slice 3	Time Slice: Total
A	30	50.0%	15.0	10.0	5.0	3.3	18.3
B	20	33.3%	10.0	0.0	5.0	3.3	8.3
C	10	16.7%	5.0	0.0	0.0	3.3	3.3
Total	60	100.0%	30.0				30.0
Session length	30			10.0	10.0	10.0	

4.3.2 Non-Resident Specific Time

The development of RUG-III case-mix indexes (CMIs) for the Skilled Nursing Facility Prospective Payment System incorporated both Resident-Specific Time and Non-Resident-Specific Time. In contrast, the work in this project has been based only

upon Resident-Specific Time. We discuss here the background and rationale for this change in methodology.

The STRIVE time study's approach to collecting time information about nursing staff (i.e., nurses and aides) was to assure that all shift time on the study unit was recorded for any staff member who spent any time delivering direct patient care during the study period. Time spent caring for the resident was classified as "Resident Specific Time" (RST). Both direct care (feeding, helping dress, giving medications) and "indirect care" (e.g., charting for that resident, calling a physician about the resident, etc.) are all part of RST. However, during the work shift nursing staff expend additional time on unit activities that are not directly attributable to any specific resident but that aid residents in general. This time was denoted "Non-Resident Specific Time" (NRST). NRST includes time spent on tasks such as stocking the medication cabinet, cleaning up a spill in the hallway, participating in training sessions, completing general charting (not associated with specific residents), performing unit administration, taking time for paid meals and breaks, etc.

Together, RST plus NRST represent all the time spent by direct-care staff on a study unit's activities. In some STRIVE facilities the study was restricted to a subset of the units in the nursing facility. Because some nurses in these facilities worked on multiple units, including non-study units, STRIVE also collected "Non-Study Time" (NST), which included staff time spent away from the study unit(s). The PDAs that were used to collect the time information required all the time of each staff member to be allocated either as RST for a specific resident, NRST (differentiating meals and breaks from all other NRST), or NST. Collecting all of a staff member's time during the work day, even if some of it was irrelevant for the study, served as a quality control check to make sure that staff recorded their time accurately. Staff members who did not care for residents on a study unit during the days that staff time was measured (e.g., the Director of Nursing) were not involved in the study and did not record their times.

For the purpose of developing the RUG-IV system, only RST was considered. The daily cost of a resident's care was computed by multiplying the RST spent by all persons in each staff role (e.g., nurses, licensed practical nurses, certified nurse assistants, medication aides, etc.) by the relative wage rate for that staff role (the wage rate standardized to a ratio of the wage rate of a certified nursing assistants). The resulting measure was called Wage-Weighted Staff Time (WWST). WWST was the dependent variable that was used for deriving the RUG-IV system and as the basis for computing the CMIs. Neither NST nor NRST was used in these computations.

This was a departure from the previous time study used to derive RUG-III. In that earlier study, the dependent variable was constructed from a combination of RST and NRST. In particular, the dependent variable used to derive RUG-III was constructed using the following formula:

[staff time] = [RST] + [meals and breaks: equal] + [other NRST: proportional]

Thus, staff time included three components:

- RST, as described above,

- Meals and breaks time equally divided across all residents in the study unit. For example, if a staff member spent 60 minutes on meals and breaks and there were 10 residents on her unit, then each resident was allocated 6 minutes of meals and breaks time from that staff member. Because meal and break time is totally unrelated to resident care, it was allocated in equal portions to all relevant residents.
- Other NRST time (excluding meals and breaks) allocated in proportion to residents' RST. In the earlier study, the assumption was made that a larger proportion of this time should be attributed to heavier care residents (who received more RST) than to lighter care residents (who received less RST). Thus, if a staff member spent 60 minutes on NRST and there were 10 residents on her unit, the 60 minutes would not be distributed equally to the 10 residents. Instead, each resident's RST across all staff members was computed first, and the 60 minutes was then distributed across the 10 residents in proportion to their RST, so that residents with higher levels of RST got more of the 60 minutes. The rationale for this allocation method was that heavier care residents might be more often responsible for the NRST time (e.g., training sessions for medical conditions requiring higher-level care) or might benefit more from this time (e.g., might utilize more of the medications that were stocked in the medication cabinet). While it was recognized in the earlier study that this assumption may not have been totally accurate in many cases, it was felt that proportional allocation of this NRST was more accurate than equal allocation.

Before discussing the rationale for STRIVE's use of only RST, it is important to understand how the RUG-IV system and its associated CMIs may be used. For Medicare, RUG-IV is used only to allocate a budgeted amount of Medicare dollars; it is not used to determine the ***size*** of that budget. Thus, the decision regarding whether or not NRST (or NST) are used in the STRIVE calculation does not affect the total amount that the federal government provides to skilled nursing facilities (SNFs), only how to allocate the payment to recognize the different cost of caring for different types of residents. Colloquially, it is not about the overall size of the "pie," only how it is sliced. The nursing and therapy rate components do include the cost of non-resident specific and administrative time.

It should also be mentioned that there is little rationale for the use of Non-Study Time (NST) in the STRIVE calculation, as these minutes are not in any way benefiting the residents on the study units. NST was never used in a previous RUG study. As noted above, the sole purpose of collecting NST was for quality control purposes, not for analytic purposes. However, NRST measurements are important for projecting staffing needs. The nursing component of the rate includes all nursing time as captured in the 1995 cost report.

Thus, the primary decision regarding the definition of the dependent variable for STRIVE's development analyses is determining whether or not to include NRST, and secondarily, if it is used, how it should be allocated (equally or proportionally).

There are six major reasons for omitting NRST from the dependent variable:

NRST Measurement Problems: The concept behind the measurement and use of NRST is that it captures staff time that indirectly benefits multiple residents. The problem is that NRST is only a partial reflection of the staff time that indirectly benefits multiple residents.

During STRIVE we only collected time for staff members who were involved in the direct care of residents on the study units. However, a facility typically employs additional staff members involved in administration or support whose time benefits multiple residents but who never directly care for any of those residents. A director of nursing, for example, performs administrative functions that benefit all residents in the nursing facility, but her time was not included in STRIVE if she did not provide any direct care on the study units during the study period. Also, no time was captured when staff from non-study units provided NRST-type activities (such as stocking a medicine cabinet) for the study unit.

Difficulty Allocating NRST Appropriately: Even if all NRST time were captured, it cannot be precisely attributed (allocated) to individual study residents. The following illustrate examples of the allocation problems involved:

- How should the time spent stocking a medication cabinet be allocated? Suppose the medication cabinet is shared by both a study unit and a non-study unit. What proportion of the nurse's time should be allocated to the study unit residents? (In STRIVE, all of that nurse's time would have been allocated to the study unit residents even though some of her time benefited residents in the non-study unit.)
- Even if the unit staff stock a medicine cabinet used only by that unit's residents, does this time benefit all study unit residents equally or should the time be allocated in proportion to their RST? Perhaps the fairest way to allocate this time would be in proportion to the number of medications that each resident uses. In theory, this could be done, but this approach might require devising different allocation strategies for different staff activities and different staff roles. There is no basis for determining these allocations with the data collected by STRIVE.
- Suppose a staff training program dealt with issues related to specific types of residents or specific types of problems. What is the most accurate way to allocate that time?
- Does time spent cleaning up a hallway spill benefit all residents equally? If the spill is directly outside a resident's room, it primarily benefits only that individual resident. Even if the spill is in a more trafficked area, the cleanup may not benefit residents who are not ambulatory.

It should be obvious from these examples that NRST cannot be precisely allocated without capturing a tremendous amount of detailed additional information about staff activities and without devising allocation rules for each

type of activity; this was beyond the scope of STRIVE. For this reason prior studies opted simply – but not fully accurately – to allocate this NRST equally or proportionally to study residents.

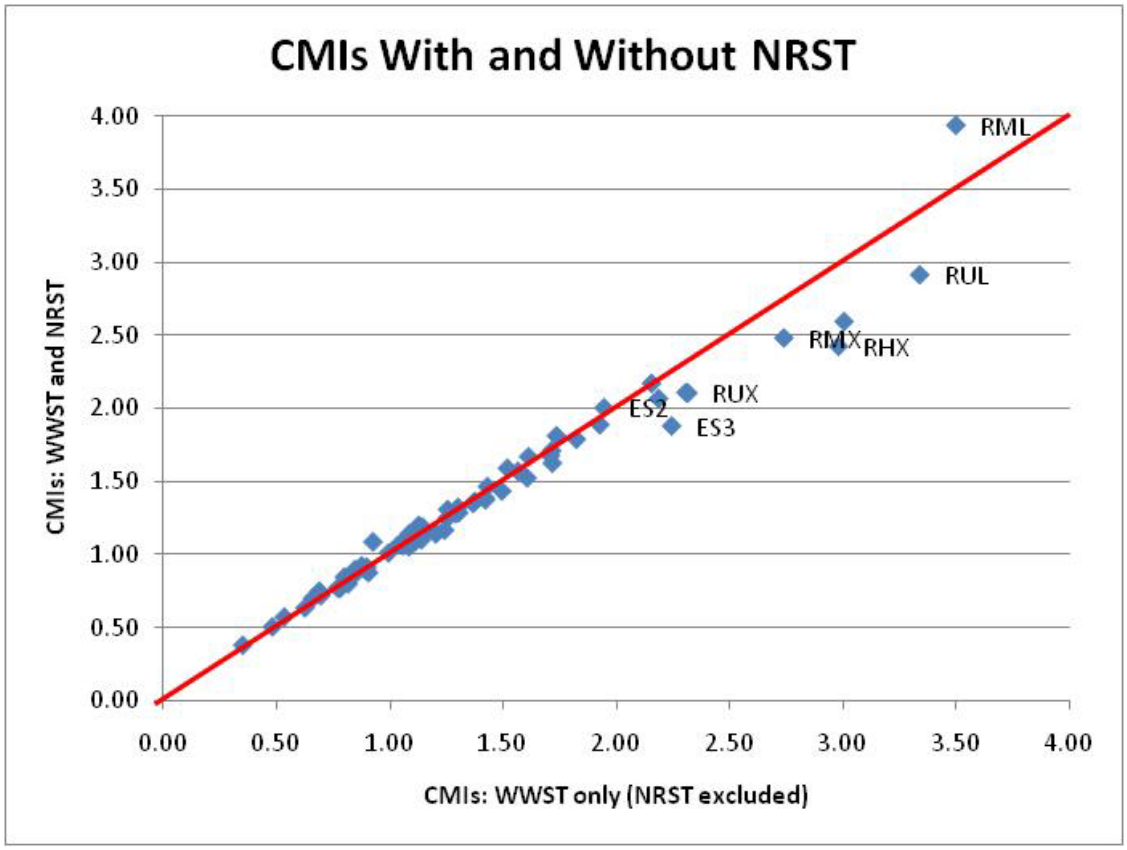
Introducing "Noise" into the Dependent Variable: As NRST comprises a large amount of time, any allocation of NRST runs the risk of introducing a considerable amount of "noise" into the dependent variable. For the STRIVE study, NRST amounted to 45.7% of the total time (RST and NRST). Of this 45.7%, only 4.0% is meals and breaks time (which is allocated equally among residents), and the remaining 41.7% is other NRST that is allocated proportionally. While it is probably appropriate to allocate some portion of this 41.7% proportionally, we do not know either what the correct portion should be or how properly to allocate the remainder. Resorting to proportional allocation (or any other method based solely upon simplifying assumptions) may introduce unwanted noise into the dependent variable without increasing its accuracy in measuring cost.

The result of adding such noise is that it makes it more difficult to derive a classification system that is sensitive to relationships between resident characteristics and direct care time. Some of the weaker relationships that exist in the data may be "washed out" by the noise that is introduced in the dependent variable. As a result, certain resident characteristics that, in reality, are related to direct care time may not be detected in the analyses that are performed. This could result in a classification system that fails to include important resident characteristics in the clinical hierarchy or that misplaces them in a less-than-optimal way.

Including NRST Attenuates or Compresses CMIs: Theoretically, including NRST in the dependent variable adds (to each resident's wage-weighted time an amount that cannot be explained by the resident's characteristics. As a result, the differences among the RUG group means and their CMIs are attenuated. For example, say the ratio of two residents' RST costs is [40 minutes]:[20 minutes] = 2:1. If we add a constant NRST to each, say of 10 minutes, the ratio changes to 50:30 = 1.66:1. Alternately stated, the range of the RUG CMIs is decreased. This makes residents in higher-cost RUG groups somewhat less attractive to facilities and (at least slightly) increases the incentives to admit and retain residents in lower-cost RUG groups. Adding NRST also reduces the overall variance explanation and runs counter to our goal of identifying the highest cost residents in groups.

After the derivation of RUG-IV was completed (see Section 5), we tested this hypothesis, computing the CMIs for RUG-IV using two dependent WWST variables: based on RST only, and based on WWST with NRST added (using the allocation approach described previously). The two sets of CMIs are compared in Figure 4-1.

Figure 4-1. CMIs With and Without NRST



The horizontal axis of the graph in Figure 4-1 shows the CMI that was produced using RST only (i.e., without including NRST in the dependent variable). The vertical axis shows the CMI that was produced with RST and NRST in the dependent variable. Each data point on the graph represents a RUG-IV group and shows the CMI using the two versions of the dependent variable. The diagonal line represents points where the two CMIs are equal. Data points falling below the diagonal line have a lower CMI when NRST is included, while points above the diagonal line have a higher CMI when NRST is included. Points falling on the line have the same CMI under both versions of the dependent variable.

Most groups fall quite close to the diagonal line, meaning that including or excluding NRST does not make much difference. However, there is a clear attenuation among almost all groups with higher CMIs when NRST is included (the exception being group RML which is above the line). All of the seven labeled groups have small sample sizes (less than 15 cases after sample weights are applied), except for ES2 and ES3 which have 57 and 102 cases, respectively. A careful examination of the graph shows the necessary slight tendency for the very lowest groups to fall above the diagonal line.

As anticipated, there is an attenuation or compression effect. Interestingly, this effect is more pronounced for the higher-paying groups, which would have received lower CMIs (and rates) if NRST is included in the dependent variable. There is a small countervailing tendency for the CMIs for the lowest groups to be higher if NRST is included. However, these groups are below the presumptive-coverage cutoff for Medicare and are rarely used in the SNF population. Thus, the major effect in the Medicare population if NRST is included is to bring down CMIs and rates for the highest-paying groups. These groups are much more homogeneous under RUG-IV than they are under the existing RUG-III model, and this effect might result in underpayments for residents in these groups and resulting access problems.

Compression Causes Inversions: One of the implications of the compressed range of RUG CMIs discussed above is an increased tendency for the mean WWST for RUG groups to "invert," a situation where a resident with one additional relevant condition (e.g., rehabilitation therapy) is classified into a higher hierarchy group with *lower* WWST which would result in both a lower CMI and payment. Instances of such inversion in the current RUG-IV model were addressed by a limited amount of "smoothing" (see Section 6.2). NRST inclusion in the calculations generally resulted in more inversions.

Producing the Best Measure of Cost: In the ideal RUG-IV model, the dependent variable would capture 100% of the costs associated with the staffing rate component. From this perspective, it might appear that including rather than excluding NRST would be better. As noted earlier, NRST was a large proportion (45.7%) of total nursing staff time and cost and should be examined and perhaps factored into the cost calculations. In the 1995 and 1997 time studies that lead to the RUG-53, NRST was allocated at the resident level (i.e., by adding it to the dependent variable). Most of this time was allocated proportionally. If every study unit in the sample had the same proportion of NRST time and if all NRST time (including meals and breaks) was allocated in proportion to RST, then there is no effect on the CMIs. Effectively, each resident's RST would be multiplied by a constant which, mathematically, would not affect either the derivation analyses or the calculation of CMIs and rates. This would also be mathematically equivalent to simply using RST in all calculations and using the resulting CMIs to adjust the staffing rate component, even though NRST was never directly considered.

However, we know that the proportion of total time that can be categorized as NRST must vary by unit and facility. Facilities that are heavily staffed, relative to their case mix, probably have higher proportions of NRST than those that are less adequately staffed. These differences can be referred to as "facility effects" because they are facility-specific (or unit-specific) cost differences. These facility effects may be due to regional factors, licensing or staffing requirements, ownership type, financial factors, management approaches, or other variables that are unrelated to case mix.

There are two ways in which NRST costs can be handled. They can be allocated at the resident level so that they are considered in model development and in producing

CMIs. Alternatively, they can be allocated when rates are calculated after the model and CMIs are produced. If the staffing component of the rate is adjusted by CMIs derived from RST only, this is equivalent to assuming that NRST costs are proportional to RST costs. This assumption is actually the same assumption that is made if NRST is allocated proportionally at the resident level. However, as will be shown, it has the advantage of not magnifying unwanted facility effects.

In previous studies most NRST was allocated proportionally with little justification. If facility effects did not exist, this would be equivalent to allocating NRST in the rate in the same proportion as RST costs, which, in turn, is equivalent to simply allocating the rate on the basis of RST alone. However, because facility effects do exist, resident-level proportional allocation serves to magnify facility effects.

Ideally, we prefer the case-mix model to be independent of facility effects. If two residents with exactly the same conditions are treated at two different facilities, relatively expensively at one facility and relatively inexpensively at a second, we would like to remove these differences from the data (i.e., these facility differences constitute another type of "noise"). In fact, in the Inpatient Rehabilitation Facility PPS model development process, a procedure is used that is intended specifically to remove such cost differences that are extraneous to case mix. Using resident-level NRST allocation for nursing facilities probably magnifies these facility effects, rather than removing them. It therefore makes little sense to allocate NRST costs at the resident level.

Instead, we performed all derivation and CMI analyses using RST. The staffing component of the rate was then adjusted on the basis of RST-derived CMIs. This approach implicitly assumes that NRST should be allocated proportionally, which is the operational assumption for resident-level allocation anyway. Further, it has the advantage of not magnifying facility effects in the way that resident-level allocations do.

4.3.3 Nursing Time

In STRIVE, nursing time was used to yield a measure of resident nursing staff resource intensity, and this measure served as the dependent variable for analyses to evaluate the RUG-III system and to develop and test a RUG-IV system. For this purpose, the appropriate nursing time is the amount of nursing staff time devoted to an individual resident.

As described in the Phase I report,[18] the nursing STM collected the time of each nursing home staff member providing care to study residents during the identified 48-hour period. Data from the Palmtop Digital Assistants (PDAs) were sorted and accumulated into the time associated with each individual resident in the study, differentiated by the different staff job roles. Nursing group time (which was rare) was allocated equally among resident in the group as described in Section 4.3.1.

[18] STRIVE Phase I report, section 2, pp. 15

For each resident, we calculated two measures of nursing staff resource intensity: per-diem minutes (raw minutes) and per-diem wage-weighted minutes (wage-weighted staff time, or WWST). Per-diem minutes were obtained by dividing the raw 48-hour times by 2; thus the per-diem minutes represented in all computations are the average daily minutes over the two days. Per-diem WWST for each resident was calculated by wage-weighting the minutes by job role (e.g. Registered Nurse [RN], Licensed Practical Nurse [LPN], Certified Nursing Assistant, Certified Medication Aide, etc.) and then summing the per-diem WWST according to job category (for nursing, these were RN, LPN, and Aide - see Table 4-9). While WWST does not directly measure the cost of care, as it is normalized to the wage of a nurse aide, it does represent the ***relative*** per diem cost of care associated with staff time.

As previously mentioned, during our analysis we found a single observation in the full (derivation plus validation) sample with a high Nursing WWST (496 WWST) relative to the rest of the study population, in combination with a very high sampling weight (it was weighted 13.9 times more heavily than the average observation). Thus this observation potentially could have had a large impact on many of our results, such as the computation of the mean Nursing WWST for the RUG group in which it would be classified. Accordingly, we dropped this single observation from the analysis and calculation of CMIs. This left an analytic sample of 9706 residents with full and "valid" STM and RUG-classifiable MDS (i.e., 9721 residents with full STM; 9707 residents with full STM and RUG-classifiable MDS, less this observation).

Table 4-9 provides descriptive STM statistics, for nursing staff by category and in total. Directly-measured raw per-diem minutes show that RNs provided 16.2% (17.6/108.5) of the time, LPNs 21.1% of the time, and the remaining bulk of the time – 62.7% – was provided by aides. When wage weighted (to WWSTs), the percentages change to 29.7%, 25.0%, and 45.3% for RN, LPNs and aides, respectively

We also examined the distribution of these variables for outliers, which could skew our analyses and potentially give misleading results. High and low outliers were identified by examining the distribution of WWST by job category (i.e., RNs, LPNs, Aides). We observed for each job category that the 99th percentile marked a sudden large increase in the values, indicating relatively high outlying values. For RNs, this corresponded to observations exceeding 369 WWST (99 observations), for LPNs 193.5 WWST (96 observations), and for Aides 212 WWST (95 observations). Residents with values that exceeded these values were further checked by CareTrack Systems in order to ensure that there were no processing errors, and none were found.

Then, upper truncation was performed, by job category. All observations in the full analytic sample (N=9,706) were retained, but values exceeding the 99th percentiles were reassigned the value at the 99th percentile. For example, observations with RN WWST in excess of 369 were assigned this value. These upper truncated WWSTs for each of the three nursing job categories were then summed with all other observations to form the total Nursing WWST. Altogether, 281 observations had one or more of the following WWST values upper truncated: RN, LPN, Aide.

Low outliers were also addressed through truncation by assuming that every study resident should have at least 10 minutes of some nursing staff care. All residents in the full analytic sample had non-zero nursing time (residents with no nursing time at all were already removed, as indicated earlier in Section 3.2). Residents in the full analytic sample who had less than 10 total minutes of nursing time were assigned 10 minutes (N=133), which were then wage-weighted using the average weight for all nursing jobs (1.387).

The results of the upper and lower truncations on WWST as well as the raw minutes for the full analytic sample (N=9,706) are shown in Table 4-9. The result of these adjustments is a small reduction in the mean WWST for each staff type and the combined WWST. For example, the total nursing WWST declined from 150.5 to 148.2.

Table 4-9. Descriptive Statistics for Per Diem Nursing Time: Effects of Upper and Lower Truncation (N=9,706)

Variable	Mean	Median	Minimum	Maximum
RNs				
Raw minutes	17.6	6	0	374
Upper Truncated Raw minutes	17.2	6	0	163.5
WWST	44.7	15.5	0	965.3
Upper Truncated WWST	43.4	15.5	0	369
LPNs				
Raw minutes	22.9	15.5	0	339.5
Upper Truncated Raw minutes	22.6	15.5	0	117.5
WWST	37.7	25.5	0	559
Upper Truncated WWST	37.2	25.5	0	193.5
Aides				
Raw minutes	68	61	0	1417
Upper Truncated Raw minutes	67.4	61	0	216
WWST	68.2	61.3	0	1419
Upper Truncated WWST	67.5	61.3	0	212
Total Nursing				
Raw minutes	108.5	95.5	0.5	1435.5
Upper/Lower Truncated Raw minutes	107.2	95.5	10	448.5
WWST	150.5	124.7	0.5	1449.4
Upper/ Lower Truncated WWST	148.2	124.7	10	712.6

4.3.4 Therapy Time

In STRIVE, two therapy time measures were used.

1. **Resident Therapy Time.** Resident therapy time refers to the amount of time that a resident was in therapy. This measure is analogous to the rehabilitation therapy time reported on the MDS 2.0. Resident therapy time was used to classify residents in therapy categories for both RUG-III and RUG-IV.

2. Staff therapy time includes the amount of therapy staff time providing individual therapy to a resident plus staff time allocated to that resident when care is provided in a concurrent or group session (allocation being proportional to the number of residents involved). Staff therapy time is a measure of therapy staff resources devoted to an individual resident, and this measure served as the dependent variable for analyses to evaluate the RUG-III system and to develop and test a RUG-IV system. When staff therapy time is wage weighted by staff role, it yields a cost measure used to produce CMIs and payment rates.

The STRIVE estimation of resident therapy time and staff therapy time are discussed in turn.

4.3.4.1 Resident Therapy Time

During the STRIVE time study, we collected seven days of therapy time. Each staff member indicated how much time was spent with each resident. This measure was used in STRIVE to classify a resident, i.e., to determine if he or she should be in a Rehabilitation or Rehabilitation Plus Extensive group.

Therapy time was generally collected on PDAs only for the first three days of a facility study and then collected on a paper tool for the remaining four days, including weekends. This occurred because the PDAs needed to be cleared of all data and shipped to a new facility for availability at the beginning of the following week. In addition, different therapy staff members were on duty during weekend days; many of these staff members did not attend the orientation for PDA use at the beginning of the week. There was substantially less oversight of the collection of the therapy data on paper.

There were three different data collection schedules:

- Schedule A - Training was done on Monday and data collection began Tuesday morning.
- Schedule B - Training was done on Monday and data collection began during the night shift on Monday for nursing. Because therapy isn't given overnight, the therapy collection between the Schedules A and B were similar).

- Schedule C - When there was a Monday holiday, training began on Tuesday and data collection started on Wednesday.

In all cases, the therapy data collection continued for a complete seven-day period.

Examining the collected therapy data from residents present for the full week of data collection, we noted that Schedule A and Schedule B showed very similar percentages of reported weekly therapy across the 7 days of data collection. In particular, roughly 75 percent of the one-week reported therapy occurred on Tuesday, Wednesday, and Thursday, at approximately 25 % each day An additional 12 percent occurred on Friday, 10-12 percent on Monday, and a minimal 0-2 percent of the weekly therapy occurred on weekend days. The therapy time for Tuesdays, Wednesdays, and Thursdays were collected by PDA for these schedules and by the paper tool on Fridays through Mondays. In contrast, Schedule C had the PDAs available Wednesday through Friday, then used paper tools on Saturday through Tuesday. While all three schedules reported similar percentages of times for Wednesdays and Thursdays (26-30%) when they all used a PDA, we observed much higher percentages of weekly therapy time on Fridays (21 percent for Schedule C when we were using the PDA, compared to 12 percent for Schedules A and B when we were using paper instruments). Similarly, Schedule C therapy time for Tuesday (the 7th day of data collection) was only nine percent, compared to the 25 or 26 percent reported for Schedules A or B on Tuesday when the PDA was used. These observations led us to believe that therapy was underreported on weekdays when the paper tool was used and there was much less supervision by study staff, as our experience with PDA-based data collection was shown to be quite accurate.

To determine if the resident therapy time collected seemed reasonable, we classified the STRIVE residents into RUG-III rehabilitation groups using resident therapy time and compared the resulting distribution of RUG-III therapy groups for STRIVE Part A residents to the national distribution of RUG-III rehabilitation groups as reported on Medicare Part A claims. The STRIVE data had far fewer patients in the Ultra High, Very High, and High rehabilitation groups and far more patients in the Medium rehabilitation groups. This corroborated our hypothesis that STRIVE therapy time was underreported because of both the use of paper forms and less intense oversight.

We thus developed a methodology to estimate the “true” amount of resident therapy time by adjusting the underreported therapy minutes from the paper tools. In the following we describe the adjustment calculations and then provide an example.

We adjusted the STRIVE resident therapy times to best match the Medicare part A claims results, developing average therapy times per day and an estimate of the number of days of therapy given. The calculations were done separately using the resident time for each of the three therapy disciplines (physical therapy, occupational therapy and speech-language pathology). Specifically:

- First we accumulated each resident's time for each discipline by adding times across the several practitioners of that discipline, for example, for physical therapy we had therapists, assistants, and aides.
- Next, the resident's average PDA therapy session for each discipline was computed by summing therapy time reported on PDA days when there were 15 or more minutes of therapy provided; this sum is divided by the PDA day count of days when there were 15 or more minutes of that therapy time provided on that day.
- Finally, we estimated the number of days of therapy for each discipline that the resident received by counting the number of days of each therapy reported on the PDA, the paper tool and the STRIVE MDS. Using the three days of PDA therapy data collection as a basis we adjusted the number of therapy days in the following ways:
  - If the resident had therapy minutes on three of three PDA days of therapy data collection, we treated that resident as if five days of therapy had been received.
  - If the resident had therapy minutes on two of three PDA days of therapy data collection, we treated that resident as if three days of therapy had been received, and added one additional day if there was therapy time of 15 or more minutes indicated on the paper form for both of the remaining weekdays.
  - If the resident had therapy minutes on one of three PDA days of therapy data collection, we treated that resident as if one day of therapy had been received, and added additional days for each of the other two weekdays when therapy time of 15 or more minutes was indicated on the paper forms.
  - If the resident had no PDA days of therapy, we counted the other weekdays when therapy time of 15 or more minutes was indicated on the paper forms.
  - In all cases, even if there was no PDA day indicating the resident received therapy we counted any weekend days when therapy time of 15 or more minutes was indicated on the paper forms.

For each resident and each therapy, the adjusted weekly resident therapy time was computed by multiplying the average PDA therapy session time by the adjusted days of therapy. The ratio of the adjusted time to the original time – a "preliminary inflation factor" – was also computed, for use in the next step.

In addition to the adjustments described above, CMS mandates a limit on the group therapy time reported on the MDS and used in RUG classification for Medicare Part A payment. In particular, group time for a discipline is limited to at most 25% of the

total time reported for that discipline[19]. In the STRIVE study, the 25% limitation on group therapy was applied to all residents, not just residents in a Part A covered stay. Once we applied this limitation, to allow for the adjustment made in the prior steps, we multiplied the limited group time by the preliminary inflation factor just described.

These calculations led to an adjusted ***resident therapy time*** for each therapy discipline – physical therapy, occupational therapy, and speech/language pathology therapy.

The result of these adjustments is a much better fit to the national RUG-III distribution for rehabilitation groups for both the Part A subpopulation (See Figure 4-2) and the entire population (See Figure 4-3). By choosing the methodology described above, we better account for actual resident therapy time, giving average credit for any day that resident therapy time was recorded for 15 or more minutes.

A sample resident therapy time adjustment calculation is provided in Exhibit 4-1.

[19] The group time limitation is not that group time is 25% or less of the total therapy time received. Rather the limitation is that adjusted group time is 25% or less of the total time remaining after adjusting group time.

Exhibit 4-1: Sample Calculation of STRIVE Therapy Adjustment of Staff Time

Consider a resident, with Occupational Therapy (OT) times, by day, as follows:

Day 1 (PDA)	**41** minutes
Day 2 (PDA)	**60** minutes – **30** group time
Day 3 (PDA)	**54** minutes – **40** group time
Day 4 (paper)	**60** minutes – **30** group time
Day 5 (paper, weekend)	**60** minutes
Day 6 (paper, weekend)	**0** minutes
Day 7 (paper)	**0** Minutes
TOTAL	**275** Minutes – **100** group time

Step 1: Calculate average therapy session:

Sum minutes on PDA days where >=15 minutes: **41**+**60**+**54**= **155** minutes.

Divide the sum by the number of PDA days where >=15 minutes: **155**/**3**= **51.67** minutes, the average therapy session for OT.

Step 2: Estimate number of PDA days received >=15 minutes of therapy:

Count number of PDA days where OT time >= 15 minutes: **3** days (Days 1, 2 and 3).

Inflate the number of PDA days where >=15 minutes according to algorithm described above: **3** out of **3** PDA days, so assign **5** days of OT therapy during the week.

Add in any weekend days where >=15 minutes: **5** + **1** = **6** days (to be used in RUG classification).

Step 3: Obtain adjusted total resident OT weekly minutes by multiplying average therapy session (step 1) with inflated days (step 2):

51.67 minutes x **6** days= **310** minutes per week for OT.

Step 4: Check if group adjustment is necessary

Total unadjusted group time: **100** (**30**+**40**+**30**) minutes group time out of the total of **275** minutes (before any adjustments); so there are **175** minutes of non-group time.

Group time as a percent of total times is **36.3%** (**100**/**275**).

As 36.3% is greater than 25%, a group adjustment is necessary to limit the group time to no more than 25% of the total time remaining after the group time limitation is applied.

When a group time limitation is required, the non-group time should be **75 %** of the total time after limitation. In this case the **175** minutes of non-group time = 75% of the total time after limitation. That is: 175 / total time after limitation = 0.75. Solving this expression for total time after limitation gives: Total time after limitation = 175 / 0.75. The total time allowed with 25% non-group time is **233.3** (175 / 0.75) minutes. The maximum group time that is allowed for the resident is **233.3 – 175 = 58.3** minutes (not all 100 group time minutes).

Step 5: Calculate the inflation factor by dividing the adjusted total resident time calculated in Step 3 (310 minutes) by the original unadjusted total resident time (275 minutes). **310/275 = 1.127** inflation factor value.

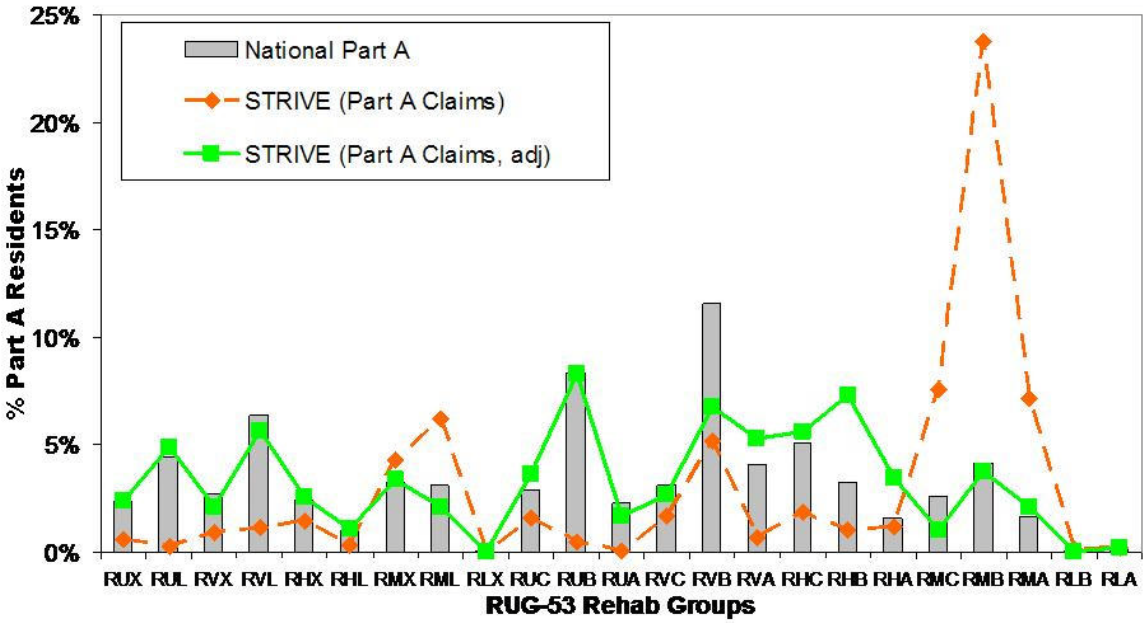
Adjust the total resident time allowed after group time limitation (233 from Step 4) for the inflation factor (1.127): **1.127***233.33**= **263.03** adjusted group-limited weekly OT minutes.

The result of these calculations is that the final OT weekly resident minutes used for RUG classification is set to **263.03** minutes, after adjustment and application of the 25% limitation on group time.

After adjusting the therapy time, we also realized that the number of extensive services reported on the STRIVE MDS was low compared to the national MDS data. As a result, the STRIVE indicators for the extensive services used in RUG-III (parenteral IV, IV medications, suctioning, tracheostomy, and ventilator/respirator) were revised based on the assumption that the STRIVE resident was receiving one of these services if the service was indicated on either their STRIVE MDS or the nearest RPT assessment (based on routine MDS assessments).

After revising the indicators for the extensive services and adjusting resident therapy time, the STRIVE distribution of all 53 RUG-III groups reasonably matched the national MDS distribution, both for Medicare Part A residents and all residents. Figure 4-2 displays the distribution of Medicare Part A residents in the RUG-III Rehabilitation Plus Extensive groups and the Rehabilitation groups. In this figure, the national distribution of Medicare Part A residents is shown with a bar representing each of these groups. Superimposed on this is the STRIVE distribution of Medicare Part A residents in these RUG-III groups based on the STRIVE unadjusted resident therapy times and unrevised extensive services (dashed line) and the STRIVE adjusted resident therapy times and revised extensive services(solid line). Figure 4-3 displays the same three distributions, but for all STRIVE residents versus all residents in the nation.

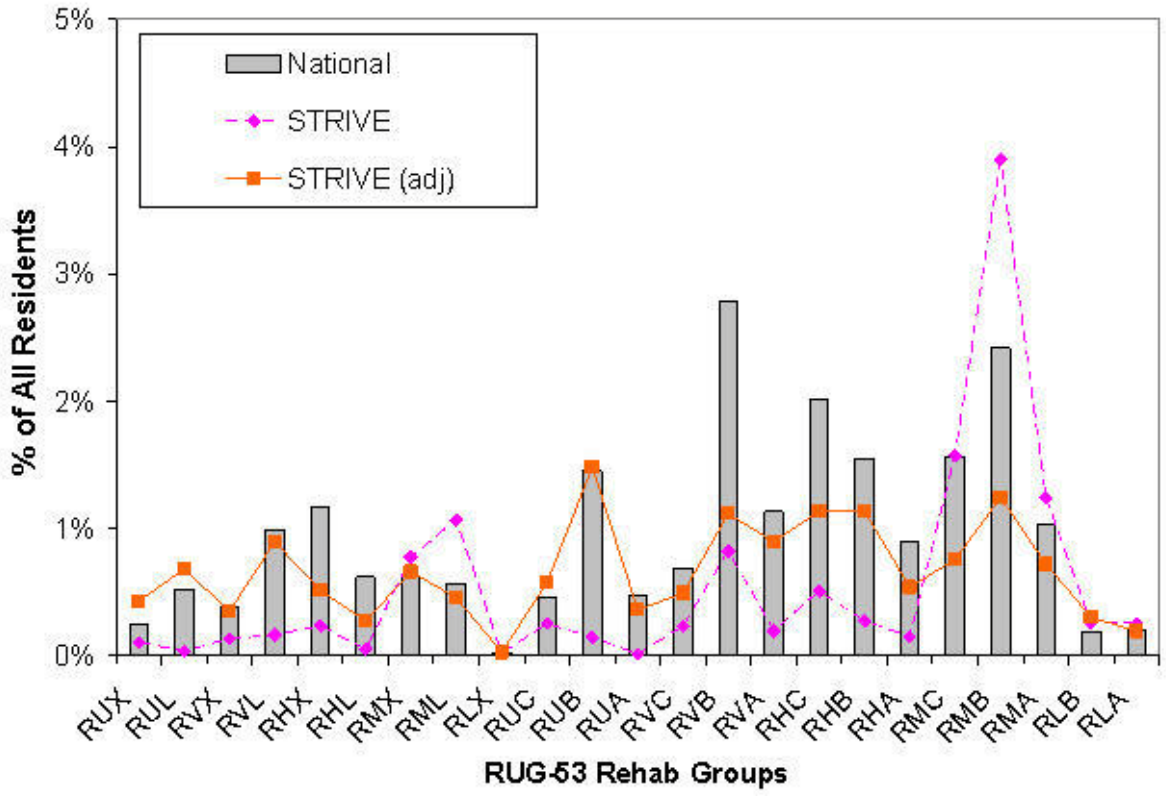
Figure 4-2. Frequency Distribution Across RUG-III Rehabilitation Plus Extensive and Rehabilitation Groups for Medicare Part A Residents (based on a Part A Claim): National Distribution and STRIVE Distributions Based on Resident Therapy Time with the Therapy Time Adjustment and Extensive Services Revision and without the Adjustment and Revision*



* The bars represent the national distribution. The dashed line is the STRIVE resident distribution without adjustment of therapy time or revision of extensive services indicators. The solid line is the STRIVE distribution after adjustment of therapy time and the revision of extensive services indicators for underreporting (see prior text).

In Figure 4-1, the match between the adjusted STRIVE Part A distribution and the national Part A distribution is quite good from RUX down to RVC and also from RMB to RLA. The match is not so good for RVB to RMC. The two groups with the largest discrepancies are RVB with STRIVE having about 5% fewer residents than the nation and RHB with STRIVE having about 5% more residents than the nation. It is as if some Very High rehabilitation residents have shifted down to High rehabilitation in STRIVE. Ongoing 2011 research by the Office of the Inspector General (OIG) is indicating that in some facilities rehabilitation level is sometimes higher during an MDS 7-day observation period than outside of an observation period, establishing a high payment rate based on the observation period and this high payment rate continues even when rehabilitation is subsequently reduced. In STRIVE, the 7-day time study period often did not correspond to an MDS observation period (basis for the national distribution). If some facilities reduce rehabilitation outside of the observation period, then STRIVE might be expected to show some shifting of residents to lower level rehabilitation. The STRIVE results for RVB and RHB are consistent with the current OIG findings. Future time study research should address the possibility that rehabilitation level during an MDS observation period may differ from rehabilitation level in the rest of a Part A stay.

Figure 4-3. Frequency Distribution Across RUG-III Rehabilitation Plus Extensive and Rehabilitation Groups for All Residents: National Distribution and STRIVE Distributions Based on Resident Therapy Time with the Therapy Time Adjustment and Extensive Services Revision and without the Adjustment and Revision*



* The bars represent the national distribution. The dashed line is the STRIVE resident distribution without adjustment of therapy time or revision of extensive services indicators. The solid line is the STRIVE distribution after adjustment of therapy time and the revision of extensive services indicators for underreporting (see prior text).

4.3.4.2 Staff Therapy Time

Recall that staff therapy time includes the amount of therapy staff time providing individual therapy to a resident plus staff time allocated to that resident when care is provided in a concurrent or group session (allocation being proportional to the number of residents involved). Staff therapy time is a measure of the therapy staff resources used by a single resident. When staff therapy time is wage weighted by staff role, it yields a cost measure used for analysis and the calculation of CMIs and payment rates.

Staff therapy time is reduced resident therapy time if there is any therapy provided in a concurrent or group session, with staff therapy time decreasing more with both the amount of time in a concurrent or group session and the number of residents in the session. Overall very little resident time was spent in group sessions where all residents received the same treatment (same HCPCS procedure code). The percentage of total resident therapy time involved in a group session was 1% physical [PT], 1% for occupational therapy [OT], and 4% for speech-language pathology [SLP]. Considerably more resident time was spent in concurrent sessions where residents received different treatments (different HCPCS codes), with 27%, 28%, and 14% for OT, PT, and SLP, respectively.

Table 4-10 displays the per diem mean and standard deviations for the resident and the staff therapy raw times (unadjusted times as reported with the PDA and paper tool).

Table 4-10. Descriptive Statistics for Per Diem Staff and Resident Therapy Raw Minutes*

Category	Number of Observations	Mean Value	Standard Deviation	Minimum Value	Maximum Value
Staff time	3024	49.5	33.9	0.3	203.9
Resident time	3024	61.9	43.2	0.5	245.2

*Only residents with therapy times are included.

The staff therapy time was used to produce a ***raw*** wage weighted staff time (WWST) measure for therapy time (therapy staff cost measure) to potentially serve as a dependent variable for RUG-III analyses and RUG-IV derivation. The raw therapy WWST per diem measure was calculated by wage weighting each staff therapy minute by a relative wage weight for the staff role involved, accumulating all wage weighted minutes for the 7-day study period, and then dividing by 7 to obtain a per diem measure.

Recall that the raw therapy times were low because of the underreporting on the therapy time paper collection tool used for 4 of the 7 days. To adjust a resident's

therapy WWST for this underreporting, we inflated the raw per diem WWST values for each discipline (OT, PT, and SLP) by the adjustment inflation factor for that discipline for that resident. The inflation factor previously calculated previously for resident therapy time was used for this calculation of ***adjusted*** WWST.

Exhibit 4-2 presents an example of this adjustment of WWST for underreported therapy time.

EXHIBIT 4-2: Example Calculation of STRIVE Therapy WWST Adjustment for a Resident

Consider a resident with per diem Occupational Therapy (OT) time of 30.43 after allocation of concurrent and group time.
Compute per-diem unadjusted WWST:

Multiply relative wage weight for OT job role by per diem allocated minutes:

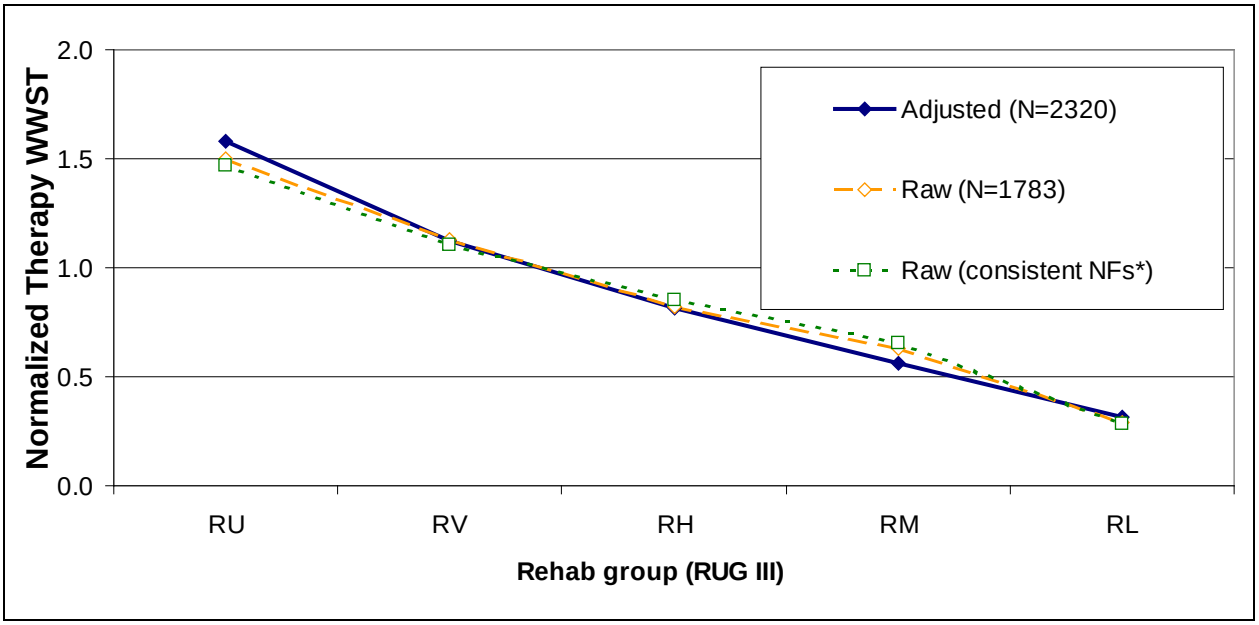
OT:	**30.43** minutes * **2.724** wage weight=	**82.9** WWST
COTA:	**0** minutes * **1.90** wage weight=	**0** WWST
OT aide:	**0** minutes * **1.13** wage weight=	**0** WWST
TOTAL		82.9 WWST

Multiply per diem total WWST by inflation factor (1.127) for this resident and therapy discipline:

82.9 unadjusted per diem WWST ***1.127**= **93.45 Adjusted WWST**

The adjustment of the therapy WWST will affect the computation of the therapy CMIs. We studied the effect of using three alternative WWST resource-use measures for the RUG-III Rehabilitation categories. The first alternative was the adjusted WWST. The second was the raw WWST (unadjusted for underreporting of therapy time). For the third alternative, we used the raw WWST from 50 "consistent" nursing facilities where the therapy time data collection appeared consistent across the entire week and probably did not involve underreporting. The three different WWST measures produced roughly the same CMIs (see Figure 4-4) because the RUG-III Rehabilitation categories use therapy time cutoffs, e.g., the Ultra High Rehabilitation category requires 720 minutes of therapy, the Very High Rehabilitation category requires 500 minutes of therapy, etc. The reader should note that while the adjustment to staff therapy time does not significantly influence change the CMIs, the adjustment to resident therapy time does change the STRIVE distribution by increasing the number of observations in the higher level rehabilitation RUG-III categories. Figure 4-4 plots the three alternative therapy WWST measures for the RUG-III rehabilitation categories, with the values for each WWST measure being normalized to have a mean value of 1.00 for all residents in these categories.

Figure 4-4. Normalized Therapy WWST for Three Therapy WWST Measures



While there is little difference due to the adjustment of therapy WWST for underreporting of therapy, we chose to use the adjusted therapy WWST as the dependent variable for analysis. For consistency, the choice was made to use the same adjustment methodology for therapy WWST as that used for resident therapy time.

4.3.5 Transformation of Dependent Variables

Many of the statistical tools used in STRIVE assume that the dependent variable has a normal distribution. However, cost-type variables, including the WWST in STRIVE, frequently have a skewed distribution. Using a natural logarithm transformation of the dependent variable usually results in a more normal distribution. A disadvantage to this is that in order to interpret the results, the estimates must be back-transformed.

We performed several of the initial STRIVE analyses using both WWST and log-transformed WWST to test whether transformation would improve our analyses. These included the analysis of pre/post-admission extensive services, the split between the Rehabilitation and the Rehabilitation and Extensive categories, and the secondary splits for the Extensive Services category. Since no substantial differences were seen in the conclusions when the log-transformation was used; we dropped this option in all further analyses.

4.3.6 Wage weighting

In prior sections, we discussed the use of Wage-Weighted Staff Time (WWST), representing the relative daily cost of care, as the primary dependent variable to

develop RUG-IV. WWST, as discussed in prior sections, was computed by multiplying minutes of staff time by a relative wage rate specific to staff roles (e.g., registered nurse, licensed practical nurse, etc.). The rates were developed by matching roles with the Bureau of Labor Statistics wage rates for 2006;[20] the weights used were determined by the ratio of that role's hourly wage rate to the hourly wage rate of a certified nurse assistant (CNA) of $10.67; for example, the wage weight for an LPN with a wage rate of $17.57 was set to 1.65 ($17.57/$10.67). The choice of the Bureau of Labor Statistics rates is discussed in the STRIVE Phase I Report Section 3 pages 49-52. The weights rates and resulting relative wage weights used in STRIVE are shown in Table 4-11.

Table 4-11. Hourly Wage, Job category, and Wage Weights Used in Calculating WWST

Description	Median Hourly (2006$)	Job Category	Wage Weight (Standardized to Median Hourly wage for CNA)
Registered Nurse (RN)	$27.54	RN	2.58
Respiratory Therapist	$22.80	RN	2.14
Licensed Practical Nurse (LPN)	$17.57	LPN	1.65
Certified Nurse Assistant (CNA) Geriatric Nurse Assistant (GNA) Resident Care Technician (RCT)	$10.67	Aide	1.00
Certified Medication Aide (CMA)	$10.67	Aide	1.00
Restorative Aide	$12.80	Aide	1.20
Bath Aide	$9.09	Aide	0.85
Feeding Aide	$9.09	Aide	0.85
Psych Aide	$11.49	Aide	1.08
Non Certified Care Technician	$9.09	Aide	0.85
Clinical Associate	$10.67	Aide	1.00
Transportation	$9.09	Aide	0.85
Respiratory Therapy Assistant	$18.81	Aide	1.76
Physical Therapist (PT)	$31.83	PT	2.98
PT Assistant	$19.88	PT	1.86
PT Aide	$10.61	PT	0.99
Occupational Therapist (OT)	$29.07	OT	2.72
Certified OT Aide (COTA)	$20.22	OT	1.90
OT Aide	$12.03	OT	1.13
Speech/Language Pathologist (SLP)	$27.74	SLP	2.60
Audiologist	$27.46	SLP	2.57

[20] Bureau of Labor Statistics, U.S. Department of Labor, Occupational Employment Statistics, [accessed: April 6,2007] [www.bls.gov/oes/].

4.4 Baseline Functioning of the RUG-III Case-Mix System

An initial evaluation of the RUG-III 53 group system was performed for several purposes:

- Evaluate the validity for the system currently in use.
- Compare its performance with that in previous studies.
- Provide the basis to which we can compare improvements to the RUG system.
- Identify any major structures of the RUG-III system that do not adequately explain resource use or differ from previous studies.

Three criteria were used to assist in evaluating the RUG-III 53 group system:

- The variance explanation of RUG-III and its principal component, the RUG-III ADL Index.
- The hierarchical ordering of mean Nursing WWST from the highest to the lowest RUG-III major category.
- Comparison of "nursing relative weights" with the CMIs currently being used for SNF payment.

We describe these evaluations as follows:

- Variance explanation: The RUG-III (53 group) system had a sample-weighted variance explanation of nursing wage-weighted staff time (WWST) equal to 29.1%; among residents who were Medicare Part A, the variance explanation was 20.5%. When therapy WWST was added to the Nursing WWST, the variance explanation increased to 55.8% (40.3% for Medicare Part A). These values are similar to the variance explanation reported in the initial development of RUG-III[21], which reported a variance explanation of 43% of total cost (therapy, nursing and other job roles).

The ADL Index used in RUG-III had a variance explanation of 20.8% of Nursing WWST for all residents which, again, was similar to the 18% variance explanation reported in the earlier study.

- Hierarchy of means: The mean WWSTs for each of the major RUG-III categories were also calculated. See Table 4-12. The means generally confirm the hierarchy of the major groups. It should be noted that the higher value for the Reduced Physical Function category compared with the Cognitively Impaired and Behavior Problem categories is not a violation of the hierarchy, as the latter groups are restricted to more ADL-functional individuals.

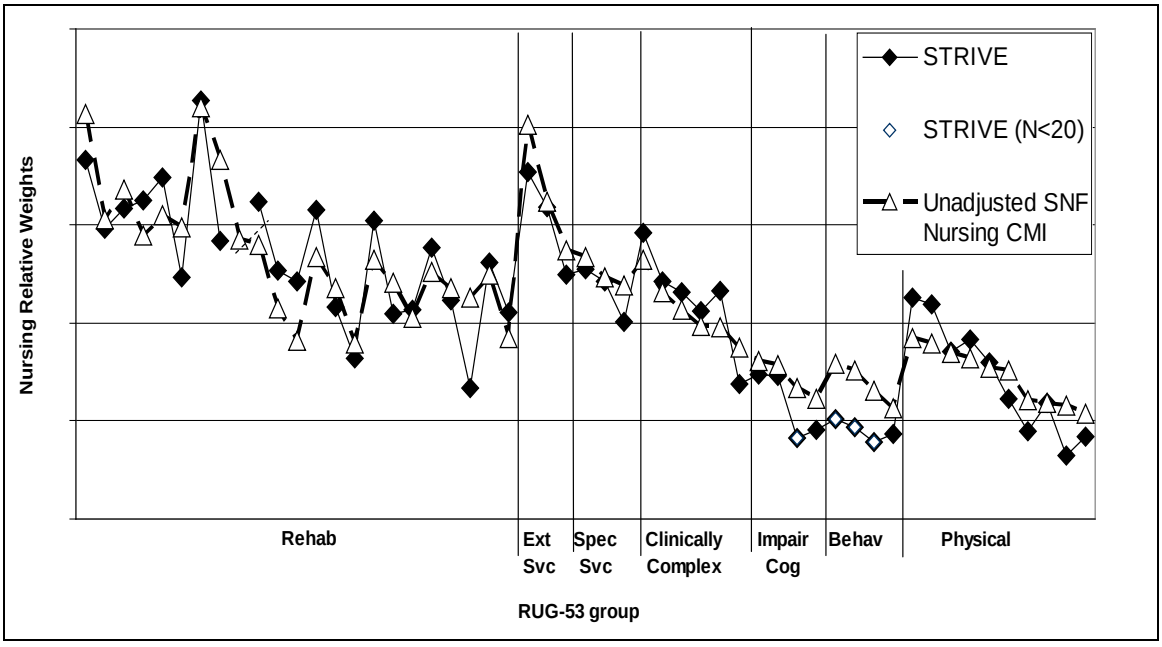
[21] Fries et al, 1994, *op cit.*

Table 4-12. Sample-weighted Nursing and Therapy Wage-weighted Staff Time by Major RUG-III (53 group) Category

Major RUG-III Category (53 group)	N	Mean Nursing WWST	Mean Nursing and Therapy WWST
1. Rehabilitation Plus Extensive	831	224.71	378.67
2. Rehab	1489	159.19	299.27
3. Extensive Services	825	216.93	220.97
4. Special Care	784	155.95	158.98
5. Clinically Complex	1701	137.60	140.25
6. Cognitively Impaired	923	88.05	89.24
7. Behavioral Problems	68	60.40	60.93
8. Reduced Physical Function	3086	116.21	118.16

- Comparison of "nursing relative weights" with SNF payment CMIs: Additionally, we examined the relationship between the unadjusted SNF nursing payment CMIs and a comparable measure, nursing relative weights (see Figure 4-5). The nursing relative weights were calculated by dividing the RUG-III group means by the overall mean and then adjusting for the differing RUG-III distributions between the STRIVE study population and the population on which the SNF payment CMIs are based.

Figure 4-5. Comparing STRIVE Nursing Relative Weights with Unadjusted SNF Nursing CMI



The STRIVE nursing relative weights are similar to the Unadjusted SNF Nursing CMIs, with the exception of a few groups (e.g. Behavioral) where small numbers of STRIVE residents result in unstable means.

Overall, the RUG-III 53-group system had good baseline predictive ability, with variance explanations comparable to the original study, and with derived relative cost measures of similar magnitudes. An implication is that while nursing facility care patterns may have changed in the decade since the derivation of the RUG-III system, the relative cost of different types of residents has not changed substantially. On the basis of these results, we decided that project efforts should focus on refinement of RUG-III to a RUG-IV system, rather than the derivation of a fully new system paradigm.

4.5. Evaluate potential of new items/scales

Another goal of STRIVE was to consider the effectiveness of new assessment items and new scales, which combine assessment items to describe an assessment domain that explains resource utilization.

The STRIVE Assessment Addendum form is provided in Appendix B. It includes items representing new concepts from the nursing home MDS 3.0 assessment that is under development, the OASIS assessment used in home health care, and interRAI assessments used in a variety of settings, in addition to items that would allow us to address the explicit goals of STRIVE (see Section 3.3) such as differentiating between pre- and post-admission services. The STRIVE Phase I Report[22] described the construction of the STRIVE Addendum, which contained these new items.

As part of the STRIVE project, we checked the feasibility and inter-rater reliability of Addendum items, as potential MDS 3.0 items. Many of these Addendum items were only included during part of the STRIVE data collection and several went through one or more versions. It was not intended that these items would be part of RUG-IV, although data on limited subsets did permit feasibility and reliability testing.

Over the course of the data collection, we used as many as 151 potential MDS 3.0 items (including revisions of items) in the Addendum. Specifically:

- 71 were items describing post-discharge plans, the size and depth of pressure ulcers, and different versions of the pain items included in the data collection for the sole purpose of MDS 3.0 testing; they were not intended for STRIVE RUG-IV use.
- Beyond these, two items were used only in a 57-resident pilot test and were dropped as being redundant with another item, (Date of Assessment) or impractical to collect (Estimated Survival).
- The Patient Health Questionnaire (PHQ) depression items for MDS 3.0 became available late in the STRIVE study. We tested these for use in RUG-

[22] STRIVE Phase I report, pp. 16

IV, but could not fully test for RUG-IV because they were available only for 51% of the sample. Instead, we used the MDS 2.0 measures to derive the depression scale for RUG-IV.

- The remaining 59 items designed for testing as potential RUG-IV items, were included in most versions of the Addendum and collected for over 96% of the sample (these were modified during or soon after the pilot test facility was completed). These items represent dimensions not included in MDS 2.0 or are items needing improvement based on documented feedback from MDS 2.0 users.
- After the reliability testing (see Section 7.2), we eliminated items likely to have poor incentives from further analysis. These items could be associated with higher resource use and controlled by a facility, e.g. catheter use.

The items in the Addendum considered relevant to case-mix analysis included:

- Procedural memory
- Abnormal thought process
- When confused
- Cognitive functioning
- Violent ideation
- Intimidation of others or threatened violence
- Violence to others
- Distance walked
- Distanced wheeled self
- Knee replacement
- Hip replacement
- Hip fracture
- Number of ADLs in which person independent prior to hip/knee replacement or hip fracture
- Post-traumatic stress disorder
- Sleep apnea
- Fatigue
- Dyspnea
- Pain interview panel (see Section 4.5.1)
- Pain control (see Section 4.5.1)
- Coughing/choking

- Changes in sound of voice
- Caregiver observations of difficulty eating
- Resident complains of swallowing difficulty
- Primary mode of nutritional intake
- Number of days in last 7 of parenteral feeding
- List of twelve treatments and services performed prior to/after admission, including six new services: modified barium swallow, BIPAP/CPAP machine, hyperbaric oxygen, isolation for inflection control, negative pressure wound therapy, and PET scan. With the exception of hyperbaric oxygen, isolation for infection control, and negative pressure wound treatment, we determined the remaining items had the potential for gaming, and dropped them from further consideration.

The Addendum items with frequency distributions are in Appendix C.

Several assessment domains were singled out for specific analyses. These include pain, depression, and pre-/post-admission services. In the following sections, we describe the analyses of these domains, followed by a description of our approach to new, relevant case-mix items in the STRIVE Addendum.

4.5.1 Pain

Because pain control is consistently identified as a major care goal in nursing facilities, it has been suggested for use in case-mix systems.

We included multiple measures and scales in the STRIVE assessments that could be the basis for a pain criterion:

- MDS 2.0 Pain Scale (derived from the two MDS 2.0 items: pain frequency and pain intensity).[23]
- MDS 3.0 Resident Pain Interview, a panel of six[24] pain items planned for MDS 3.0 and based solely on the interview of a resident (see STRIVE Addendum – Appendix B). By design, this panel of items was not performed on residents unable to communicate. In addition, as part of the MDS 3.0 development process, the items in the panel changed three times during the STRIVE data collection, so analyses involving these items were performed only on a subset of the sample. For those responding to the last version

[23] Fries BE, Simon SE, Morris JN, Flodstrom C, Bookstein FL. "Pain in US Nursing Homes: Validating a Pain Scale for the Minimum Data Set" *Gerontologist* 41(2):173-179, 2001.

[24] A seventh item, XJ5f in the STRIVE Addendum (see Appendix B), was not considered here because it involves pain management rather than pain itself.

(N=2,049), the Cronbach's Alpha[25] showing concordance of this panel was 0.77, close to the standard 0.80 considered acceptable.

- MDS 3.0 Facility Pain Indicator, a panel of five items planned for MDS 3.0, to be based solely on facility assessors' determination. Although the MDS 3.0 intends to have these items collected only on those individuals who cannot respond to the Resident Pain Interview (see above), the STRIVE data collection used these items for all individuals. However, as with the Resident Pain Indicator, these items were changed during the STRIVE data collection, so the analyses involving these items were performed only on a subset of the sample. For those responding to the latest version (N=5,410) the Cronbach's Alpha was 0.71.
- A combination of the MDS 3.0 Resident Pain Interview and MDS 3.0 Facility Pain Scale to simulate the pain items that would be available on MDS 3.0 if these two panels of items were used. If a resident could self-report on pain then we indicated pain based on the MDS 3.0 Resident Pain Interview Scale; if not, we used the MDS 3.0 Facility Pain Scale.
- STRIVE Pain Control item that indicates both whether there is pain and whether it is controlled by a therapeutic regimen or tolerated (See STRIVE Addendum Items– Appendix B).

All of these scales had acceptable internal consistency of 0.80 or more as measured by Cronbach's Alpha: that is, they all seem to measure the single issue of pain.

Nevertheless, we were concerned about the validity of the MDS 3.0 pain interview. For example, 197 (138 + 59) residents reported that they had pain even though they had severe dementia (as indicated by Cognitive Performance Scale[26] values of 5 or 6) – see Table 4-13. Also, a cross-tabulation of pain control and the MDS 2.0 Pain Scale showed little agreement as evidenced by considerable numbers of observations for every combination of levels of these two items, not just those for which *a priori* we might expect agreement (data not shown).

[25] Cronbach's Alpha is a statistic that measures the internal consistency of a set of measures of the same concept. Applied here, it measures whether the several items in the Resident Pain Interview were measuring a common concept.

[26] Morris JN, Fries BE, Mehr DR, Hawes C, Phillips C, Mor V, Lipsitz LA. "MDS Cognitive Performance Scale" *J. Geront: Medical Sciences* 49(4):M174-M182 (July) 1994.

Table 4-13. Pain Self Report (Item XJ4) by Cognitive Performance Scale (CPS)

CPS	Self-report no pain	Self-report pain
0 (intact)	89	1108
1	68	625
2	96	703
3	215	1262
4	133	250
5	157	138
6 (very severely impaired)	295	59
TOTAL	1053	4145*

*Missing observations dropped.

We tested whether each of the five potential pain criteria (MDS 2.0 Pain Scale, MDS 3.0 Resident Pain Interview, MDS 3.0 Facility Pain Indicator, Combined MDS 3.0 Resident and Facility Pain Indicators, and STRIVE Pain Control Item) was able to explain the variations in WWST (either Nursing WWST or Nursing Plus Therapy WWST) remaining in the data after fitting the RUG-III system.[27] For each potential pain criterion, the variance explanation added to that of RUG-III was very small (0.01).

In addition, we were unable to develop an approach to include pain in RUG-IV that would provide appropriate incentives. A primary problem was that including pain in any manner as a criterion in RUG-IV would provide an incentive to keeping a resident in pain when pain control was possible or if the person decided to accept their pain level (e.g., to preserve mental clarity at the end of life). For all of these reasons, we did not include pain as a criterion in RUG-IV.

4.5.2 Depression

Depression is used as a tertiary split for Special Care and Clinically Complex in RUG-III. With additional mood items designed for MDS 3.0 included in the STRIVE Addendum, we had several alternative constructions of a criterion.

We began by considering two measures of depression derived from MDS 2.0:

- A count of the 16 mood symptom items from MDS 2.0 Section E that were exhibited five or more days per week. This is the depression criterion used in RUG-III.

[27] To control for general clinical and functional status, WWST residuals were created by running the regression WWST = a+b*RUG, where RUG was the categorical variable of all 53 RUG-III groups and computing the residual WWST (actual WWST minus predicted WWST) for each observation. The test, run for each pain criterion, examined the variance explanation of the regression: residual WWST=c+d*PAIN, where PAIN was the 0-1 dichotomous (indicator) variable indicating the absence or presence of pain.

- The Depression Rating Scale (DRS),[28] formed by summing 7 selected mood indicators.

We examined various alternatives for a depression criterion derived from the two "counting" measures by setting alternative thresholds. Of these, the measure with the best characteristics (a combination of larger variance explanation and larger difference between the Nursing WWST for individuals with depression vs. those without depression) was the 16-item Mood definition, with a count of three or more to indicate depression (see Table 4-14).[29] Using this measure and threshold, 21% (1,359 out of 6,454 in the derivation sample) of residents were deemed depressed; those with depression had a mean Nursing WWST of 164 in contrast to 144 for all residents without depression. As the MDS 3.0 would be using the PHQ-9 Mood Scale, we converted the MDS 2.0 16 item scale threshold to an equivalent threshold for the PHQ-9 Mood Scale.

Table 4-14. Identifying a Threshold for the MDS 2.0 16-Item Mood Scale: Mean Nursing WWST and RUG-IV Residuals, by Thresholds of the 16-Indicator RUG-III Mood Count

Alternative Thresholds (Depression Indicated When Count Meets or Exceeds Threshold)	**Depression***		**Mean Nursing WWST**			**Mean Residuals**		
	Number of Observations	Prevalence	Depressed Residents	All Others	Difference**	Depressed Residents	All Others	Difference**
1	3,272	51%	154.9	141.3	13.6	6.5	-6.7	13.2
2	2,080	32%	159.3	142.9	16.4	11.6	-5.5	17.1
3	1,359	21%	164.0	144.0	20.0	17.1	-4.6	21.7
4	868	13%	165.9	145.5	20.4	20.7	-3.2	23.9
5	563	9%	169.5	146.2	23.3	23.8	-2.3	26.1

* Analysis done for derivation sample (N=6,454)
** Differences statistically significant at p<.05

We also used partial data to test a third measure. The RAND Corporation recommended incorporating the Patient Health Questionnaire (PHQ) into the MDS 3.0. The PHQ includes nine depression items obtained from a resident interview and ten similar items from a staff assessment. RAND suggests that depression is indicated by a resident interview PHQ summary score at or exceeding the threshold of 9.5, where the

[28] Burrows AB, Morris JN, Simon SE, Hirdes JP, Phillips C. "Development of a Minimum Data Set-base Depression Rating Scale for Use in Nursing Homes." *Age and Ageing,* 29(2):165-172, 2000.
[29] To control for general clinical and functional status, WWST residuals were created by running the regression WWST = a+b*RUG, where RUG was the categorical variable of all of all RUG-IV Version 7 groups, and then computing the residual WWST (actual WWST minus predicted WWST) for each observation. The test, run for each depression measure and threshold, examined the variance explanation of the regression: residual WWST=c+d*DEPRESS, where DEPRESS was the 0-1 dichotomous (indicator) variable indicating the absence or presence of depression.

summary score is formed by the summation of the frequency scores on the nine interview items[30]. RAND also suggests that when resident interview items are missing, the PHQ score is to be computed in the same way from the ten staff assessment items. PHQ items were tested in the STRIVE data collection as part of the STRIVE Addendum. In the STRIVE data, adjustment was made for missing PHQ frequency items by using the average non-missing item score rather than the sum, with this average being the summary score divided by the number of non-missing items. As the panel of addendum items changed over the course of the data collection (see STRIVE Phase I Report[31]), usable PHQ data was collected only on 43% of the sample (2,743 observations from the derivation sample of 6,454).

RAND used a depression threshold of 9.5 applied to the sum of the PHQ symptom frequency items. The corresponding threshold for the average item score would be about 1 (average sum of 9.5 divided by 9 items). However, this threshold of 1 cannot be applied to the average item score used in the STRIVE study. The RAND PHQ frequency item scaling was 0 = occurrence on 0 to 1 days of the last 14 days, 1 = occurrence on 2 to 6 days, 2 = occurrence on 7 to 11 days, and 3 = occurrence on 12 to 14 days. In contrast the STRIVE PHQ frequency item scaling was 0 = 0 days, 1 = 1 day, 2 = 2 to 6 days, 3 = 7 to 11 days, and 4 = 12 to 14 days. A specific symptom frequency above 0 days will yield a value 1 higher for the STRIVE scaling than the RAND scaling. The appropriate threshold should be twice as high for the STRIVE data than the RAND data. The depression threshold for the STRIVE PHQ average item score was therefore set at 2. The STRIVE PHQ with a average item score threshold of 2 identified 25% (673 out of the 2,743 observations with PHQ data) of the STRIVE derivation sample, and had the highest differentiation in Nursing WWST between those identified as depressed (192) and all others (168). At 25%, this rate of depression is also similar to that seen with the MDS 2.0 Mood Scale with a threshold of two (21%), but the two measures are not concordant (see Table 4-15). This lack of concordance does not indicate which is the better measure of depression.

Table 4-15. Concordance of Two Measures of Depression: the PHQ and the DRS

	No Depression (PHQ score<2)	Depression (PHQ score >= 2)	Total
No Depression (DRS < 3)	2,070	133	2,203
Depression (DRS >=3)	0	540	540
Total	2,070	673	2,743

[30] The STRIVE Addendum (Appendix B) has the nine interview items as XE2a-I, and scored 0-4, with additional codes for frequency not known (5) and no response (9); these additional codes were considered missing for the analysis. The PHQ score was thus the summation of the individual scores of these nine items, divided by 9. If fewer than nine items were available, then the score was produced by those available, divided by the number of items available (e.g., if only 7 items were available, the sum of the scores of the seven items was divided by seven). The RAND documentation indicated that a score of 9.5 or greater – the threshold – would indicate depression.

[31] STRIVE Phase I report, pp. 16

In conclusion, the MDS 3.0 PHQ appeared to be the best measure to identify depressed individuals with higher WWST. Unfortunately, it was unavailable for over half of the STRIVE sample precluding its broad use in the analysis. As a result, for the vast majority of the analyses described here we used the MDS 2.0 Mood Scale (with a threshold of 2 – the original RUG-III criterion), to model depression.

4.5.3 Frailty

Earlier work during the RUG-III study sought, but did not find, a measure of medical fragility or acute needs, as a possible classification concept for case mix. In STRIVE, we considered developing such a measure using the newer concepts about frailty as a geriatric syndrome. "Frailty" is defined as a chronic condition acquired with aging and associated with adverse outcomes, such as ADL impairment, falls, institutionalization, and death. The MDS-based frailty measure we developed in STRIVE was analogous to a research-based frailty measure by Bandeen-Roche *et al.,*[32] which covers five areas: shrinking, weakness, exhaustion, slowness, and low activity. We identified MDS items from each of these five areas and checked them for internal consistency, i.e., whether these items occur concurrently sufficiently often that they plausibly measure a single construct. The fourteen MDS or STRIVE Addendum items identified for each of these five areas were:

- Shrinking: body mass index < 20 (MDS items K2a and K2b), weight < 125 lbs. (K2b), recent weight loss (K3a).
- Weakness: distance walked no more than 50 feet (Addendum item XG1), fully dependent walk in corridor (G1da), fully dependent for locomotion on unit (G1ea), unable to balance standing without assistance (G3a), unsteady or needs support to balance sitting (G3b).
- Exhaustion: unable to finish normal daily activities (Addendum XJ1), dyspnea during normal activity or at rest (Addendum XJ2), shortness of breath (J1l).
- Slowness: slow performing tasks (G8c).
- Low Activity: reduced social interaction (E1p), periods of lethargy (B5e).

The Cronbach's Alpha statistic of 0.65 demonstrated that these fourteen items did frequently occur together.[33] We also checked for internal consistency using principal component analysis, a method that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables (principal components). Either

---

[32] Bandeen-Roche, K., Xue, Q.L., Ferrucci, L., Walston, J., Guralnik, J.M., Chaves, P., Zeger, S.L., Fried, L.P. Phenotype of frailty: characterization in the women's health and aging studies. J Gerontol: Biological Science, 61(3):A262-A266, 2006.

[33] The test performed was Cronbach's Alpha, applied to 5,222 residents in the derivation sample after having excluded those a) receiving hospice care, b) receiving ventilator/respirator care, c) comatose, d) with end-stage disease, or e) fully dependent in bed mobility.

as fourteen individual items or combined into single indicators for each of the five areas, these fourteen MDS items were associated as a single principal component.[34]

In the same manner as the Bandeen-Roche frailty measure, we formed a six-level MDS-based frailty scale (0 to 5) by counting the presence of the five areas (i.e., counting one point for each area in which any of the characteristics were present). We tested this scale for convergent validity by correlating it with several measures of physical function, psychological function, social function, pain, and prospective mortality. In all cases, the MDS Frailty Scale correlated significantly and in the theoretically-expected directions with these conceptually-related health status measures[35] (see Table 4-16).

Table 4-16. Spearman Correlation of MDS Frailty Scale with Health Status Measures

Type of scale	Name of Scale	Correlation with MDS Frailty Scale*
Physical Function	ADL scale	0.44
	ADL Hierarchy	0.45
	Barthel Index	-0.45
Psychological Function	Cognitive Performance Scale	0.20
	Communication Scale	0.18
	Depression Rating Scale	0.13
Social Function	Social engagement scale	-0.21
Pain	Pain Scale	0.09
Prospective mortality	CHESS**	0.28
	PSI***	0.48

* All correlations statistically significant ($p<.05$).
** Changes in Health, End-Stage Disease, and Symptoms and Signs, see Section 4.5.4
*** Personal Severity Index, see Section 4.5.4.

The frequency of the MDS Frailty Scale in the full STRIVE derivation sample (including those excluded during development of the measure) is shown in Table 4-17. Most (68%) of these residents had difficulty in one or two areas; few (less than 4%) had difficulty in four or more frailty areas.

Table 4-17. Frequency of MDS Frailty Scale

Frailty Scale	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	862	13.4	862	13.4
1	2209	34.2	3071	47.6

[34] Principal component analysis performed on same sample as for Cronbach's Alpha, in prior footnote.
[35] Tested by Spearman Correlation, performed on the same sample as for Cronbach's Alpha, in prior footnote.

2	2136	33.1	5207	80.7
3	1023	15.9	6230	96.5
4	206	3.2	6436	99.7
5	18	0.3	6454	100.0

We conducted the preliminary research on the potential usefulness of the Frailty Scale. We recommend that additional research be done in this area to fully evaluate its potential as a criterion for future revisions. (See Sections 4.7.3 [Special/Clinically Complex] and 4.8.3 [Tertiary Splits]).

4.5.4 Other New Items and Scales

In Section 4.5 we provide a list of the case-mix relevant items included in the STRIVE Addendum. As well, we included in the STRIVE analytic database a variety of scales, i.e., combinations of items known to validly represent some dimension of assessment. These included:

- ADL Hierarchy[36] – a seven-category summarization of the MDS 2.0 ADL measures;
- Cognitive Performance Scale[37] – a seven-category scale used in RUG-III;
- Augmented Cognitive Performance Scale – a modification of the CPS, using the new item on procedural memory, which is more sensitive to early losses of cognition[38];
- Severe mental illness indicator – an indicator of the presence of any of 11 specific psychiatric problems, based on diagnoses and symptoms;
- 18 Resident Assessment Protocols (RAP) triggers for the MDS 2.0 indicating the presence or risk of a problem, for example, a fall;
- 17 Clinical Assessment Protocols (CAP) triggers,[39] algorithms derived by the interRAI group to identify individuals with a problem in a domain (such as incontinence) that can be the target for successful intervention;
- Social Engagement Scale[40] – an indicator of problematic involvement in activities and unsettled relationships;

---

[36] Morris JN, Fries BE, Morris SA. "Scaling ADLs Within the MDS" *J Geront: Medical Science* 54A(11):M546-553, (November) 1999.
[37] Morris JN et al., 1994, *op cit.*
[38] The scale is formed by following the logic of the original Cognitive Performance Scale, except that the values of the Impairment Count and the Severe Impairment Count are increased by one if the resident has procedural memory problems, an item included in the STRIVE Addendum..
[39] Morris JN, Fries BE, Bernabe R et al. *interRAI Clinical Assessment Protocols*. (Washington DC: interRAI). In press, 2010.
[40] Mor V, Branco K, Fleishman J, Hawes C, Phillips C, Morris J, Fries B. "The Structure of Social Engagement Among Nursing Home Residents" *J. Geront: Psychological Science* 50B(1):P1-P8 (Jan.) 1995.

- CHESS[41] – a measure identifying individuals at serious risk of physical decline;
- Body Mass Index – a measure set to the weight (in kilograms) divided by the square of height (in meters);
- Barthel ADL Index – an alternate ADL summary measure;
- Patient Severity Index – an experimental index predicting early death.[42]

In addition, we worked in STRIVE to develop measures to identify persons who were frail or instable (See Section 4.5.3).

Throughout the analysis to derive RUG-IV, we considered all of the new items and scales in the following approaches:

- As additional criteria to identify hierarchy categories ,e.g., to add resource intensive issues such as infection isolation to the Extensive Services category, or as a qualifier for a criterion, such as "tube feeding with BMI<16";
- As criteria to define a new hierarchy category, e.g., to identify a Violent Behavior group to be placed in the hierarchy either just above or below the current RUG-III Behavior category;
- As a tertiary split for a category; secondary splits were almost always best based on the ADL Index;
- As a criterion for case-mix classification not otherwise noted. We systematically examined, at several points in the derivation of RUG-IV, whether any of the long list of MDS 2.0 items, Addendum items, or new scales were able to explain the residual WWST, after adjusting for the then-current RUG-IV system version. We examined the existence of a positive relationship with resource intensity. Although rare, there were a few instances where this "broadside" approach provided clues to items worthy of further investigation, such as the identification of infection isolation as an indicator of high resource intensiveness.

As discussed in the following sections, few of these new items and scales proved useful, although they were routinely tested.

[41] Hirdes JP, Frijters D, Teare G. "The MDS-CHESS Scale: A New Measure to Predict Mortality in Institutionalized Adults." *J Geriat Soc* 51(1):96-100 (2003).
[42] Morris J, personal communication, 2008.

4.6 Specific targets identified for improvement

4.6.1 Pre-admission services

As discussed in Section 3.3, one explicit goal of this project was to determine whether there were significant differences in resource use between residents who received specific services before nursing facility admission compared with those that used these services after admission. The seven services included in RUG-III are: ventilator/respirator; tracheostomy care, suctioning, oxygen therapy, transfusion, IV medication, and parenteral/IV feeding. For each of these services, the MDS 2.0 records services in the past 14 days. As a result, residents assessed within the first 14 days of their nursing facility stay will have services recorded that occurred before admission, for example, in a prior hospital stay. The STRIVE Addendum was designed to differentiate whether each of six services were provided before or after admission to the nursing facility; the "look-back" period was 7 days.

In the following, it should be noted that the services recorded on the STRIVE MDS were supplemented by information in proximate MDSs, using the RPT overlay discussed in Section 3.2. Comparing the RUG-III distributions in the STRIVE sample and in the national distribution, there was a lower prevalence of those with an extensive service. For example, for Medicare Part A patients in the sample, 25.4% were in the Extensive or Rehabilitation Plus Extensive categories in contrast to 36.5% in Part A Claims data. However, when STRIVE residents were classified according to extensive services that were indicated on either the STRIVE MDS or the RPT, the prevalence of residents who had an extensive service rose to 32.9% among residents who were Medicare Part A. Since this is close to the prevalence in the Part A Claims data, we decided to alter the overlay of the extensive services to count extensive services that were indicated on ***either*** the STRIVE MDS or the associated RPT for that person.

We performed several analyses to investigate the appropriate measure, associated with resource use, to use in RUG-IV.

The first analysis focused on residents assessed within seven days of admission to the nursing facility, here called "early stayers." The restriction to early stayers was placed to examine the possibility that early stayers may be more expensive than residents later in their stay, regardless of whether they received an extensive service. Due to the small percentage of early stayers in the sample, the analysis was performed on the full STRIVE sample, resulting in 802 residents for analysis. Our approach examined resource use (nursing plus therapy WWST) for early-stay residents receiving one of the following six services:

- Ventilator/respirator
- Tracheostomy care
- Suctioning
- Oxygen therapy

- Transfusion
- IV medication

We examined when these services were provided in each of the following venues of service:

- Pre-admission only;
- Post-admission only;
- Pre- and post-admission;
- Services only identified in the MDS 2.0, with a look-back period of 14 days; or
- None of these services.

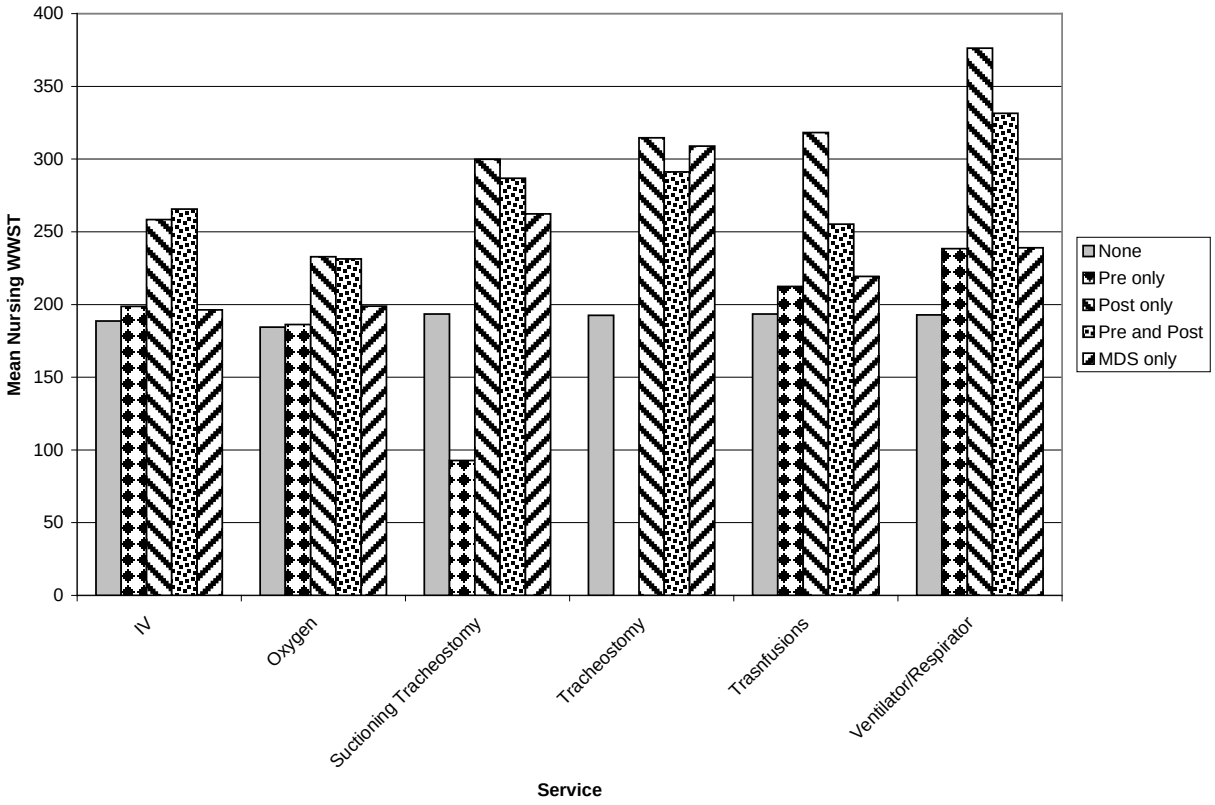
Analyses evaluated the impact of the venue of these services on nursing WWST adjusted for RUG-III group.[43]

Among early stayers, we found the following:

- Residents who received these services post-admission had quantitatively greater (16% to 60%) resource use than those who received these services pre-admission only. After adjusting for the RUG-III group, those receiving IV medications, oxygen therapy, and suctioning had statistically significantly higher WWST than those receiving these services pre-admission only (see Figure 4-6).
- Residents who received a service pre-admission only were usually comparable in resource intensity to those who did not receive that service. Differences in resource intensity for those who received a pre-admission service only were not significantly different from those who did not receive that service.

[43]. We adjusted for RUG-III 53 group by constructing a regression model predicting nursing WWST using RUG-III 53 groups (modeled as a categorical variable) and venue of service (modeled as a categorical value). The model was used to output least-squares means for each venue of service within each RUG-III 53 group and differences between these RUG-III adjusted means for venue of service were tested using the Type-III F-test for differences in means

Figure 4-6. Mean Nursing WWST of Services Among Early Stayers, Adjusting for RUG-III Group



These results were mirrored when we compared resource use either by venue of service or by day of stay for the first 20 days of nursing facility care.

Focusing on the four services involved in defining the RUG-III Extensive category – specifically ventilator/respirator; tracheostomy care, suctioning, and IV medication – we compared the RUG-III classification based on post-admission services rather than, as in the standard RUG-III system, based on any presence of the services. With this change, the variance explanation of Nursing WWST among early stayers increased from 22.9% (with classification based on the presence of any service) to 27.2% (with classification based only on post-admission services). Most of the increase in variance explanation was due to using post-admission IV medication. An examination of WWST residuals from the RUG-III groups[44] showed that classification based upon post-admission services resulted in residents being classified into groups with more similar resource use, i.e., smaller absolute residuals, and reduced overall group heterogeneity. In addition, using post-admission extensive services for classification resulted in higher

[44] Residuals were computed after fitting the regression model (without sample weights): WWST=a+b*RUG, where WWST is the nursing WWST and RUG is a categorical variable representing the 53-group RUG-III system. Residuals were computed for each observation as the nursing WWST for that observation minus the mean nursing WWST for the assigned RUG-III group.

resource use (Nursing WWST) for the Rehabilitation Plus Extensive and Extensive groups.

As a result, an initial change made in developing RUG-IV was to base the presence of these six services within the first seven days of the resident's admission on the presence of the service only after admission to the nursing facility.

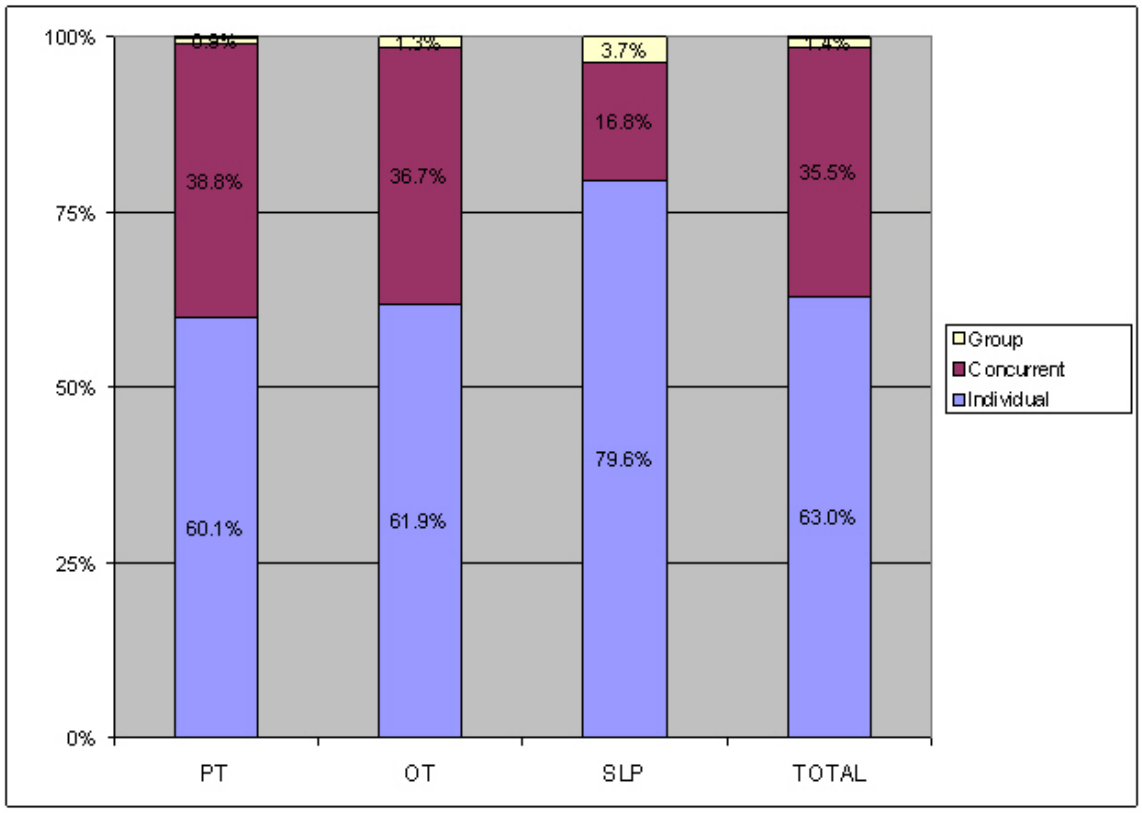
We were unable to evaluate whether there was a difference in the WWST for pre- and post-admission parenteral/IV feeding, as this concept was not incorporated into the STRIVE Addendum (we did collect the number of days post admission that it was given, but there was a high percentage of missing data). Therefore, parenteral/IV feeding was not restricted to post-admission use.

4.6.2 Employing Concurrent and Other Therapy

As discussed in Section 2.2, the RUG-III system classifies residents using the time they spend in three disciplines of therapy: physical, occupational, and speech therapy. However, it does not identify the modality of the therapy.

Figure 4-7 presents the percent of therapy time that was individual, concurrent involving residents receiving different services, and group involving residents all engaged in the same activity. Overall in the STRIVE data, 35.6% of all therapy time was concurrent. By discipline, PT and OT had larger amounts of concurrent therapy (38.8% and 36.7%, respectively) than SLP (16.8%).

Figure 4-7. Distribution of Therapy Time, by Discipline and Type



NOTE: PT=Physical Therapy, OT=Occupational Therapy, SLP=Speech and Language Pathology

There was a positive relationship between total therapy time and the percentage of therapy given concurrently: alternately stated, when more therapy was given, it was more likely to be concurrent time. For example, in the Ultra High Rehabilitation groups, approximately 51.2% of therapy time was concurrent therapy, while in the Low Rehabilitation groups this percentage was 22.8%.[45]

We also examined whether membership of the facility in a large nursing home chain affected concurrent therapy. For PDA days only, 50.0% of the therapy was concurrent in SNFs that were part of a large chain, versus 43.8% for all others.

There was very little group therapy time (1.4% overall, and primarily SLP where it represented 3.7% of the time collected). We decided that group therapy time would not be allocated; it would, however, continue be capped at 25% of total therapy time by discipline (see Section 4.3.4) as it was in RUG-III classification.

[45] These numbers are for the PDA days only, where it is believed that the concurrent therapy is most accurately defined. However the results do not differ substantially if we use all the therapy time collected. Also, the calculations were performed before allocating concurrent time, as discussed later in this section.

We examined the effects of classifying residents for concurrent therapy using five alternate methods, by contrasting the characteristics of RUG-IV[46] computed using each:

- Concurrent therapy fully allocated using the "time slice" methodology ("100% Allocation");
- Allocating 50% of the concurrent therapy using the "time slice" methodology ("50% Allocation");
- Placing a cap on the unallocated concurrent therapy, i.e., no greater than 25% of the total unallocated time, by discipline ("Capped Allocation");
- Allocating concurrent therapy assuming a specific "group size" (n=2) ("Fixed Group Size");
- No allocation of concurrent time.

The contrast was made by comparing for each option: variance explanation,[47] therapy and Nursing WWST means for the RUG groups, and distribution of residents across the rehabilitation categories. In the analysis, we combined individuals in the Rehabilitation and the Rehabilitation Plus Extensive groups.

Compared to the results when concurrent therapy was unallocated, the first four options above resulted in mean WWST that were less compressed, i.e., had a larger range in WWST across the RUG-IV rehabilitation groups. The four options had WWST ratios of the highest to lowest Rehabilitation groups ranging from 3.6 to 4.2, compared to a ratio of 3.4 for the unallocated methodology (see Table 4-18). The four alternative methods also had higher rehabilitation group means of nursing and therapy WWST (data not shown). Less compression in the mean WWST provides evidence that classifying using allocated therapy identifies residents with higher WWST.

Table 4-18. Comparison of Characteristics of Five Methods to Allocate Concurrent Therapy

Methodology for Allocating Concurrent Therapy	Ratio of mean Nursing + Therapy WWST in the highest to lowest Rehabilitation group	RUG-IV Variance Explanation (nursing+ therapy WWST)
No (zero) allocation	3.4	41.2%
100% allocation	4.2	41.8%
50% allocation	4.2	41.9%
Capped allocation	3.9	41.7%
Fixed group size	3.6	41.7%

[46] Comparisons were made using RUG-IV Version 7 and on the entire sample (N=9706) using sample weights.

[47] Variance explanation was the R^2 statistic for the regression: WWST=a+b*RUG, where WWST was the nursing plus therapy WWST and RUG was the categorical variable of all groups in the RUG-IV Version 7 system. Analysis was performed on the entire sample (9,706) using sample weights.

Among the four new methods of allocating concurrent therapy, the variance explanation was similar (from 41.7% to 41.9%), but all were higher than that seen with no allocation (41.2%).

Based on both variance explanation and WWST range (WWST ratio of high to low group), some variety of allocation is preferable to no allocation. Also, the four allocation methods did not show great differences in statistical properties. We therefore selected the 100% allocation as the most appropriate because of its logical appeal, allocating the actual concurrent staff time over the actual number of residents receiving concurrent therapy. The other allocation methods either allocate only a part of the concurrent time (50% allocation and capped allocation) or allocate too much concurrent time (allocating to a fixed group size of 2 residents when there are more than 2 residents in a session). In the final RUG system, using this allocation method for concurrent therapy resulted in about a 2% drop (from 15.5% to 13.2%) in the percentage of residents classified into the Rehabilitation and Rehabilitation Plus Extensive categories. In particular, smaller percentages were observed in upper rehabilitation groups – Ultra-High Rehabilitation (3.6% to 1.1%) and Very High Rehabilitation (3.8% to 2.6%) – and an increase was observed in a lower rehabilitation group, Medium Rehabilitation (3.9% to 5.0%) – there was no change in the lowest rehabilitation group, Low Rehabilitation, which was infrequently seen.

4.6.3 Special Units

Prior research has shown that, after adjusting for case mix, resource utilization for residents on special care units (SCU) is the same as for other residents in the same facility, but higher than for residents with the same characteristics in facilities without a special care unit.[48] We tested this to assure data on resource use in special care units could be combined with the rest of the STRIVE data.

After facilities were selected for data collection, the STRIVE project team worked with facility administration to determine the types of units at each facility. Ten different types of units were identified. We describe in the STRIVE Phase I Report[49] the protocols used to select facilities from 5 sampling strata that include special units so that there was a substantial representation of SCUs. The final sample included the following types of units (with percentage of the STRIVE sample[50]): long term care (27.7%), skilled nursing (49.4%), rehabilitation (9.5%), Alzheimer's (3.8%), open mental health (0.7%), locked mental health (0.8%), ventilator (3.4%), traumatic brain injury (0.3%), AIDS/HIV (3.1%), and mixed (1.4%) units. The SCUs involved in this analysis were those with differentiated services and sufficient sample size to permit analysis. These included the following:

---

[48] Mehr DR, Fries BE. "Resource Use on Alzheimer Special Care Units" *Gerontologist* 35(2):179-184 (Apr.) 1995.

[49] STRIVE Phase I Report, pp. 7 -12.

[50] Percentages are based on 9,707 STRIVE residents who have a full STM (48 hours), classify on RUG-III. Note that the analysis described in this section were performed prior to the analysis leading to RUG-IV Version 7 and thus includes the single high WWST outlier observation.

- Alzheimer's including Dementia (14 facilities in 8 states);
- Ventilator/Respirator (12 facilities in 8 states);
- Traumatic Brain Injury (TBI) (1 facility);
- AIDS/HIV (2 facilities in 1 state).

To reproduce prior analyses,[51] for each of the four types of SCU, we identified whether each STRIVE resident had the following:

- Medical condition (e.g., dementia or Alzheimer's Disease);
- Presence in SCU was for that condition;
- Was present in a facility that had that type of SCU.

For each of the four types of SCUs, this produced a three-digit coding scheme, one digit for each of the three dimensions just listed, and with six possible unique combinations. See Table 4-19. [52]

Table 4-19. Six Combinations of Ventilator/Respirator Special Care Unit Coding

Code	With Condition	On SCU	Facility has SCU
111	Yes	Yes	Yes
101	Yes	No	Yes
100	Yes	No	No
011	No	Yes	Yes
001	No	No	Yes
000	No	No	No

To contrast resource use on each type of SCU across the six possible SCU codes (representing the possible combinations of condition, unit, and facility in Table 4-19), we calculated a mean Nursing WWST adjusted for RUG-IV.[53] As some of the sample sizes were small, the full sample (derivation and validation) was used in these analyses.

The results for each of the four types of SCUs are shown in Figures 4-8 through 4-12, with each histogram bar denoting the mean for an SCU code. For comparison,

---

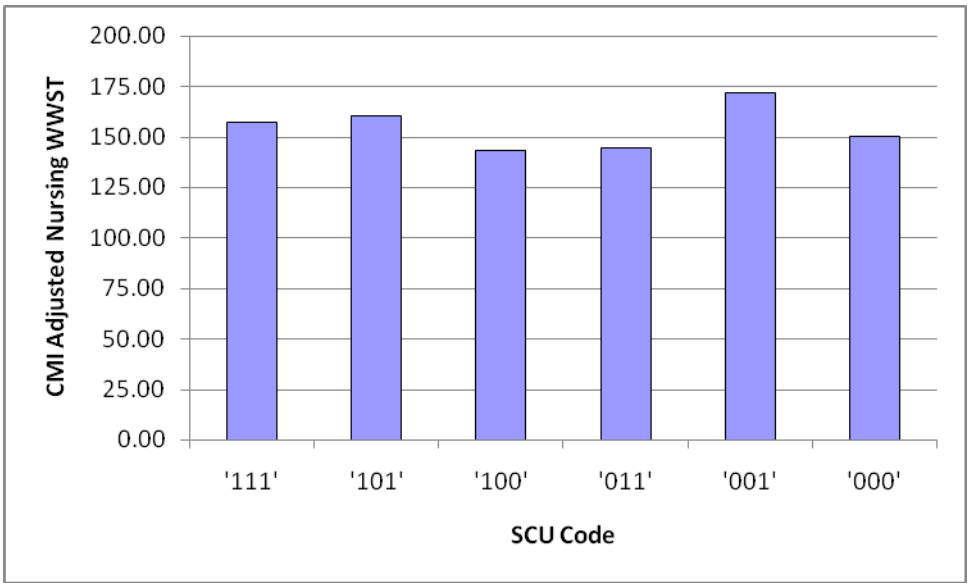
[51] Mehr DR, Fries BE, *op cit.*

[52] Both unit type and MDS items define each of the SCU groups. Resolved ICD-9 MDS variables for Alzheimer's/Dementia, TBI, and AIDS/HIV use MDS pick lists plus codes from Section I3 code listings. Ventilator/Respirator is MDS item P1al. For comparison, Alzheimer's/Dementia is also based on cognitive performance scale scores (CPS). Residents with a score of greater than or equal to 3 are defined as cognitively impaired or having Alzheimer's/Dementia.

[53] For each RUG-IV (Version 2) group, we computed the Case Mix Index (CMI) as the ratio of the overall nursing cost mean for each RUG-IV group and the overall mean WWST. A regression model was constructed to predict nursing WWST from the CMI for each observation and the predicted value was taken as a RUG-IV CMI adjusted nursing WWST for each observation. The significance of any adjusted WWST differences between the SCU codes was tested using the z-statistic for means.

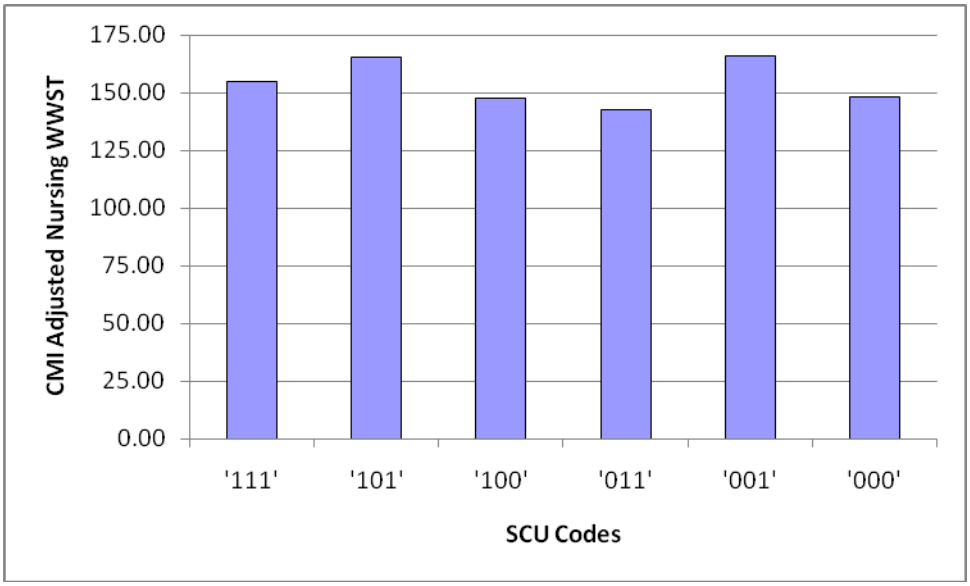
each graph also shows the overall adjusted nursing WWST mean for all residents as a dashed line (148.3, for N=9,707). Figures 4-8 and 4-9 show the results when Alzheimer's/dementia is indicated either by diagnosis or by CPS score. Both graphs (Figures 4-8 and 4-9) show the same pattern for each set of SCU codes: there is little difference between CPS and Diagnosis groups. Overall, the six groups of residents have relatively similar WWST (means range: 142.5 to 172.2) and the means are very similar to the overall 148.3 WWST mean for all residents. In particular, there is no statistical evidence that residents on an SCU (SCU codes 111 and 011) have substantially higher WWST that those not on an SCU.

Figure 4-8. Adjusted Nursing WWST by Dementia Special Care Unit Code* – Dementia derived from Diagnoses



*See Table 4-19 for codes.
Blue bars represent mean CMI-Adjusted Nursing WWST
Dashed line represents overall mean WWST for all residents

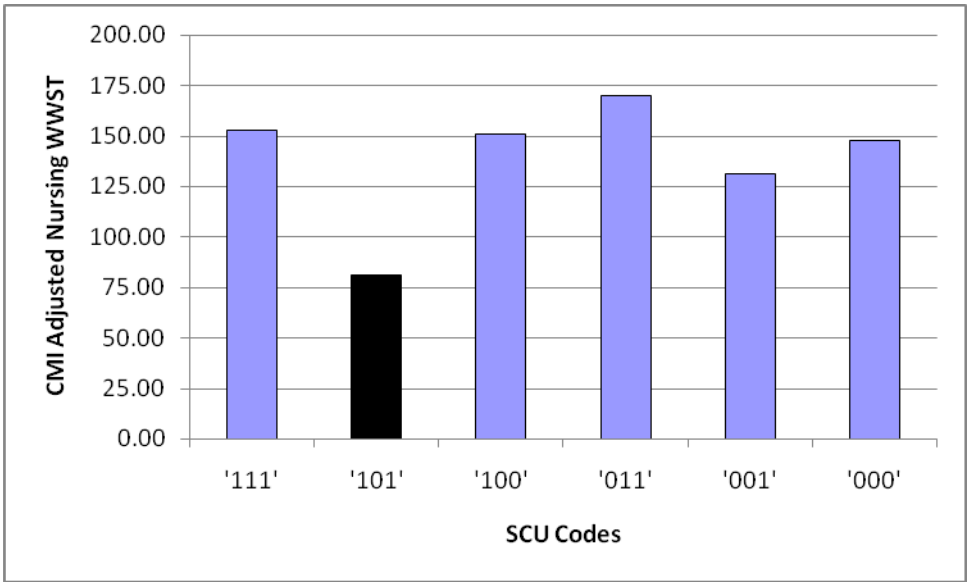
Figure 4-9. Adjusted Nursing WWST by Dementia Special Care Unit Code* – Dementia derived from Cognitive Performance Scale



*See Table 4-19 for codes.
Blue bars represent mean CMI-Adjusted Nursing WWST
Dashed line represents overall mean WWST for all residents

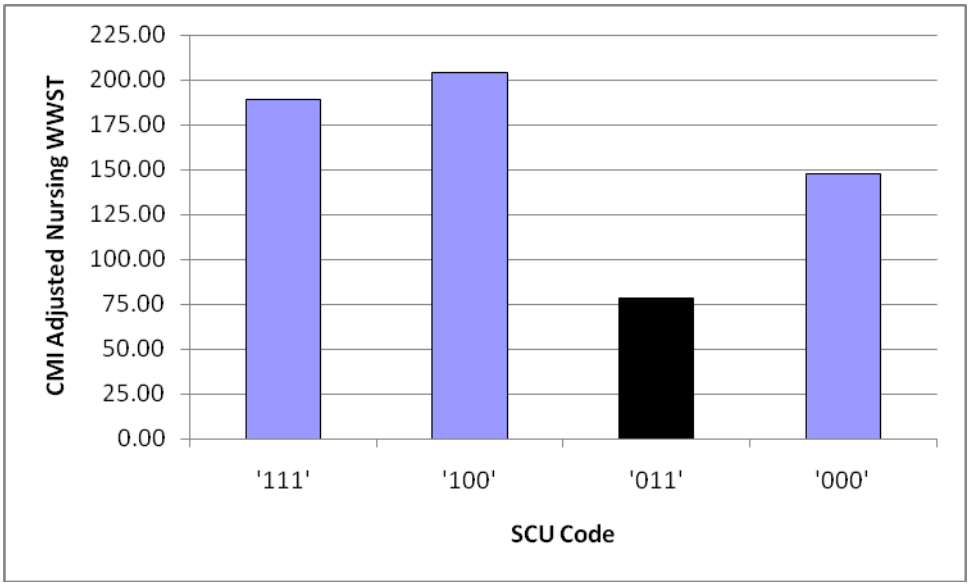
The remaining graphs, Figures 4-10 through 4-12, show similar results for the other three types of SCUs: little substantial difference between the SCU code categories. Note that black bars denote groups with less than 20 residents; these means can be expected to be unstable due to small numbers.

Figure 4-10. Adjusted Nursing WWST by Ventilator or Respirator Special Care Unit Code*



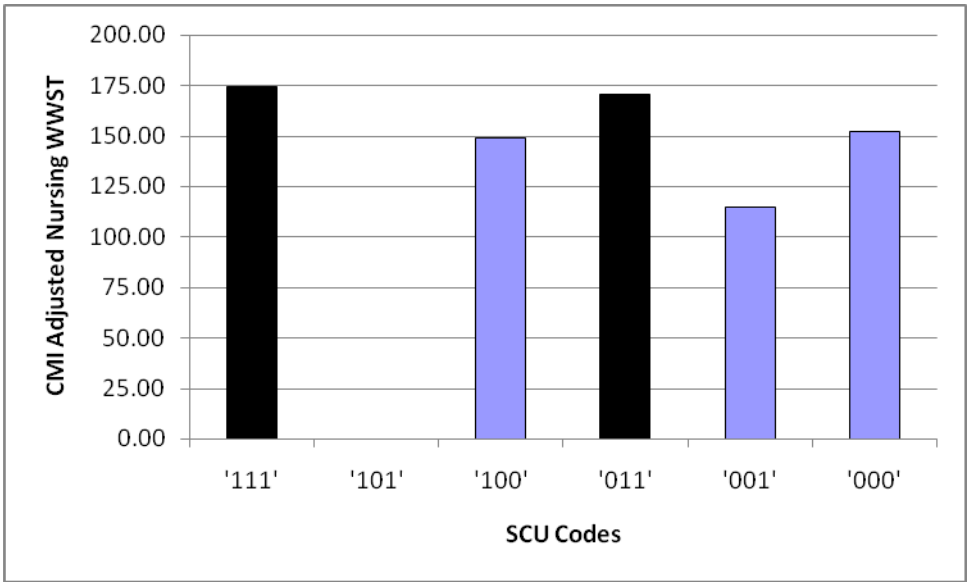
*See Table 4-19 for codes.
Blue bars represent mean CMI-Adjusted Nursing WWST
Black bars represent mean CMI-Adjusted Nursing WWST with a sample size < 20
Dashed line represents overall mean WWST for all residents

Figure 4-11. Adjusted Nursing WWST by AIDS/HIV Special Care Unit Code*



*See Table 4-19 for codes.
Blue bars represent mean CMI-Adjusted Nursing WWST
Black bars represent mean CMI-Adjusted Nursing WWST with a sample size < 20
Dashed line represents overall mean WWST for all residents

Figure 4-12. Adjusted Nursing WWST by Traumatic Brain Injury Special Care Unit Code*



*See Table 4-19 for codes.
Blue bars represent mean CMI-Adjusted Nursing WWST
Black bars represent mean CMI-Adjusted Nursing WWST with a sample size < 20
Dashed line represents overall mean WWST for all residents

Based on this analysis, it appears that WWST for residents in special care units or residents of facilities with SCUs are not substantively different from nursing facility residents in general. This finding, in combination with the complexities of determining an appropriate adjustment methodology led to the decision not to include any special care unit adjustments in the STRIVE analyses.

4.7 Update ADL Index

Extensive of case-mix research show that Activities of Daily Living (ADLs) serve a central role in predicting the cost of care. This result is mirrored in the STRIVE data, where RUG-IV, like RUG-III, uses a scale measuring ADLs to permit identification of residents with similar levels of physical function. A scale, rather than individual items, allows for an efficient summarization of multiple measures of performance and support. The RUG-III scale (the RUG-III ADL Index) uses the MDS measures of performance and support for each of four ADLs, specifically, toilet use, eating, transfer, and bed mobility, to develop component values which are then summed. In RUG-III, the resulting Index is used to split each of the major hierarchical categories except Extensive Services. It is also used as part of the qualification criteria for many of the hierarchical categories and for some specific criteria within these categories.

The RUG-III ADL Index was one of the best individual predictors of STRIVE WWST. Still, we investigated whether improvements could be made that would increase variance explanation and fit better across the whole scale range. We decided

to retain the four component ADLs used in RUG-III, as these represent functions lost later in life. In addition, these functions are less prone to facility effect. For example, bathing is not used since some facilities, as a practice, do not permit residents to bathe themselves without assistance, regardless of their ability. We also retained the requirement that "supervision" would not be differentiated from "independent" as it would be easy for a facility to provide minimal supervision if it would achieve a higher case-mix classification.

To make the new ADL index easier to understand, we changed its base value to zero for a person independent in all four ADLs; the RUG-III ADL Index for such a person is 4.

To improve the correspondence of the ADL scale with resource use, we considered over 30 alternatives. The alternatives assigned differing weights for combinations of the performance and support items for each component ADL. We also considered several approaches to incorporate artificial feeding items into the eating ADL component, such as differing combinations of tube feeding, parenteral/IV feeding, percentage of nutrition supplied, and amount of fluid intake. For each alternative, we evaluated the effect of each combination using clinical and statistical criteria such as variance explanation (R^2) and "linearity," i.e., that each additional "point" on the constructed scale is associated with an approximately equal increase in WWST (since linear indexes are more easily interpreted).

For each component ADL score, the best RUG-IV index retained the same five-point range used for most RUG-III scale components, although the responses are coded differently. For the RUG-IV ADL Index, referred to hereafter as the "ADL-IV Index," component values start at 0 and run from 0 to 4, whereas they ran from 1 to 5 for the RUG-III Index. This change simplifies interpretation of the Index while changing none of its properties. As well, the new ADL-IV Index Eating component scale runs from 0 to 4, two points more than for the RUG-III ADL Index, resulting in greater index range.

The ADL-IV Index is the sum of component scores for four ADLs: Bed Mobility, Transfer, Toilet Use, and Eating. Each component score is derived from information from both the Self-Performance and Support MDS items for all four ADLs. This is a minor change from the RUG-III ADL Index, which did not use the support item for Eating. Many specific combinations of MDS items remain the same in RUG-IV as in RUG-III, although the corresponding component scores can be slightly different.

As with the RUG-III ADL Index, in the ADL-IV Index bed mobility, transfer, and toilet use are treated identically. Eating, however, continues to be scaled differently from the other three ADLs. In addition, we considered different definitions of artificial feeding. In RUG-III, all residents receiving artificial feeding had the same component score. In the ADL-IV Index, artificial feeding is not used to determine the Eating component value, since setting a constant score for all residents receiving artificial feeding reduced the explained variance in WWST below that achieved by using only a combination of Eating Performance and Support items.

Tables 4-20 and 4-21 provide the details of the coding for each component of the RUG-III and ADL-IV Indexes. To compute the ADL-IV Index, component scores for bed mobility, toilet use, transferring, and eating are summed. Higher scores represent greater functional dependence and more need for assistance.

Table 4-20. Scaling ADLs for the ADL-IV Index: Bed mobility, transfer, toilet use*

	RUG-III			RUG-IV		
	Support			Support		
Performance	**None/Set up**	**1-person**	**2-person**	**None/Set up**	**1-person**	**2-person**
Independent/ Supervision	1	1	1	0	0	0
Limited Assistance	3	3	3	1	1	1
Extensive Assistance	4	4	5	2	2	4
Total Dependence	4	4	5	3	3	4

*Note: RUG-IV scaling starts as 0, so shading adjusted to match similar scale values on RUG-III

Table 4-21. Scaling ADLs for the ADL-IV Index: Eating*

	RUG-III			RUG-IV		
	Support			Support		
Performance	**None/Set up**	**1-person**	**2-person**	**None/Set up**	**1-person**	**2-person**
Independent/ Supervision	1	1	1	0	2	2
Limited Assistance	2	2	2	0	2	2
Extensive Assistance	3**	3**	3**	2	3	3
Total Dependence	3**	3**	3**	2	4	4

*Note: RUG-IV scaling starts as 0, so shading adjusted to match similar scale values on RUG-III
** Including Parenteral/IV/Tube Feeding

The resulting ADL-IV Index has a range of 17 points (from 0 to 16), greater than the RUG-III ADL Index range of 15 points (from 4 to 18). This allows greater distinction in physical function. Crosswalks between these two indexes are provided in Tables 4-22 and 4-23. The improvements resulted in better variance explanation of Nursing WWST: an increase to 11.1% for the ADL-IV Index from 10.5% for the RUG-III ADL Index.[54] As well, the relationship between ADL-IV index values and mean Nursing

[54] Variance explanation was based on regression models of the form NURSING WWST = a+b*RUG-IV+c*ADL, where RUG-IV represents a categorical variable indicating the RUG-IV Version 2 major

WWST has better fit across the full range of the index (see Figure 4-13) and has a more constant percentage of residents at each index level (see Figure 4-14): both these features of the ADL-IV Index can be expected to increase the variance explanation of the RUG-IV system. Also Figure 4-13 indicates the ADL-IV index has close to a linear relationship.

Table 4-22. Crosswalk from the RUG-III ADL Index to the ADL-IV Index

RUG-III ADL Index	Number of Observations*	Mean ADL-IV Index	Minimum	Maximum
4	1185	0.02	0	2
5	10	1.80	0	2
6	298	1.03	0	4
7	128	2.31	2	5
8	240	2.21	1	5
9	142	3.36	2	7
10	485	3.33	2	6
11	268	4.53	3	7
12	358	5.49	3	10
13	653	6.58	4	11
14	437	8.57	5	12
15	569	10.39	6	13
16	669	12.50	8	14
17	496	14.18	10	15
18	516	15.56	12	16

* Total number of observations = 6,454 (the full derivation sample)

Table 4-23. Crosswalk from the ADL-IV Index to the RUG-III ADL Index

ADL-IV Index	Number of Observations*	Mean RUG-III ADL Index	Minimum	Maximum
0	1195	4.03	4	6
1	275	6.04	6	8
2	324	7.55	4	10
3	534	9.63	6	12
4	296	10.38	6	13

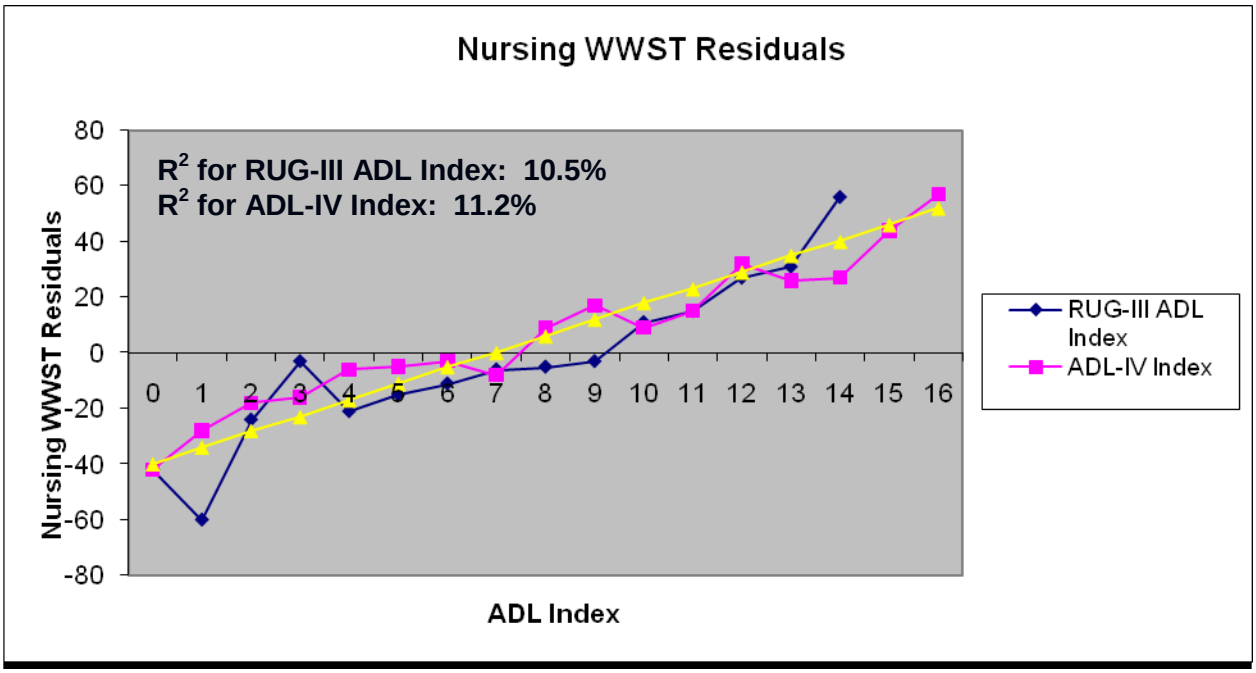
---

hierarchical categories and ADL is an ADL Index treated as a numeric variable (linear). Separate models were run for the RUG-III ADL index and the ADL-IV index. The regression analyses were performed, unweighted, on the derivation sample, excluding residents classified into the Extensive Services category (as research showed that this category was not very sensitive to ADL – see Section 4,8.1), for a total sample size of 6,165 observations. For further analyses, residual scores (observed minus predicted WWST) were calculated for both models.

ADL-IV Index	Number of Observations*	Mean RUG-III ADL Index	Minimum	Maximum
5	330	11.49	7	14
6	549	12.78	9	15
7	175	12.73	9	15
8	315	13.78	12	16
9	241	14.17	12	16
10	297	14.82	12	17
11	188	15.10	13	17
12	455	15.94	14	18
13	211	16.04	15	17
14	344	16.57	16	18
15	406	17.45	17	18
16	319	18	18	18

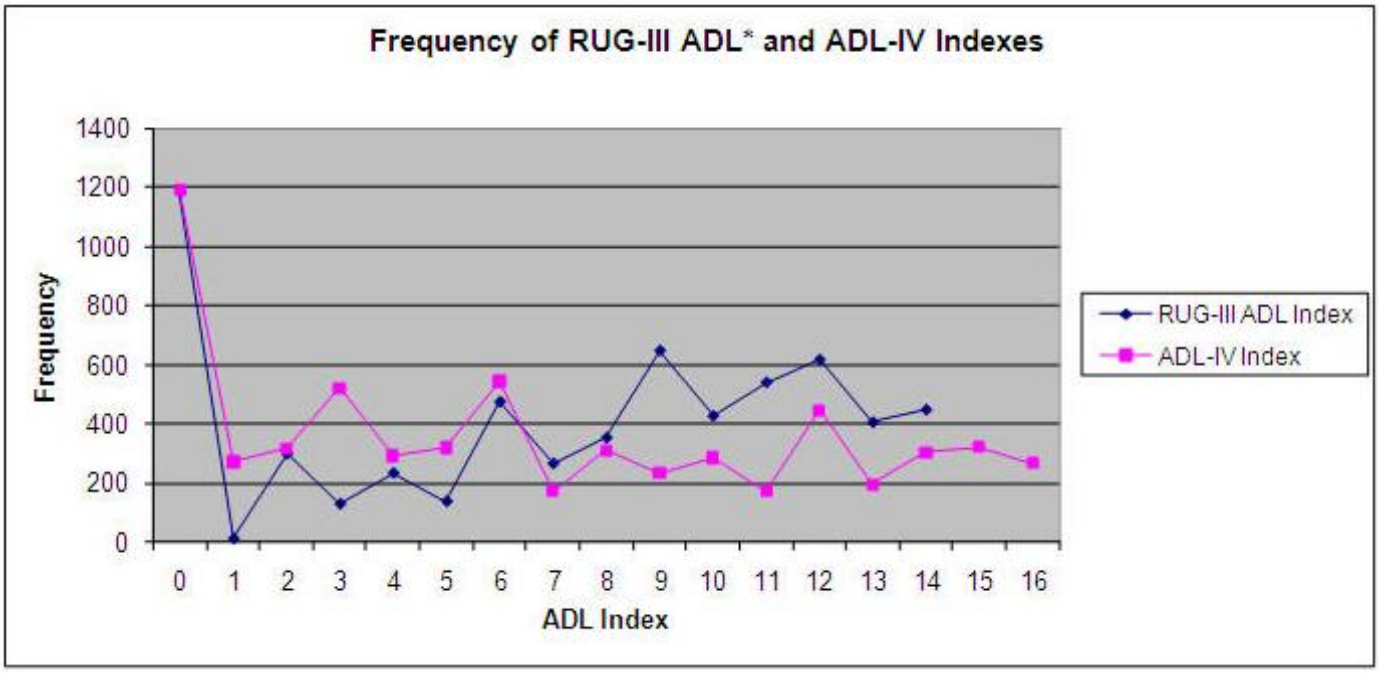
* Total number of observations = 6,454 (the full derivation sample)

Figure 4-13. Nursing WWST Residuals (adjusted for RUG-IV Category), by RUG-III ADL and ADL-IV* Indexes



*Note: The RUG-III ADL Index is shifted in this graph from a range 4-18 to the range 0-14 to make comparison here easier. The yellow linear is a linear fit for ADL-IV.

Figure 4-14. Frequency of RUG-III ADL and ADL-IV Indexes



*Note: The RUG-III ADL Index is shifted in this graph from a range 4-18 to the range 0-14 to make comparison here easier.

4.8 Revising RUG Hierarchy Categories

The bulk of analysis to refine the RUG system centered on revision of the RUG-III hierarchy categories (see description in Section 2.2) which are subdivided into the individual RUG groups. In the following sections, we describe the analyses performed to determine whether the criteria for each category should be changed. As the RUG algorithm classifies an individual hierarchically, i.e., into the first (highest) category for which he or she qualifies, the decisions made about "higher" groups potentially affect analyses for lower groups. As a result, we made all decisions about the groups sequentially, starting with the highest groups. Many of the categories also involve an ADL Index threshold, which we discuss later in Section 4.8.5.

4.8.1 Rehabilitation and Rehabilitation Plus Extensive Categories

The RUG-III system classifies residents into the Rehabilitation (R) and Rehabilitation Plus Extensive (RE) categories based primarily on the provision of services by three disciplines: physical therapy, occupational therapy, or speech-language pathology services. This decision to use service-based criteria, made in earlier versions of RUGs (and discussed in Section 3.3), was retained. Also, as in prior RUG research, the analysis used measured therapy time for classification rather than

the time reported on the MDS. It is acknowledged that this approach artificially inflates any variance explanation of a dependent variable including therapy times (such as "nursing plus therapy WWST"). As described earlier, in STRIVE the therapy time was adjusted to address problems in data collection (see Section 4.3.4) and concurrent therapy time was allocated (see Section 4.6.2). When RUG-IV is used in practice, however, classification into the R and RE categories would be based on the MDS items describing the number of days and total weekly time (items in MDS 2.0 Section P1b).

The thresholds of rehabilitation time for inclusion in each of the five R and RE categories (for example, the threshold for Ultra High Rehabilitation in RUG-III was "at least 720 minutes per week with a least one discipline 5 days a week and a second discipline at least 3 days a week") were determined in RUG-III based on clinical evidence. This evidence included how therapy was delivered in nursing facilities, since there were no evidence-based criteria indicating "break points" in the continuous distributions seen in practice. While altering thresholds would change the composition of subcategories and change category mean therapy WWSTs, there would be no difference in the overall case mix of a group for residents. For example, consider the effect of changing the threshold for Rehabilitation-High from 325 total minutes, after allocation of concurrent therapy to all persons in the session. While lowering the threshold to, say, 250 minutes would qualify more individuals for the Rehabilitation-High group, the inclusion of these additional persons with therapy times of 250-325 minutes would lower the mean therapy staff time cost and therefore the Case-Mix Index and payment rate for the Rehabilitation-High group. Nationally, the lower rate would exactly balance the increased inclusion, resulting in no change to the average payment. At the facility level, the result of such a change is, therefore, likely to be minimal. Similar results would occur if any threshold were raised, with higher payment being balanced by fewer persons and no change in average payment. Following this logic, we decided not to alter any of these thresholds.

The Rehabilitation Plus Extensive (RE) categories were split out from the other Rehabilitation (R) categories by the presence of extensive services, and an ADL Index threshold, discussed later. In RUG-III, the criterion for this split was the same as that for identifying the RUG-III Extensive Services (E) category: receipt of any of five extensive services including ventilator/respirator, tracheostomy care, parenteral/IV feeding, IV Medication, or suctioning.

For RUG-IV, as in RUG-III, an *a priori* decision was made to use the same extensive service criteria for RE as for Extensive Services (E) (see next section). As we developed RUG-IV and revised the shared extensive service criteria for defining both the RUG-IV RE and E category, we tested the residents who qualified for the Rehabilitation category as to whether there were different amounts of resources provided if they also met the criteria for the Extensive Services category. Specifically, we compared the WWST variance explained (fit) in models that compared residents that qualified for a Rehabilitation category to residents that qualified for the comparable

Rehabilitation Plus Extensive Services category, controlling for ADL group.[55] This test was performed several times while developing RUG-IV for two reasons: a) as the definitions for the Rehabilitation category changed (e.g., using allocated concurrent therapy); and b) as different possible criteria were used for qualification to the Extensive Services category. All the tests supported the need for Rehabilitation Plus Extensive Services categories (see Table 4-24). In the final model, with three criteria for Extensive Services (ventilator/respirator, tracheostomy care, and isolation for infection control) and a Rehabilitation definition based on allocated concurrent therapy time (RUG-IV, version 8, full sample, N=1,989, weighted), residents in the RE category used, on average, 168 Nursing WWST more than residents in the corresponding R category.

Table 4-24. Resource use difference between Rehabilitation and Rehabilitation Plus Extensive Services Categories for alternate Extensive criteria

RUG version	Extensive Criteria	Concurrent Therapy Time	Type of WWST	Difference in WWST	p
1	Ventilator/respirator, tracheostomy care, parenteral/IV feeding, IV Medication, suctioning	Unallocated	Therapy + Nursing	34	<0.0001
3	Ventilator/respirator, tracheostomy care, isolation for infection Control, Septicemia	Unallocated	Nursing	119	<0.0001
8	Ventilator/respirator, tracheostomy care, isolation for infection Control	Allocated	Nursing	168	<0.0001

On the basis of these results, the Rehabilitation and Rehabilitation Plus Extensive categories (and subcategories: Ultra High, Very High, High, Medium, and Low) were determined (see Table 2-1).

4.8.2 Extensive Category

The Extensive Services (E) category identifies those residents who are the most expensive to care for in a nursing facility. As discussed in the prior section, there were five criteria used to identify this category in RUG-III (ventilator/respirator, tracheostomy care, suctioning, parenteral/IV feeding, and IV medication). The same criteria were used to identify the residents in the Rehabilitation category who also had extensive services and would, therefore, qualify for the very high WWST RE categories. For RUG-IV, we sought criteria that would accomplish the same dual purpose, i.e., be criteria for the E category as well as serve as the extensive criteria for the RE category. However, as there were very few individuals both receiving substantial rehabilitation

[55] Run as a regression of WWST = a+b*EXTEN+c*ADL, where EXTEN was the dichotomous variable of qualification for the Extensive group (alternate definitions tested – see text), and ADL was a categorical variable indicating mutually exclusive ranges of the ADL-IV Index.

treatment and an extensive service, our approach was to determine the criteria for the Extensive Services category, then only check that these criteria did not provide aberrant results defining the RE category. It should be noted that once residents are classified into the RUG-III Extensive Services category, other information is used to subdivide the groups, including membership in other categories (special care, clinically complex, and impaired cognition).

All analyses to determine the criteria for the E category were performed on the derivation sample (N=6454) and, based on previous findings that extensive services used post-admission are more resource-intensive than services used only pre-admission, are limited only to post-admission extensive service use (see Section 4.6.1).

We first considered whether the five RUG-III criteria should be retained. As residents with a particular RUG-III extensive service are classified into one of the three E subcategories, we examined Nursing WWST deviations (residuals) from subcategory means for each service, i.e., how well each service "fit" each subcategory. Subcategories were determined only using a count of extensive services, since preliminary evaluation showed that RUG-III use of co-qualification for other hierarchy categories was not effective in subdividing this category (see Section 4.9.1). IV medication was inappropriately placed in Extensive Services, as indicated by the largest negative residuals– these residents who received IV medications were 79.5 (ES1), 5.5 (ES2), and 0 (ES3) Nursing WWST below the subcategory mean (in ES3 all residents received IV medication, but they also received other extensive services). When IV medications was dropped both as a E category criterion and in the count used to subdivide the E category, the variance explanation (R^2) improved from 34.0% to 36.1%,[56] and further improved when IV medication was dropped as a criterion for the R + E groups (to 37.0%).

Receving parenteral/IV feeding also corresponded to large residuals. However, dropping parenteral/IV feeding as a criterion led to mixed results: for the E category, R^2 increased slightly (from 37.0% to 37.3%),[57] but dropping it as a criterion for the RE category led to a marginal loss in R^2 diminished (from 37.3% to 37.2%). Given these relative weak findings, we used the next analysis to decide whether parenteral/IV feeding was an appropriate criterion for Extensive Services.

Residents frequently receive more than one extensive service, so we ranked residents by the most expensive service they received. This type of analysis is independent of any particular service count. We considered each service used as a RUG-III criterion for Extensive Services, ordered in mutually exclusive categories by

[56] Run as a regression of nursing WWST = a+b*RUG, where RUG was the categorical variable of all RUG-IV, Version 1 groups, with the E category split on a count of extensive services; derivation sample (n = 6454).

[57] RUG-IV, Version 1 with the E category split on a count of extensive services; IV medications dropped as an E category qualifier and not used in the services count to split the E category, derivation sample (n – 6454).

descending mean Nursing WWST.[58] To do this, we first selected the criterion that identified the most resource-intensive extensive residents. Then, we considered all the remaining residents (i.e., those not having this first characteristic) and looked for the extensive service associated with the (next) highest WWST. We continued until all five criteria were ranked.

Residents using ventilators or respirators clearly had the highest Nursing WWSTs (see rankings in Table 4-25), followed by tracheostomy care, suctioning, parenteral/IV feeding, and finally IV medication. There was a substantial drop in mean Nursing WWST between the first two criteria between the second and third criteria, and then again between the fourth and the last criteria on this list. On the basis of this analysis, we made the provisional decision to use ventilator/respirator use and tracheostomy care (both post-admission) as criteria for the RE and E category. With only these two criteria, the E category became less populous – it dropped from 8% to 3% of all residents. But this approach also was more efficient at identifying the very high-cost resident: the mean WWST increased substantially, from 257 to 362.

Table 4-25. Extensive Service Criteria, ranked by resource use (see text for description) for residents in the RE and E categories.

Rank	Criterion	Number of Observations*	Therapy + Nursing WWST	Nursing WWST
1	Ventilator/respirator	147	417	391
2	Tracheostomy care	79	373	329
3	Suctioning	18	291	249
4	Parenteral/IV feeding	192	305	233
5	IV medication	558	253	176

* Total number of observation = 994, based on RUG-IV Version 1.

We then sought additional criteria for the E (and RE) categories. We adopted liberal criteria to identify a large set of criteria with high resource use that could increase overall WWST variance explanation. The search was performed using more than 500 potential criteria, as described in Section 4.5 (MDS 2.0 items, STRIVE Addendum items, MDS-based scales, and other algorithms).[59] We computed statistics for the ***additional*** residents that would be added to the RE and E category if we added the particular criterion.

[58] Analysis performed with RUG-IV Version 2, N=994 residents in RE or E category or with IV medications and ADL RUG-III Index above seven

[59] Analysis performed with residents in the derivation sample not already classified into the R + E or E or R categories, RUG-IV Version 2, with RUG-III Index above seven (N=3524).

Empirically, after examining the list generated, it was determined that criteria demonstrating one of the following four characteristics would be examined further:

- Mean therapy + Nursing WWST greater than 215;
- Increased variance explanation of WWST;
- Higher means for the E groups;
- Increased homogeneity within E groups.

We placed an additional requirement that the criteria must be present for greater than 10 residents, as to help avoid examining spurious results. Six potential additional criteria for the E category were identified: isolation for infection control, bi-level/continuous positive airway pressure (BIPAP/CPAP), stage 4 pressure ulcer, septicemia, internal bleeding, and dehydration (see Table 4-26).

Table 4-26. Potential Additional Extensive Services Criteria

Potential Criterion	Number of Observations	Mean Nursing WWST
Isolation for infection control	21	231
BIPAP/CPAP	11	230
Stage 4 Pressure Ulcer	100	216
Septicemia	15	224
Internal bleeding	19	219
Dehydration	11	224

We considered these six criteria, as well as suctioning, IV medications and IV/parenteral feeding, as potential candidates, and subjected them to a more rigorous examination as part of the determination of the Special and Clinically Complex categories (see following section).

During the examination process described in the next section, one criterion – isolation for infection control – performed most like the other two E category criteria already determined (specifically, ventilator/respirator and tracheostomy care) and differently than other criteria. We reviewed these three criteria to ascertain all three also had high Nursing WWST in the RE category. As a result, RUG-IV has three criteria for the E category: ventilator/respirator, tracheostomy care, and isolation for infection control.

4.8.3 Special/Clinically Complex Categories

For residents not qualifying for Extensive Services, the RUG-III system has two major hierarchy categories devoted to persons with major medical problems and receiving substantial services: Special Care (S) and Clinically Complex (C). Initial evaluation of STRIVE data showed there was the potential to substantially revise which criteria differentiated residents in these two categories. In order to derive RUG-IV, we

first considered which criteria should be used for both the Special Care and Clinically Complex categories together, then how they should be differentiated. As part of that search, we considered whether new criteria could be found that would identify individuals consuming these relatively high levels of care resources. As always, all choices were reviewed for clinical significance and relevance as well as clarity of clinical indication for services or diagnosis of conditions. Also, historically, several of the criteria for these categories have specific "qualifiers" that more strictly limit those who are assigned to the category. An example of this is the criterion of tube feeding which has as a qualifier that the resident also exceeds thresholds for the numbers of calories and ounces of fluid.

The RUG-III system has nine clinical criteria for the Special Care category, several with other qualifiers. RUG-III also includes residents who meet clinical criteria for the Extensive Services category but fail to meet the additional qualification of a RUG-III ADL Index value at or exceeding 7 (see Table 2-1 for list of criteria). The Clinically Complex ("Complex") category has 15 criteria (again, some with qualifiers) and includes all residents meeting the clinical criteria for Special Care category but not the qualifier of an ADL threshold of 7 or more.

We began by investigating whether each RUG-III Special Care or Clinically Complex criterion corresponded to sufficient resource use (WWST) to be part of an appropriate assignment. As with the Extensive Services category, we evaluated each individual criterion (and its qualifiers) by the following:

- Mean Nursing WWST for individuals who would be brought into the Special Care or Clinically Complex categories by that criterion as compared to mean Nursing WWST for individuals brought into these categories by all other criteria;
- Differences in overall RUG variance explanation with or without that criterion.[60]

For these categories some RUG-III criteria such as fever, dehydration, diabetes, internal bleeding, physician orders and changes, and transfusions, appeared associated with high resource use compared to their category. However, no criterion appeared sufficiently resource-intensive to be used as a criterion for the Extensive Services category, i.e., around 300 Nursing WWST. Conversely, while other criteria, such as radiation therapy, second- and third-degree burns, chemotherapy, and qualification for the Extensive Services category except for a RUG-III ADL Index threshold below 7, appeared associated with low resource use, none were so low as to be omitted, i.e., below 120 Nursing WWST. Therefore, all RUG-III Special Care and Clinically Complex category criteria remained under consideration for these categories in RUG-IV.

---

[60] Using RUG-IV v3, in the derivation sample (n-6464) mean Nursing WWST was calculated separately for each qualifier within the combined Special and Clinically Complex categories, using all individuals meeting that qualifier. The difference was calculated between the mean Nursing WWST for that qualifier and the mean Nursing WWST of the category without that qualifier. R Squared was calculated with and without each qualifier.

We then identified additional criteria, proceeding as follows:

- Among residents who were not already classified into the Special Care, Clinically Complex, or any higher category, we identified all criteria for which the mean Nursing WWST was greater than 120 WWST.[61] A list of over 500 criteria was screened, including RUG-III criteria formerly used or considered for the Extensive Services category (see Section 4.8.2), individual MDS 2.0 items, STRIVE addendum items, MDS-based research scales, and MDS-based Clinical Assessment Protocols (CAPs) (see Section 4.5.4). For each candidate criterion, we calculated the mean Nursing WWST and the change in R^2 if it were added to either the Special Care or Clinically Complex categories.
- We distributed the full list of potential criteria to a Clinical Panel and reviewed the list internally to develop other criteria.
- All newly-developed criteria were reviewed for the following:
  - Clinical significance , specifically importance, and incidence;
  - Clarity of clinical indication for services or diagnosis of conditions;
  - Effect on resource use
- Many criteria were considered with alternative definitions and clinical qualifiers. Following this review, 25 criteria were selected for possible use in RUG-IV (see Table 4-27).

Table 4-27. Potential RUG-IV Criteria for Special Care (S) or Clinically Complex (C) Categories

Variable	RUG-III Category	Definition Changed***	Number qualifying with that criterion alone	Mean Unadjusted Nursing WWST
Comatose (confirmed by ADLs of 4 or 8)	C	X	3	250.6
Isolation for Infection Control	-		12	229.7
Fever (with vomiting or weight loss or pneumonia or dehydration)	S	X	14	229.2
Quadriplegia*	S		39	213.1
Ulcers (2 or more Stage 2 ulcer or 1 Stage 3 or 4 pressure ulcer) and skin treatment	S	X	247	205.9

[61] Using RUG-IV, version 3 and limited to qualifiers met by at least 10 residents with a RUG-III ADL Index score of 7 or greater (the RUG-III criterion for classification into Special Care),

Variable	RUG-III Category	Definition Changed***	Number qualifying with that criterion alone	Mean Unadjusted Nursing WWST
Parenteral/IV Feeding	E		52	198.9
Respiratory therapy (7 days)	S		33	196.8
Tube feeding (with sufficient intake)	S		131	184.8
Surgical wound or open lesion with wound care or dressings or ointment	S		166	183.7
Multiple sclerosis*	S		65	183.3
Cerebral palsy*	S		37	168.6
Diabetes (with injections on 7 days and 2 or more physician order changes)	C		184	168.5
Transfusion	C	X	20	166.1
Shortness of breath with Emphysema/COPD	-		94	165.8
IV medication	E	X	145	165.1
Oxygen therapy	C	X	383	149.7
Hemiplegia*	C		350	149.5
Pneumonia	C		73	142.2
Burns (2^{nd} or 3^{rd} degree)	C		9	137.8
Parkinson's disease*	-		246**	139.7**
Foot infection or lesion with dressings	C		43	131.9
Dialysis	C		47	128.7
Radiation therapy	S		4	120.9
Septicemia	C		3	108.7
Chemotherapy	C		14	88.3

*Clinical condition with qualifier of ADL-IV Index >= 5
**Count and mean WWST without ADL-IV Index qualifier
***An "X" in the "Definition Changed" column indicates a criterion that was used in RUG-III but with a changed definition for RUG-IV.

This list includes 22 criteria used for RUG-III (those with an "S" for Special Care or a "C" for Clinically Complex). For six of these (with an "X" in the "Definition Changed" column), the RUG-III definitions were altered slightly to the following RUG-IV definitions:

- Fever and one of the following: vomiting, weight loss, pneumonia, tube feeding (dehydration qualifier omitted for RUG-IV)
- Stages 3 or 4 Pressure Ulcer or more than one Stage 2 ulcers of any type (dropped consideration of Stage 1 ulcers for RUG-IV)
- Comatose (dropped "time awake" qualifier for RUG-IV)

- Oxygen therapy (post-admission only for RUG-IV)
- IV medication (post-admission only for RUG-IV)
- Transfusion (post-admission only for RUG-IV).

In RUG-III, criteria associated with chronic conditions such as quadriplegia, cerebral palsy, multiple sclerosis, and hemiplegia, had a RUG-III Index ADL qualifier of 10 or more. We tested and sought a threshold for the ADL-IV Index that would work across these chronic conditions plus the newly-identified Parkinson Disease, to identify high resource usage. The value was determined by testing all ADL threshold values to identify which would maximize variance explanation of resource use, i.e. Nursing WWST, within each condition and within all. The best threshold for the ADL-IV Index was found to be 5; this corresponds to a RUG-III ADL Index value of 11 (see Section 4.7), not far from the value of 10 used in RUG-III.

Several criteria used in RUG-III to classify residents into the Special Care and Clinically Complex categories were eliminated for use in RUG-IV. Some were eliminated because they were deemed unreliable and difficult to diagnose, or could not be audited, while others provided inappropriate incentives. These included the following RUG-III criteria:

- Internal bleeding: too difficult to diagnose reliably
- Dehydration: too difficult to diagnose reliably
- Suctioning: no clear indications for use
- Feeding tube with sufficient nutrition/fluid intake and aphasia (used for RUG-III Special Care): not more resource intensive than the same criterion without aphasia (feeding tube with sufficient nutrition/fluid intake without consideration of aphasia was retained as a criterion)
- Physician visit/order changes: no verifiable definition found for a sufficient number of clearly resource-intensive residents (more details follow later in this section).

For other criteria, we often tested multiple versions, with differing qualifiers, to find a definition that successfully identified a reasonable number of high resource use residents with a consequential medical issue with good reliability. In a few cases, we were unable to find a successful definition and no version of criterion was used. The following are criteria considered but not used (with the number of qualifying residents in the derivation sample, and their mean Nursing WWST):

- Stage 4 pressure ulcer: combined with stage 3 pressure ulcer to reduce the incentive to upcode the stage of a severe pressure ulcer, and recognizing the complexity of differentiating between stage 3 and 4 ulcers (100; 216);
- Abdominal feeding: used with feeding tube (86; 192.7);
- Respiratory failure (ICD-9): used with oxygen therapy (11; 168.8);
- Sleep apnea: too few residents and poorly verifiable (15; 193.9);

- BIPAP/CPAP: potentially gameable (11; 230);
- Dyspnea: most residents identified by the oxygen therapy criterion (187; 178.2);
- No voluntary neck movement: poorly verifiable (24; 189.3);
- Deaf-blind (moderately or severely impaired in hearing plus highly or severely impaired in vision [62]): discarded after discussion with CMS (66; 182.5);
- Hip fracture: discarded after discussion with CMS (51; 160.0);
- Inability to lie flat due to shortness of breath: used with Emphysema/ COPD (108; 177.0);
- Surgical wound care: used with treatment (11; 175.4);
- Vomiting: poorly verifiable (24; 130.3);
- Missing limb: low resource use (36; 121.3);
- Diabetes and dialysis: used separately (41; 152.3);
- Diabetes and renal failure: diabetes used and dialysis used instead of renal failure (159; 158.5);
- Internal bleeding and transfusions: poorly verifiable and too few residents (1; 191.0);
- Inability to lie flat due to shortness of breath with respiratory therapy: inability to lie flat due to shortness of breath with Emphysema/COPD used instead (10; 180.0);
- Inability to lie flat due to shortness of breath with respiratory failure: inability to lie flat due to shortness of breath with Emphysema/COPD used instead (2; 147.4);
- Inability to lie flat due to shortness of breath with respiratory infection: inability to lie flat due to shortness of breath with Emphysema/COPD used instead (15; 165.1);
- Respiratory failure and Emphysema/COPD: inability to lie flat due to shortness of breath with Emphysema/COPD used instead (12; 181.1);
- Respiratory failure and Asthma: oxygen therapy with respiratory failure used instead (4; 128.4);
- Respiratory failure and respiratory therapy for 7 days (2; 217.6): respiratory therapy for 7 days used instead;

[62] The definition for deaf-blind provided by the CAN-STRIVE project and based on analysis and evaluation by a panel of deaf-blind individuals (John Hirdes, personal correspondence, 2008)

- Respiratory failure with inability to lie flat due to shortness of breath: inability to lie flat due to shortness of breath with Emphysema/COPD used instead (2; 147.4);
- Respiratory failure with respiratory infection: oxygen therapy with respiratory failure used instead (2; 94.5);
- Respiratory failure and one of Emphysema/COPD or respiratory therapy for 7 days: oxygen therapy with respiratory failure used instead (13; 175.5);
- Respiratory failure and Emphysema/COPD and respiratory therapy for 7 days: oxygen therapy with respiratory failure used instead (1; 327.7);
- Asthma with inability to lie flat due to shortness of breath: oxygen therapy used instead (22; 159.8);
- Foot infection and open foot lesions and foot treatment: foot infection or open foot lesions with foot treatment used instead (20; 206.9);
- Foot infection and open foot lesions and no foot treatment: foot infection or open foot lesions with foot treatment used instead (3; 136.7);
- Speech clarity and voice change and difficulty swallowing and caregiver observation of difficulty eating: poorly verifiable (22; 200.8);
- Speech clarity and voice change and difficulty swallowing and resident complaint of difficulty swallowing: poorly verifiable (22; 200.8);
- Speech clarity and voice change and difficulty swallowing and choking/coughing during intake: poorly verifiable (23; 197.7);
- Vomiting and dehydration: no residents with criterion.

While many of the criteria that classified residents in the RUG-III system into either the Special Care or Conically Complex Categories were validated in the STRIVE data as resource-intense, the magnitude of their resource use was considerable different in the RUG-III and STRIVE analyses. Thus, criteria that in RUG-III would, for example, classify a resident in the Special Care category might in STRIVE be substantially lower than other RUG-III Special Care criteria. As well, we had identified additional criteria to be considered. We therefore performed a comprehensive analysis of the relative resource use of all possible criteria. Our goals were to:

- Determine whether any criteria were sufficiently resource intensive for the Extensive Services category or insufficiently resource intensive for the Clinically Complex category;
- Evaluate how many categories were needed to divide the criteria into groups with similar resource use; and
- Specify the criteria for each category.

The approach taken had to be more sophisticated than simply ranking the potential criteria by their average resource use, i.e., Nursing WWST. Many residents

had more than one of the criteria. It followed that the evaluation of a criterion had to be performed on everyone not otherwise classified into the category by some other criteria. To determine relative resource use, we used a sequential hierarchical analysis to order criteria from most resource-intensive to least resource-intensive. In this analysis, criteria are repeatedly ranked by resource usage, from most to least, and any resident having the highest-ranked criterion is removed from the analysis. Since the most resource-intensive residents are removed, each successive criterion identified corresponds to less resource use than earlier criteria. Eventually, the most resource-intensive criterion identified is barely higher than the average resource use of the remaining residents and may well not be sufficiently resource-intensive for the Clinically Complex category. This approach has the desirable property of classifying each resident on the basis of his/her most resource-intensive condition. Because resource use may differ among criteria due to associated ADL limitations, we (re)adjusted resource use for ADL index score before ranking criteria.

The analysis included all residents in the derivation sample who a) did not qualify for a Rehabilitation category or qualify for Extensive Services via the two criteria already identified (ventilator/respirator or tracheostomy care), and b) who had an ADL-IV Index of 2 or more, the threshold we eventually found for these categories (see Section 4.8.5). This resulted in a sample for analysis consisting of 1755 observations.

The analytic procedure was:

1. To adjust for ADL differences between criteria, we created an ADL-adjusted resource use measure by dividing each resident's Nursing WWST by the mean Nursing WWST for all residents with that ADL-IV Index score. For interpretability, we multiplied this ratio by the mean ADL index score across all observations (the grand mean).
2. Then, we performed 24 iterations of these four steps:
    a. Estimated the mean ADL-adjusted resource usage for all as-yet unordered criteria;
    b. Ordered all remaining criteria by their mean ADL-adjusted resource use;
    c. Add the most resource-intensive criterion to the ordered list of criteria;
    d. Removed from the analysis all residents with that criterion.

The resulting ordered list of criteria is displayed in Table 4-28. In general, as one moves down the list, Mean ADL-Adjusted Nursing WWST (the measure we used in ranking) decreases. However, in some cases, a criterion can have greater Mean ADL-Adjusted Nursing WWST than other higher-ranked criteria. Such situations occur when two criteria are associated and residents who share both criteria are less resource intensive than residents with one or the other criteria, but not both. One such inversion in rank occurs for comatose and quadriplegia. So, on average, quadriplegia is more expensive than comatose, so it is ranked above comatose. But, among comatose residents, the less expensive are quadriplegic. When these comatose, quadriplegic residents are removed from the analysis, average comatose resource use rises - enough to appear more expensive in this list than quadriplegia.

Although isolation for infection control is less resource intensive than the ventilator/respirator (393 Nursing WWST) and tracheostomy care (332 Nursing WWST) criteria for the Extensive Category, it is clearly substantially more resource-intensive than other criteria evaluated here. We determined isolation for infection control to be a criterion for Extensive Services. The other criteria spanned a wide range of resource intensity, from 223.1 to 127.7 ADL-adjusted WWST, and none were considered too low for Clinically Complex.

Table 4-28. List of Criteria for Extensive (E), RUG-III Special Care (S), RUG-IV Special Care High (H), RUG-IV Special Care Low (L), or Clinically Complex (C) Categories, Ordered by Adjusted Resource Use, with RUG-IV Category Assignment.

Order	Variable	RUG-III Category	RUG-IV Category	N	Mean ADL-IV Index	Mean ADL-Adjusted Nursing WWST*	Mean Unadjusted Nursing WWST
1	Isolation for Infection Control	-	E	31	8.7	285.2	279.9
2	Fever*	S	H	25	12.3	223.1	241.5
3	Septicemia	C	H	19	9.0	216.3	210.6
4	Parenteral/IV Feeding	E	H	90	9.8	212.8	213.1
5	Diabetes	C	H	185	9.0	205.1	200.8
6	Respiratory therapy	S	H	27	9.0	207.7	198.8
7	Shortness of breath with Emphysema/ COPD	-	H	58	7.5	195.8	183.5
8	Quadriplegia**	S	H	34	13.0	192.0	212.9
9	Comatose	C	H	4	14.5	193.5	229.8
10	Ulcers*	S	L	156	10.7	186.4	190.1
11	Cerebral palsy**	S	L	31	11.1	184.0	186.3
12	Dialysis	C	L	31	7.2	175.3	165.8
13	Multiple sclerosis**	S	L	51	11.4	174.2	180.3
14	Tube feeding	S	L	168	13.1	164.3	180.1
15	Oxygen therapy***	C	L	231	8.6	160.0	154.8
16	Radiation therapy	S	L	2	8.5	155.1	164.2
17	Foot infection or lesion	C	L	26	8.3	152.0	145.2
18	Parkinson's disease**	-	L	144	10.5	148.8	152.3
19	Surgical wound or open lesion	S	C	82	7.3	149.4	139.7
20	Pneumonia	C	C	28	6.5	146.1	130.4
21	Hemiplegia**	C	C	252	10.0	145.2	145.1
22	IV medication	E	C	70	8.5	139.6	135.1
23	Transfusion	C	C	5	6.8	129.9	116.2

Order	Variable	RUG-III Category	RUG-IV Category	N	Mean ADL-IV Index	Mean ADL-Adjusted Nursing WWST*	Mean Unadjusted Nursing WWST
24	Chemotherapy	C	C	1	2.0	154.5	122.9
25	Burns*	C	C	4	4.8	127.7	109.8

*See description in text before Table 4-27
** With qualifier: ADL-IV Index <= 5
*** Note: item changed in final system published

To obtain relatively homogeneous resource intensity within categories, we decided to form three clinical categories for RUG-IV instead of the two categories (Special Care and Clinically Complex) in RUG-III. We denoted these three categories as Special Care High (H), Special Care Low (L), and Clinically Complex (C). Tentatively, Special Care High included the eight criteria between 225 and 190 ADL-adjusted Nursing WWST: fever (with pneumonia, or vomiting, or weight loss, or feeding tube), septicemia, parenteral/IV feeding, diabetes, respiratory therapy, shortness of breath with emphysema/COPD, quadriplegia, and comatose. Special Care Low included nine criteria between 190 ADL and approximately 150 ADL-Adjusted Nursing WWST: ulcers, cerebral palsy, dialysis, multiple sclerosis, tube feeding, oxygen therapy, radiation therapy, foot infection or lesion, and Parkinson's disease. Clinically complex included the remaining seven criteria, with ADL-Adjusted Nursing WWST from 150 down to approximately 125: surgical wound or open lesion, pneumonia, hemiplegia, IV medication, transfusion, chemotherapy, and burns. As we describe later in this section, a few of these criteria were reevaluated based on comments made in response to the publication of these criteria in the Notice of Proposed Rulemaking.[63]

In the derivation of RUG-III, there was a significant attempt to develop a usable criterion to identify the Clinically Complex category that would represent resident vulnerability or instability. No clinical criteria could be found, and RUG-III instead utilized measures of physician involvement (visits and order changes) as a surrogate. In particular, RUG-III classifies residents into the Clinically Complex category using a two-part criterion: 1) one or more physician visits (MDS 2.0 item P7) and four or more physician order changes (MDS 2.0 item P8); or 2) two or more visits and two or more order changes. It has been reported that physician order changes are easily manipulated when included in a payment system. Thus, in the current analysis we focused on the issue of vulnerability in two efforts: first, to test a criterion based on physician visits alone, and second to see if our Frailty Scale (see Section 4.5.3) would be effective.

[63] Federal Register / Vol. 74, No. 90 / Tuesday, May 12, 2009 / Proposed Rules. 42 CFR Part 483 [CMS–1410–P] RIN 0938–AP46; 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

Physician visits are associated with higher WWST for residents not classified into the Rehabilitation Plus Extensive, Rehabilitation, and Extensive categories[64] (e.g., the 1210 residents with 2+ visits in 14 days had a mean Nursing WWST of 199.6). However, many of these residents have other characteristics that would classify them into the H, L, or C categories. Thus, after determining all other Special or Clinically Complex criteria, we examined resource use related to physician visits for all residents not yet classified.[65] We consider three thresholds for days with physician visits: 2+, 3+ and 4+ of the last 14 days, and compared their characteristics with those of the RUG-III criteria that were not selected (see Table 4-29). A 2+ days with visits threshold identified many residents (242), but average resource use for these residents was too low for the Clinically Complex category (115.8 Nursing WWST) and using this criterion reduced model fit (i.e., variance explanation).[66] Raising the threshold to 3+ or 4+ days with visits did identify residents with high WWST, but the number of observations was few (59 and 19 residents, respectively) and thus had negligible effect on model fit. We concluded that a criterion based on physician visits alone was not appropriate.

Table 4-29. Physician Visits as a Criterion for the Clinically Complex Category

Criterion	Additional residents identified	Mean Nursing WWST	Variance Explanation (R^2)		
			Including Criterion	Without Criterion	Difference
RUG-III: Physician order changes on 4 or more days AND physician visits on 1 or more days OR Physician order changes on 2 or more days AND physician visits on 2 or more days.	206	123.28	39.16%	38.96%	-0.20%
Physician visits on 2 or more days	242	115.8	39.03%	38.96%	-0.07%
Physician visits on 3 or more days	59	123.5	38.96%	38.96%	0.00%
Physician visits on 4 or more days	19	137.9	38.96%	38.96%	0.00%

Next, we considered the Frailty Scale as a way to identify vulnerable and potentially unstable residents. As with physician visits, among those not classified into

[64] Based on RUG-IV, Version 3, 4718.observations in the derivation dataset in categories below Extensive.
[65] Based on RUG-IV, Version 7, 2900.observations in the derivation dataset in categories below Clinically Complex.
[66] Run as a regressions of WWST=a+b*RUG where WWST was nursing WWST and RUG was the categorical variable of RUG-IV groups, Version 7. In the first model, no changes were made to RUG-IV; in the second, a physician visit criterion was added for the Clinically Complex category

the Rehabilitation Plus Extensive, Rehabilitation, and Extensive categories,[67] frail residents had reasonably high Nursing WWST (e.g., the 3,383 residents scoring 2 or more on the MDS Frailty Scale had a mean Nursing WWST of 123.25). However, again as with the physician visit criterion, many of these residents had other comorbid characteristics that would already classify them into the Special Care High, Special Care Low, and Clinically Complex categories. We thus tested the MDS Frailty Scale in the residents not classified into any hierarchy category above or including Clinically Complex.[68] Our test included two Frailty Scale thresholds: 2^+ and 3^+. Model fit (variance explanation[69]) was minimally improved for the threshold at 3+ and mildly worsened for the threshold at 2+; both criteria had Nursing WWST comparable with criteria at the low end of the Clinically Complex category (see Table 4-30). If a frailty scale threshold of 3+ were used as a Clinically Complex qualifier, then the mean nursing WWST for Clinically Complex would decrease from 140.89 to 135.74 and 48% of Clinically Complex residents would qualify solely on the basis of frailty. If a frailty scale threshold of 2+ were used as a Clinically Complex qualifier, then the mean nursing WWST for Clinically Complex would decrease from 140.89 to 127.93 and 73% of Clinically Complex residents would qualify solely on the basis of frailty. Use of frailty as a Clinically Complex qualifier would redefine the Clinically Complex category to one dominated by frailty. This was seen as undesirable, especially given that inclusion of frailty did not improve model performance. Therefore, frailty was not used a Clinically Complex qualifier and was considered as a possible criterion for lower RUG categories.

Table 4-30. Frailty as a Criterion for the Special Care or Clinically Complex Categories

Criterion	Additional residents identified	Mean Nursing WWST	Variance Explanation (R^2)		
			Including Criterion	Without Criterion	Difference
MDS Frailty Scale 3 or more	402	130.08	38.99%	38.96%	0.03%
MDS Frailty Scale 2 or more	1225	123.25	38.85%	38.96%	-0.11%

Responding to comments made in response to the Notice of Proposed Rulemaking, several criteria were redefined or changed, primarily on clinical grounds, and some new ones were added. Each was analyzed separately to ascertain that the model improved, but the complete hierarchical categorizing analysis was not repeated for all criteria.

- The ulcer criterion (Stages 3 or 4 Pressure Ulcer and more than one Stage 2 ulcer (of any type)) was expanded to include stasis ulcers recorded on the MDS as Stage 3 or 4. This change added 27 observations and the criterion remained in Special Care Low.

[67] Based on RUG-IV, Version 3, 4718.observations in the derivation dataset in categories below Extensive.

[68] Based on RUG-IV, Version 7, 2900.observations in the derivation dataset in categories below Clinically Complex.

[69] Models were run with and without each criterion, as for physician visits; see prior footnote 66.

- Oxygen therapy was changed from a single criterion to two criteria. The first, oxygen therapy and respiratory failure, is a criterion for Special Care Low. The second, oxygen therapy (without respiratory failure), is a criterion for Clinically Complex. No additional observations were identified by the change.
- Fever was redefined to include tube feeding as a qualifier, but not to include dehydration. This criterion remained in Special Care High and the change moved four residents from Special Care Low to Special Care High.

Therefore, the final RUG-IV model has three categories devoted to persons with significant medical problems and receiving substantial services: Special Care High, Special Care Low, and Clinically Complex. Special Care High includes eight criteria (full specifications, including qualifiers, are given in Table 2-1): fever, septicemia, parenteral/IV feeding, diabetes, respiratory therapy, shortness of breath with emphysema/COPD, quadriplegia, and comatose. Special Care Low includes nine criteria: ulcers, cerebral palsy, dialysis, multiple sclerosis, tube feeding, oxygen therapy with respiratory failure, radiation therapy, foot infection or lesion, and Parkinson's disease. Clinically complex includes eight criteria: oxygen therapy (without respiratory failure), surgical wound or open lesion, pneumonia, hemiplegia, IV medication, transfusions, chemotherapy, and burns.

4.8.4 Behavioral Symptoms and Cognitive Performance

Moving down the hierarchical categories in RUG-III (for developing RUG-IV), the next two categories are Cognitive Impairment and Behavior Problems. It should be noted that the final category, Reduced Physical Function (P), is determined by all individuals not classified into any of the other higher categories.

The RUG-III Impaired Cognition category is indicated by a score of 2 or more on the Cognitive Performance Scale (CPS). For RUG-IV, we examined whether procedural memory should be incorporated into the CPS. Prior research on other projects has shown that an alternative formation of the CPS, incorporating the STRIVE Addendum item on procedural memory, better correspondence to the widely used Folstein Mini-mental Status Examination.[70] We tested whether there was an improvement to RUG-IV if this new CPS scoring was used as the criterion for the Impaired Cognition category. The test was performed by evaluating whether the variance explanation of RUG-IV would increase substantially with this single change. After fitting RUG-IV, the variance explanation of Nursing WWST was changed insignificantly by including procedural memory into the CPS (R^2=23.1% including

[70] The change in the CPS is that the algorithm to compute the impairment count and the severe impairment count are changed: each are one point greater if procedural memory is impaired – otherwise, the CPS algorithm is not altered

procedural memory, 23.0% not including it).[71] As procedural memory does not appear to improve explanation of resource use, we kept the RUG-III CPS criterion for RUG-IV.

Historically, very few residents are categorized in the RUG-III Behavior Problem (B) category. This is mirrored in the STRIVE data, where only 41 (0.7%) observations are classified into this category.[72] The project's Technical Advisory Panel suggested violent behavior or developmental disability might be important concepts that are missed in defining this category.

The STRIVE Addendum included three items on violence: violence ideation, violence threats, and violence acts. When we added these as criteria, the RUG-IV Behavior Problem category increased by 12 residents, to 53 (0.8%), but there was no improvement in variance explanation.[73] In a similar manner, we examined the use of both violence and Mental Retardation/Developmental Disability (MR/DD) items (MR/DD Setting, Down Syndrome, and other organic conditions) as criteria for Behavior Problem category. The inclusion of these items increased the size of the Behavior Problem category to 58 (1.9%) from 41 (derivation sample observations in or below the Impaired Cognition category with 2,820 observations), but still the overall variance explanation did not increase. Therefore, we found no evidence for RUG-IV to use different criteria than RUG-III for identifying a behavior problem category.

Since there were few residents in the category corresponding to the RUG-III Behavior Problem category and the resource use was similar to that in the category corresponding to the RUG-III Impaired Cognition category, we combined these two categories to the new RUG-IV Behavior Symptoms and Cognitive Performance category. This combination caused almost no change in variance explanation (R^2 for Nursing WWST decreased from 22.39% to 22.31%[74]).

---

[71] Run comparing regressions of WWST = a+b*RUG, where RUG was the categorical variable of all RUG groups, based on RUG IV Version 3, and where the Impaired Cognition category was determined either by the CPS including or excluding procedural memory. The models were run on derivation sample observations in or below the Impaired Cognition category (N=2,820).

[72] Based on RUG IV Version 3 and using the full derivation sample.

[73] Run as a regression of WWST = a+b*RUG, where RUG was the categorical variable of all RUG groups, based on RUG IV Version 3. Model run on derivation sample observations in or below the Impaired Cognition category (N=2,820).

[74] Same model as directly above.

4.8.5 ADL Thresholds for Categories

To create more homogenous RUG groups, RUG-III used ADL Index thresholds as criteria for the Rehabilitation Plus Extensive, Extensive, Special Care, Impaired Cognition, and Behavior Problem categories. ; Rehabilitation Plus Extensive, Extensive, and Special Care used a lower RUG-III ADL Index threshold of 7 and Impaired Cognition, and Behavior Problem used an upper threshold of 10. We tested this concept using the new ADL-IV Index (see Section 4.7).

First, we sought evidence whether any threshold would be appropriate for RUG-IV. Analysis indicated that lower thresholds were appropriate for the Rehabilitation Plus Extensive, Extensive, Special Care High, and Special Care Low categories. An upper threshold was found to be appropriate for the Behavior Symptoms and Cognitive Performance categories. It was determined clinically that an ADL threshold for the Rehabilitation category was not appropriate. As with RUG-III, no threshold was sought for the Clinically Complex category.

Second, we looked for a single lower threshold that would work across the Rehabilitation Plus Extensive, Extensive, Special Care High, and Special Care Low categories. A single lower ADL threshold for all of these categories will simplify the RUG-IV system.

Using the RUG-IV Version 4 applied to the derivation sample (N = 6,464), lower ADL thresholds were tested for the Rehabilitation Plus Extensive, Extensive, Special Care High, and Special Care Low categories. Upper thresholds were tested for the new Behavior Symptoms and Cognitive Performance category.

Table 4-31 displays the results of testing for an ADL threshold for Extensive Services criteria, used in both the Extensive Services category and the combined Rehabilitation Plus Extensive Services category. There were a total of 293 observations from the derivation dataset in these categories. For each value of the ADL-IV Index (range 0 – 16), the table displays both the number of observations and the mean Nursing WWST for observations a) at that ADL level, b) ***below*** that level (i.e., not qualifying for the category at this threshold level), and c) at or above that level (i.e., qualifying for the category at this threshold level). We also tested the variance explanation (R^2)[75] for each ADL threshold value. The choice of an optimal threshold was made using three criteria. First, we looked where there was a substantial change across the thresholds in the mean WWST. Second, we looked for higher variance explanations. Finally, we sought similar values across all of the categories Rehabilitation Plus Extensive Services, Extensive Services, Special Care High, and Special Care Low. An appropriate ADL threshold was found for each of these categories and the threshold of 2 or more on the new ADL-IV Index was chosen as the

[75] Run as a regression of WWST = a+b*Threshold, where Threshold was a dichotomous variable indicating whether the observation met the selected ADL threshold. Model run on derivation sample observations in the Rehabilitation Plus Extensive and Extensive categories.

best common value; no alternative for specific categories was substantially superior. Table 4-31 presents the analysis for the combined Rehabilitation Plus Extensive Services and Extensive Services categories. The same type of analysis was used for the Special Care High and Special Care Low categories.

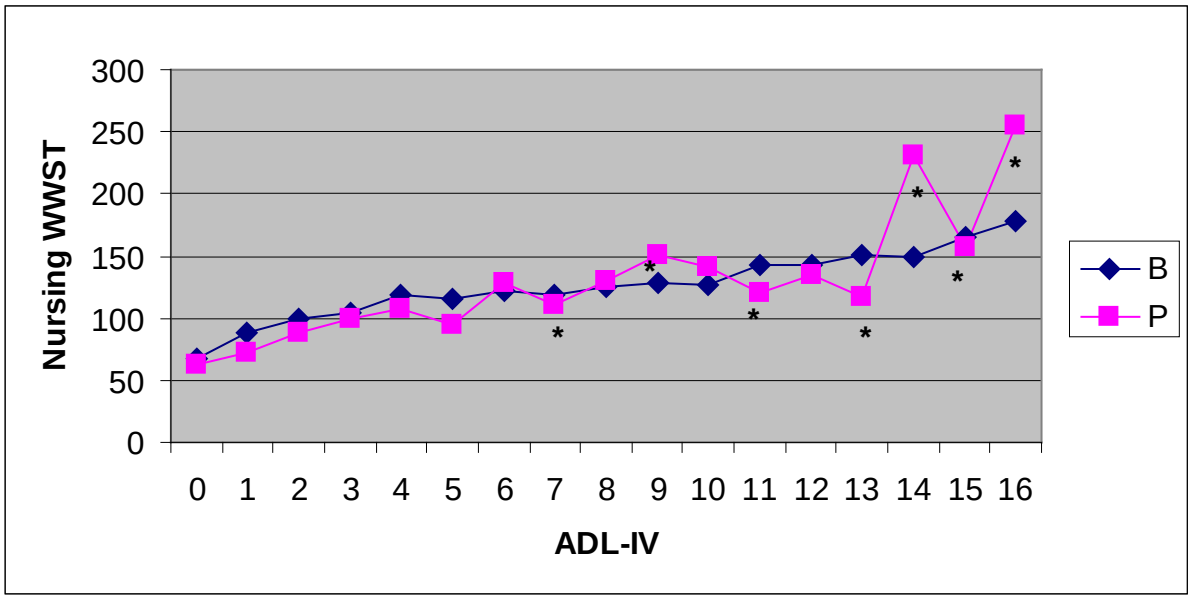
Table 4-31. Mean WWST for observations qualifying and not qualifying for Extensive Care (Combined Rehabilitation Plus Extensive and Extensive categories)* and variance explanation, by alternative ADL thresholds

ADL-IV Threshold	Number of observations at threshold value	Mean Nursing WWST for observations at threshold value	Number of observations below threshold	Mean WWST for observations below threshold	Number of observations at or above threshold	Mean WWST for observations at or above threshold	R2
TOTAL	293	345.1					
0	8	140.9			293	345.1	
1	4	120.0	8	140.9	285	350.9	5.6%
2**	10	229.3	12	133.9	281	354.2	9.2%
3	10	277.2	22	177.3	271	358.8	11.0%
4	8	256.0	32	208.5	261	361.9	11.0%
5	12	286.2	40	218.0	253	365.3	12.3%
6	10	332.8	52	233.7	241	369.2	12.9%
7	5	344.7	62	249.7	231	370.8	11.8%
8	6	394.1	67	256.8	226	371.4	11.1%
9	15	326.2	73	268.1	220	370.7	9.5%
10	18	291.0	88	278.0	205	374.0	9.3%
11	16	401.5	106	280.2	187	382.0	11.5%
12	14	288.7	122	296.1	171	380.2	8.3%
13	11	378.2	136	295.3	157	388.3	10.3%
14	31	376.5	147	301.5	146	389.1	
15	71	387.1	178	314.6	115	392.5	
16	44	401.1	249	335.3	44	401.1	

*Based on RUG-IV, Version 4.
**Chosen value

The final ADL threshold decision was for the new Behavior Symptoms and Cognitive Performance category. No clinical criteria were examined, as all such effects would have already been incorporated into the criteria for "higher" hierarchy categories. In RUG-III, all residents with a RUG-III ADL Index more than 10 were omitted from both the Impaired Cognition and Behavior Problem categories. We therefore sought an upper threshold for the new RUG-IV combined Behavior Symptoms and Cognitive Performance category. Using the RUG-IV version 9 on the derivation sample, we plotted the mean Nursing WWST for each level of the ADL-IV Index values separately for the Behavior Symptoms and Cognitive Performance category and for the Reduced Physical Function category observations (N=2900) (Figure 4-15). Values with few observations, and thus unstable estimates, are indicated. (Note that the RUG-III threshold of 10 corresponds to an ADL-IV Index value of 3.3.) For ADL Index values below 5, those in the Behavior Symptoms and Cognitive Performance category generally have increasing higher mean WWST values. Above the threshold of 5, this effect appears to go away, although the evidence is not strong. Based on this information, we determined that employing a threshold of 5, slightly higher than the equivalent of the RUG-III threshold, for this Behavior Symptoms and Cognitive Performance category would be appropriate.

Figure 4-15. Mean Nursing WWST by ADL-IV Index, for the RUG-IV Behavior Symptoms and Cognitive Performance (B) and the Reduced Physical Function (P) Categories**



* 25 or fewer observations
** Based on RUG-IV Version 9

These analyses completed the specification of all RUG-IV hierarchy categories.

4.9 Secondary splits of RUG-IV Categories

The major hierarchical categories of RUG-III were subdivided with "secondary splits," and at times with tertiary splits as well, to increase the similarity, or homogeneity, in resource use within the resulting groups. We discuss here the analysis to determine appropriate secondary splits below the Rehabilitation category (we discussed the issue of splitting the Rehabilitation Plus Extensive and Rehabilitation categories on the basis of therapy intensity earlier, in Section 4.8.1).

4.9.1 Secondary splits for the Extensive Services Category

The RUG-IV Extensive Services category has both very high average and very heterogeneous resource use. Thus, appropriate splits in this category are important for placing residents with similar resource use in the same group to achieve a good case-mix system.

In RUG-III, the Extensive Services (E) category is split into three RUG groups (SE1, SE2, and SE3), using a count of five criteria: the presence of two extensive services (parenteral/IV feeding, IV medication) and three RUG category qualifications (for Special Care, Clinically Complex, and Impaired Cognition categories).

When we applied the RUG-III logic for splitting the E category to the STRIVE data, however, the results were disappointing. The RUG-III count showed poor distinction among Extensive Services groups, including a slight inversion in the mean Nursing WWSTs: 265, 269, and 205 for ES3, ES2, and ES1 groups, respectively.[76] [Note: Extensive Services groups were renamed "ES" instead of the RUG-III nomenclature of "SE".] Preliminary analysis demonstrated that the inclusion of the count of category qualifiers, when added to Version 1 of the RUG-IV system, was a major problem. Therefore, after three extensive services were finalized as clinical criteria for the Extensive Services category (tracheostomy care, ventilator/respirator, isolation for infection control), we sought a superior method for splitting the E category.

Overall, eight alternatives for categorizing Extensive Services were considered, using different criteria and alternative numbers of 'cut points': (a) count of Extensive Services alone (in two variants: one with three services, one with two); (b) count of RUG category qualifications; (c) count of both Extensive Services and RUG category qualifications; (d) counts of all 28 individual criteria that qualify for the Extensive, Special Care High, Special Care Low, or Clinically Complex categories (in two variants: 3 or 10 groups); and (e) the ADL-IV Index (in two variants: 2 or 4 groups). The alternatives were evaluated by variance explanation of Nursing WWST (R^2)[77] and distinction among

[76] Mean nursing WWST for Extensive groups, after classifying using RUG-IV, version 1 on the derivation sample (N=405).

[77] Run as a regression of nursing WWST = a+b*RUG where RUG was a categorical variable, and involving only the observations in the Extensive Category at determined by RUG-IV, Version 4, with ADL-IV Index of 2 or more (N=219),

the Extensive subcategories in mean Nursing WWST. The results are shown in Table 4-32.

The largest variance explanation was achieved by splitting into 10 Extensive Services subgroups based on the count of the 28 individual criteria. However this resulted in an unusable number of very small groups and a model that was likely over-specified. The largest usable variance explanation was achieved by the third construct (c), an analogue to the RUG-III count with the same cut points, that included two extensive services and three RUG category qualifications (R^2 of 15.7%). The second-greatest variance explanation was achieved by the three-category count of all 28 individual criteria, with variance-optimized cut points of 1-2, 3-4, and 5-28 (R^2 = 15.4%).[78] However, both of these alternatives had unacceptably skewed distributions, with only 9 or 12 residents (respectively) in the ES1 category: in practice, they had only two useful categories. Also, the second of these options was considered to be over-fit, with a large number of possible cut points that could be considered. Models that have a large number of categories relative to the number of observations used to determine these categories usually cannot be replicated with other data. These models are called 'over-fit' or 'over-specified' and are not preferred compared to models with fewer categories, even though they appear to fit better with the current dataset.

The chosen alternative was the count of two extensive services (tracheostomy care, ventilator/respirator). It had an acceptably high variance explanation (R^2 = 13.3%) and no category with few residents. Thus, in RUG-IV, the Extensive Services category is split into three categories (ES1, ES2, and ES3), using a count of only two extensive services: ventilator/respirator and tracheostomy. The 31 residents qualifying for Extensive Services but with a count of "0," those with only isolation for infection control, form the lowest (ES1) group.

Table 4-32. R^2 for Extensive Services Count Alternatives

Alternatives for Extensive Services Count	Number of groups	R^2
3 Extensive Services (tracheostomy, ventilator/respiratory, and isolation for infection control)	3	11.5%
2 Extensive Services (tracheostomy, ventilator/respiratory)	3	13.3%
3 Category Qualifiers (H, L, and C)	3	3.2%
2 Extensive Services (tracheostomy, ventilator/respiratory) and 3 Category Qualifiers (H, L, and C)	5	15.7%
28 Individual Condition Qualifiers*	10	21.4%
28 Individual Condition Qualifiers*	3	15.4%
ADL-IV Index	4	10.6%
ADL-IV Index	2	10.1%

*All 28 individual criteria for the Extensive, Special Care High, Special Care Low, or Clinically Complex categories.

[78] Optimization was performed using Automatic Interactions Detection (AID) in the SAS Data Miner package, with nursing WWST as the dependent variable

4.9.2 Secondary Splits for RUG-IV Categories other than Extensive Services

In RUG-III, all categories except Extensive Services are subdivided ("split") by the RUG-III ADL Index. For RUG-IV, we considered many other alternatives for these parallel categories, but found that splitting them with the ADL-IV Index provided better statistical properties, including higher variance explanation, than the alternatives considered. First, we describe these comparisons, and then explain how we determined the cut points within the categories involved for the ADL-IV Index splits.

The Rehabilitation Plus Extensive category is initially split by the intensity of rehabilitation therapies. These splits are critical to describing resource differences across a wide range of therapy services. Within each of these subcategories, we preferred a two-group split, as in RUG-III, to maintain adequate group size in this sparsely populated category. We considered two alternative bases for splitting each rehabilitation intensity subcategory: the ADL-IV Index or the count of extensive services already determined for the Extensive category (see prior section). The best fit was obtained with the Extensive Count, with values 0-1 assigning residents to the lower subgroup and 2 assigning them to the upper subgroup within each rehabilitation intensity (R^2 = 20%).[79] However, many inversions were created by this split. The next best measure, the ADL-IV Index (R^2 = 16%), (see Table 4-33) was used. For parallelism, as with RUG-III, we decided that the R category would be best split by the ADL Index.

We then considered the appropriate measure for splitting the RUG-IV Special Care High, Special Care Low, and Clinically Complex categories other than the Extensive category. In RUG-III, the equivalent categories are subdivided ("split") by the RUG-III ADL Index. For RUG-IV, we also considered subdividing these categories by clinical complexity instead of the ADL-IV Index, an approach suggested by the project's Technical Expert Panel. We considered four alternatives for the Special Care High, Special Care Low, and Clinically Complex categories: (a) the best ADL split for that category; (b) a count of all conditions that would qualify a resident for that category; (c) a count of all conditions that would qualify a resident for any of those three categories lower on the hierarchy; and (d) a count of all conditions that would qualify a resident for any of those three categories.[80] The results are shown in Table 4-33. Of the four alternatives considered, using the ADL-IV Index to divide categories explained most variance for each of the Special Care High, Special Care Low, and Clinically Complex categories.

---

[79] Run as a regression of nursing WWST = a+b*RehCat where RehCat was a categorical variable describing the five levels of rehabilitation (i.e., Ultra High, Very High, etc.) subdivided by the selected measure (e.g., the ADL-IV Index or levels of the Extensive Count). The analysis was performed using RUG-IV, Version 4 and applied only to residents in the Rehabilitation Plus Extensive category (N=62).

[80] Run for each category as a regression of nursing WWST = a+b*VAR, where VAR was a categorical variable that represented one of the splitting alternatives displayed in Table 4-33. The analysis was applied only to residents in that specific category based on the category specification in RUG-IV Version 4.

Table 4-33. Variance Explanation (R^2) of Alternative Measures for Splitting the Categories other than Extensive Services (Number of Observations included in the last row)

Splitting Alternatives for Category *	Rehabilitation Plus Extensive Subcategories	Special Care High	Special Care Low	Clinically Complex
ADL-IV Index in categories (number of categories)	16% (2)	5% (2)	7% (3)	11% (4)
Condition Qualifiers for that Category	NA	3%	3%	3%
Category Qualifiers (Special Care High, Special Care Low, Clinically Complex) for that category and all below	NA	2%	1%	NA
28 Condition Qualifiers in 3 categories	NA	4%	4%	NA
Extensive Count (number of categories)	20% (2)	NA	NA	NA
Number of observations (Derivation sample)	62	384	987	654

*Based on RUG-IV, Version 4.

For the RUG-IV Behavior Symptoms and Cognitive Performance and Reduced Physical Function categories, we considered all variables in the database as possible candidates for secondary splits in these categories[81], looking for a variable that both demonstrated significant explanation of Nursing WWST and was associated in the correct "direction," i.e., with higher values for residents with a problem or troubling condition. While a few options emerged, all were inferior to using the ADL-IV Index as a secondary splitting measure; we discuss the other measures in greater detail in a following section about tertiary splits.

RUG-III used a variety of ranges of the ADL index to define subgroups, with from two to five splits for each category (see Figure 4-16). As the split points for each category were chosen to maximize WWST variance explained, those for one category often did not coincide with those for other categories. This variety of inconsistent splits was considered a positive attribute, as it made it more difficult to discern opportunities where up-coding an ADL would move a resident into a higher-paying group. However, with the substantial automation now seen in nursing facilities, this possible advantage is no longer at issue and superseded by the clarity of having a coordinated set of ranges. Experience with the use of RUG-III classification in payment systems has indicated that inconsistent ADL splits allow inversions, which occur when the WWST for a RUG group is less than the WWST for the RUG group with the same ADL value in a lower hierarchical category (see Section 6.1). Since payment rates are based upon WWST, an inversion can mean that a resident will receive a lower payment if they qualify for a higher group rather than a lower group. In Figure 4-16, note that the WWST for RHX is based on an ADL range of 13 to 18 but the WWST for RMX is based on a range of 15 to 18. Residents in RMX have higher average ADL dependence and more resource need

[81] Run as a regression of nursing WWST = a+b*RUG+c*VAR, where RUG was a categorical variable differentiating the Behavior Symptoms and Cognitive Performance and Reduced Physical Functions categories for RUG-IV Version 4 and VAR was one of a series of tested variables. The model was run on derivation sample observations in these two categories.

than those in RHX and, for that reason, the WWST for RMX may be higher than RHX. A resident in RHX with an ADL score of 15 may receive a lower payment than a resident in RMX with an ADL of 15. Such inversions are undesirable and require more complex classification systems that do not stop with the highest group.

Figure 4-16. RUG-III ADL Ranges for Secondary Spits

RUG-III Category	Rehabilitation Subcategory	ADL												
		4-5	6	7	8	9	10	11	12	13	14	15	16	17-18
Rehabilitation Plus Extensive	Ultra High			RUL									RUX	
	Very High			RVL									RVX	
	High			RHL						RHX				
	Medium			RML								RMX		
	Low			RLX										
Rehabilitation	Ultra High	RUA				RUB							RUC	
	Very High	RVA				RVB							RVC	
	High	RHA			RHB					RHC				
	Medium	RMA			RMB							RMC		
	Low	RLA									RLB			
Extensive Services		NOT ADL												
Special Care				SSA								SSB		SSC
Complex Conditions		CA							CB					CC
Impaired Cognition		IA	IB											
Behavior Problems		BA	BB											
Reduced Physical Function		PA	PB		PC		PD						PE	

For RUG-IV, our overall plan was to develop a standard, consistent set of ranges for the new ADL-IV Index: all ADL secondary splits would be based on these ranges, or combinations thereof. [Note: We have discussed earlier (in Section 4.8.1) that RUG-IV retains the subcategory splits of the Rehabilitation Plus Extensive and Rehabilitation categories based on rehabilitation intensity. We discuss here further ADL subdivisions of these ten rehabilitation intensity subgroups (five each in Rehabilitation Plus Extensive and Rehabilitation).] Figure 4-17 presents the final ADL ranges derived for RUG-IV.

Figure 4-17. RUG-IV ADL-IV Ranges for Secondary Spits

RUG IV Category	Rehabilitation Subcategory	ADL-IV				
		0-1	2-5	6-10	11-14	15-16
Rehabilitation Plus Extensive	Ultra High		RUL		RUX	
	Very High		RVL		RVX	
	High		RHL		RHX	
	Medium		RML		RMX	
	Low		RLX			
Rehabilitation	Ultra High	RUA		RUB	RUC	

	Very High	RVA		RVB	RVC	
	High	RHA		RHB	RHC	
	Medium	RMA		RMB	RMC	
	Low	RLA			RLB	
Extensive		ES1, ES2, ES3 (based on Extensive Count)				
Special Care High			HB	HC	HD	HE
Special Care Low			LB	LC	LD	LE
Clinically Complex		CA	CB	CC	CD	CE
Behavioral Symptoms and Cognitive Performance		BA	BB			
Reduced Physical		PA	PB	PC	PD	PE

The ADL-IV ranges for RUG-IV were determined by evaluating the effects of alternatives on Nursing WWST variance explanation and on the number of inversions created. Variance explained was obtained for many possible divisions of the 17-point ADL-IV Index (0-16, inclusive). Although statistical fit (e.g., variance explanation) can be expected to increase as the number of divisions increased, a case-mix system with too many groups is both unwieldy and likely to have small groups with unstable estimates of group mean resource use. We examined only divisions having at least two consecutive values in a subdivision (eight splits). After ascertaining that variance-maximizing divisions occurred at roughly the same ADL values regardless of the number of splits, we concentrated on five or fewer divisions to maintain a manageable number of RUG-IV groups. Thereafter, variations with five divisions were intensively investigated across all main RUG categories in several subsamples: all observations in the derivation sample, all observations below Extensive Services, Rehabilitation alone, Rehabilitation and Rehabilitation Plus Extensive Services, and all except Extensive Services. [82]

The final alternative chosen has the ADL-IV Index divided into the following ranges: 0-1, 2-5, 6-10, 11-14, and 15-16. This alternative had the highest variance explanation in two of the three subsamples, below Extensive Services (R^2 = 27.1%) and all except Extensive Services (R^2 = 29.7%), and tied for highest in the whole sample (R^2 = 37.3%). Cut points for this alternative usually coincided with cut points in the highest-variance-explanation alternative with 6, 7, or 8 subdivisions.

While we limited ourselves to these five ADL ranges across the RUG-IV system, there were two reasons that some categories had fewer than five subcategories. First, the Rehabilitation and Rehabilitation Plus Extensive categories had few residents in some ADL ranges; combining the five subcategories into fewer subcategories provided a larger sample size and a more stable estimate for the group mean WWST, and the Case Mix Index based on mean WWST. Second, some RUG categories include ADL restrictions (see Section 4.8.5), which forces some RUG categories to have empty ADL splits above or below these values.

[82] Run as a regression of nursing WWST = a+b*RA where RA was a categorical variable representing RUG categories subdivided by the candidate ADL splits. This model was run on multiple samples, as specified, based on RUG-IV, version 4.

The final set of RUG-IV subcategories is shown in Figure 4-17. For both the Rehabilitation Plus Extensive (RE) and Rehabilitation (R) categories, the highest subcategory had ADL-IV Index values of 11 or more. This alternative had the highest two-division variance explanation for all residents in the RE and R subsample (R^2 = 16.2). For the Rehabilitation Plus Extensive category, all ADL-IV Index values below 10 and greater than or equal to the lower threshold of 2 formed single groups. Due to the small number of residents qualifying for the Low Rehabilitation Plus Extensive subcategory, it was appropriate to create only one ADL group appropriate for the entire range from 2 to 16. For the Rehabilitation category, the chosen three-group alternative (ADL-IV Index divided as 0-5, 6-10, and 11-16) had the highest 3-division variance explanation for Rehabilitation category residents (R^2 = 18.1%) Again, fewer groups (two) were formed in the Low Rehabilitation subcategory due to small sample size.

All of the remaining categories, from Special Care High through Reduced Physical Function, were split into all five ADL ranges unless they had thresholds that prohibited this. Thus, the Special Care High and Special Care Low categories have four divisions, as the ADL threshold prevents inclusion of any residents with an ADL Index of 0 or 1, and the Behavior Symptoms and Cognitive Performance category has two divisions (here the ADL threshold is 5 or fewer). All five divisions are used for the Clinically Complex and Reduced Physical Function categories. This entire system of divisions, before tertiary splits, resulted in only a single mean WWST inversion below Extensive Services. This inversion was corrected as part of the process of developing the case-mix indexes (see Section 6-1).

4.10 Testing the Use of a Frailty Measure

One variable that demonstrated potential utility in explaining resource use in the Behavior Symptoms and Cognitive Performance and Reduced Physical Function categories was our derived Frailty Scale (see Section 4.5.3). Residents in these categories[83] have slightly fewer difficulties related to frailty than all nursing facility residents (see Table 4-34). Nevertheless, as in other RUG-IV groups, most (65%) of these residents had difficulty in one or two areas and few (2%) had difficulty in four or more frailty areas.

Table 4-34. Distribution of Frailty Scale in Behavior Symptoms and Cognitive Performance (B) and Reduced Physical Function (P) Categories and All Residents

Frailty Scale	Prevalence in B and P Categories				All*	
	Frequency	Percent	Cumulative Frequency	Cumulative Percent	Percent	Cumulative Percent
0	578	20.5%	578	20.5%	13.4%	13.4%
1	1022	36.2	1600	56.7	34.2	47.6
2	827	29.3	2427	86.1	33.1	80.7

[83] RUG-IV Version 3, derivation sample, n = 2820 below Clinically Complex.

3	335	11.9	2762	97.9	15.9	96.5
4	50	1.8	2812	99.7	3.2	99.7
5	8	0.3	2820	100.0	0.3	100.0

* All residents in derivation sample. Duplicated from Table 4-17

We explored several alternative configurations for identifying and locating a category of frail residents within RUG-IV. The first was to place a new Frailty category above the Behavior Symptoms and Cognitive Performance and Reduced Physical Function categories. It could be identified by a threshold of either 2 or 3 on the Frailty Scale, and would be split by the standard ADL-IV ranges described in Section 4-9-2 (see Figure 4-17). In our testing, other than the ordering of categories, the Behavior Symptoms and Cognitive Performance category was identified by the same criteria as before (see Section 4.8.4) and the Reduced Physical Function category was composed of the remaining residents not qualifying for any other category. For each configuration, we computed model fit.[84]

The results show a small improvement by adding a Frailty (F) category above the Behavior Symptoms and Cognitive Performance category (e.g., for Nursing WWST the variance explanation increased from 22.8% to 23.0%) (see Table 4-35), but at the expense of an additional 5 subcategories (and potentially an additional 10 or more RUG-IV groups, depending on use of tertiary splits). Placing Frailty between the B and P categories was much worse, with no increased variance explanation and inversions throughout (results not shown). With the decision to retain tertiary splits based on restorative nursing, splitting of the Behavior Symptoms and Cognitive Performance and Reduced Physical Function categories on both frailty and restorative nursing would create an unacceptably large and unmanageable increase in the number of RUG-IV groups. Thus, we decided not to include frailty as a criterion to either define a new category or splits for the existing categories in RUG-IV, but recommend that future study of this issue might be appropriate.

Table 4-35. Mean Nursing WWST for RUG-IV With a Frailty Category, Defined by Two Thresholds, and Placed Above the Behavior Symptoms and Cognitive Performance Category

RUG-IV Subcategory*	No Frail Category		With Frailty Category			
			Frailty Scale Threshold = 2		Frailty Scale Threshold = 3	
	N	Nursing WWST	N	Nursing WWST**	N	Nursing WWST**
FE			342	158.6	124	154.0
FD			517	127.3	183	130.6
FC			120	103.3	35	110.3
FB			113	91.8	30	90.1

[84] RUG-IV, Version 3, derivation sample. Models were run with and without each criterion, as for physician visits; see prior footnote for physician visits.

FA			128	68.8	21	72.6
BB	387	97.7	254	97.3	344	97.3
BA	303	65.1	251	62.3	296	65.5
PE	505	156.2	163	151.1	381	156.9
PD	875	123.8	358	119.1	692	122.0
PC	111	97.8	58	96.0	98	96.8
PB	139	81.1	92	75.3	130	80.1
PA	500	60.5	424	60.1	486	59.7
R Square (%)						
Nursing plus Therapy	22.2%		22.5%		22.4%	
Nursing	22.8%		23.1%		23.0%	

* Subcategory code: first character identifies category (F for Frail; B for combined B and I, and P as before)
** Inversions highlighted

4.11 Tertiary Splits

With the large sample sizes available in STRIVE, it was often possible to subdivide further (i.e., make "tertiary splits" to) the major groups shown in Figure 4-17, thereby increasing the homogeneity of groups in resource use.

The Rehabilitation Plus Extensive, Rehabilitation, and Extensive categories were sufficiently small that additional splits were not seen as appropriate. In the derivation sample, several of the groups in these categories had 30 or fewer residents (all nine Rehabilitation Plus Extensive groups, four rehabilitation groups, and one Extensive Services group). Splitting such small groups would not yield reliable results and and the decision was made not to further split the groups in these categories. We thus focused on further splitting in the other five categories: Special Care High, Special Care Low, Clinically Complex, Behavior Symptoms and Cognitive Performance, and Reduced Physical Function .

RUG-III uses depression as a tertiary split in the Clinically Complex category. We described in Section 4.5.2 the selection of a depression indicator that is predictive overall in the STRIVE sample. Within the RUG-IV system, a similar finding was achieved: residents in the Clinically Complex category with depressive symptoms having higher Nursing WWST than those without these symptoms. Depression also was seen to be an effective tertiary split for the Special Care High and Special Care Low categories[85] ($p<.01$). On this basis, the decision was made to develop pairs of final RUG-IV groups for each of the 13 subcategories for the Special Care High, Special Care Low, and Clinically Complex categories (see Figure 5-1 and Table 5-1). For example, the lowest resource Special Care High category (HB, with ADL Index values

[85] Run as separate regressions, one each for the observations in the Special Care High, Special Care Low, or Clinically Complex categories (N=669, 959, and 1432 respectively) as identified by RUG-IV Version 9. The model was: nursing WWST = a+b*CAT+c*DEP, where CAT was a categorical variable describing secondary ADL splits and DEP was a dichotomous variable determined by mood scale values of 3 or more.

of 2-5) is split into HB2 for those with depressive symptoms, and HB1 for those without such symptoms. As we describe later, in Section 6.2, small sample sizes lead to instability in estimating the mean Nursing WWST for many of the groups, so assigning group weights involved additional calculations.

Finally, we address tertiary splits in the Behavior Symptoms and Cognitive Performance and Reduced Physical Function categories. In RUG-III, these categories are split using restorative nursing (originally called nursing rehabilitation). We sought here evidence to include either restorative nursing or an alternative to split these RUG-IV categories.

We began by examining whether the original RUG-III scale is the best possible. Nine restorative nursing criteria used in RUG-III are created from 12 MDS items. For residents to qualify, they must receive 2 or more of these activities for 15 or more minutes on 6 or more days in the past week. These criteria are as follows (with MDS 2.0 item identifiers):

- Passive range of motion (p3a) or active range of motion (p3b)
- Bed mobility (p3d) or walking (p3f) training
- Splint or brace assistance (p3c)
- Transfer training (p3e)
- Dressing or grooming training (p3g)
- Eating or swallowing training (p3h)
- Amputation/prosthesis care (p3i)
- Communication training (p3j)
- Scheduled toileting plan (h3a) or bladder retraining program (h3b).

We computed the Cronbach's alpha statistic for these activities to determine how well they measure a consistent restorative nursing dimension (see Section 4.5.3 for a brief discussion of the purpose of this statistic). Only moderate item consistency was found (alpha = 0.6) for all items. Deleting passive range of motion, splint or brace assistance, and toilet/bladder plan improved the item consistency (alpha = 0.7) but the results still remained at best moderate. Still, there was clinical rationale to retain all these dimensions in a measure to encourage restorative nursing.

Restorative nursing was initially tested for the RUG-III model. To test the effectiveness of using restorative nursing as a tertiary splitting variable, we ran separate analyses in the combined RUG-III Impaired Cognition and Behavior Problems categories (N=672) and the RUG-III Reduced Physical Function (N=2,044) category[86],

[86] Run as a regression of WWST = a+b*RA+c*RN where WWST was nursing plus therapy WWST, RA was a categorical variable representing the RUG-III combined Impaired Cognition and Behavior Problems categories and the Reduced Physical Functions category split by the ADL ranges as shown in Figure 4-16 and as defined by the RUG-III model (using STM times rather than MDS time to identify therapy

using the derivation sample residents and the RUG-III restorative nursing qualifications of two or more restorative nursing activities for six or more days. The results were inconclusive. There was virtually a zero increase in variance explanation (VE) when restorative nursing was included in models run on the combined Impaired Cognition and Behavior Problems categories (VE = 0.1%) and on the Reduced Physical Function (VE=0.0%) category (see Table 4-36). As well, the difference in the average Nursing WWST for residents receiving restorative nursing within each category was not statistically significant and quite small (see Table 4-36). While the average Nursing WWST for the combined RUG-III Impaired Cognition and Behavior Problems categories was 85, the difference between those receiving and not receiving restorative nursing was only 6.6 (difference not significant, p=0.46). For the Reduced Physical Function category (where the average Nursing WWST was 115), the difference was 1.4 (again, a difference that was not significant, at the p=0.74 level). When the differences between those with and without restorative nursing were examined within the ADL subcategories of the combined RUG-III Impaired Cognition and Behavior Problems categories and the Reduced Physical Function category (i.e., the categories after splitting by the ADL-IV Index), the results were no better. For example, in splitting the five subcategories of Reduced Physical Function, three created restorative nursing groups that had *lower* Nursing WWST (inversions shown as shaded cells in Table 4-37).

Table 4-36. Restorative Nursing As a Tertiary Split for RUG-III

Categories	Number of Observations	Increase* in R^2	Nursing plus Therapy WWST		
			Average for Category	Difference Between Restorative Nursing and All Others	Test of Significant Difference (p)*
Combined Impaired Cognition & Behavior Problems	672	0.1%	85.1	6.6	0.46
Reduced Physical Function	2,044	0.0%	114.5	1.4	0.74

* See text for description

groups), and RN is restorative nursing as defined in the text. Separate models were run for the combined RUG-III Impaired Cognition and Behavior Problems categories (N=672) and the Reduced Physical Functions (N=2,044) category, using data from the derivation sample. The difference due to restorative nursing was the coefficient of the RN term, and it was tested for significance. The difference in variance explanation was the difference in the R^2 statistic for two models, one including the RN term and one excluding it.

Table 4-37. Restorative Nursing Splits for the RUG-III Reduced Physical Function Category Groups

RUG-III Reduced Physical Function ADL Subcategory	Restorative Nursing	N	Average Nursing WWST*
Physical E	Yes	82	157.7
	No	396	159.0
Physical D	Yes	153	129.6
	No	691	125.5
Physical C	Yes	17	105.4
	No	88	100.5
Physical B	Yes	14	73.1
	No	117	82.6
Physical A	Yes	24	60.7
	No	462	62.2

* Categories where inversions occurred are shaded.

These analyses were rerun after the major changes in RUG-IV were determined, as described above. Overall, similar results were found for RUG-IV, i.e., that restorative nursing added little to the variance explanation and created some inversions in the RUG-IV Reduced Physical Function category (see Table 4-38). Similar RUG-IV results were also found when we restricted our sample only to facilities that reported some restorative nursing.

Table 4-38. Restorative Nursing Splits for the RUG-IV Reduced Physical Function (P) Category Groups*

RUG-IV P ADL Subcategory	Restorative Nursing	N	Average Nursing WWST*
Physical E	Yes	24	177.7
	No	153	173.9
Physical D	Yes	59	153.1
	No	300	147.6
Physical C	Yes	105	124.6
	No	503	131.9
Physical B	Yes	46	109.6
	No	263	102.1
Physical A	Yes	32	62.4
	No	560	66.9

*Performed on the Derivation sample (N=6454)
** Categories where inversions occurred are shaded.

While restorative nursing was one possible (legacy) splitting variable to form tertiary splits in the lowest categories; we considered all other variables in the database as possible alternative candidates, looking for a variable that both demonstrated significant explanation of Nursing WWST[87] and was associated in the "correct direction," i.e., with higher values for residents with a problem or troubling condition or residents receiving a service. As no other splitting variables were found, it was decided to continue to support restorative nursing which has been a part of the RUG system and provides incentives for the provision of these services in RUG-IV.

It follows that in RUG-IV each of the seven Behavior Symptoms and Cognitive Performance and Reduced Physical Function ADL subcategories is split into two, one for those qualifying as receiving restorative nursing, the other not. For example, the lowest ADL Reduced Physical Function subgroups would be PA1 (without restorative nursing) and PA2 (with restorative nursing

The result of all of these specifications is a RUG-IV system with 66 groups, as shown in Figure 5-1 and Table 5.1.

4.12 Special populations

One of the goals of STRIVE was to determine if emerging "Special Populations" of nursing facility residents would be appropriately classified by the RUG system. The following populations were considered: residents with severe mental illness, younger age, AIDS/HIV, bariatric, and traumatic brain injury. In all cases, there were no specific criteria in RUG-III to identify these populations. We investigated whether these subpopulations themselves or characteristics of these subpopulations would be more useful case-mix predicators.

While most analyses in STRIVE were performed only on the derivation sample, the analyses were performed on the full sample (derivation plus validation) in cases where the special populations were relatively rare. It should also be noted that the analysis to identify these population was performed early in STRIVE, so that the variables to identify these special populations were always available and tested at all stages of the RUG-IV development discussed earlier in Section 4.

[87] Run as a regression of WWST = a+b*RUG+c*VAR, where WWST was nursing plus therapy time, RUG was a categorical variable identifying the seven subgroups defined by RUG-III, but with the Impaired and Behavior groups combined (i.e., with subgroups BA, BB, PA, PB, PC, PD, and PE) and VAR was the tested variable. The model was run on derivation sample observations. The results from a retest using RUG-IV V8 showed no major differences.

4.12.1 Severe Mental Illness

Two criteria were examined to identify nursing facility residents with severe mental illness (SMI). The first criterion identified residents with any of 10 diagnostic categories, based primarily on the ICD-9-CM diagnoses included in the MDS 2.0 :(drug or alcohol psychoses/dependence; schizophrenia; affective psychoses; paranoid states; other non-organic psychoses; personality disorders; other organic psychotic conditions; psychoses with an origin in childhood; anorexia or sleep disorder or eating disorder; or, organic brain damage. By this criterion, 1,406 (22.0%) out of 6,464 residents had severe mental illness.[88] The second SMI criterion used the same 10 diagnostic categories but excluded residents with severe dementia (identified by levels 5 and 6 of the CPS). By this criterion, 1,127 (17.6%) out of 6,464 residents had SMI. Both options are displayed in Table 4-39.

Table 4-39. Number and Mean Nursing WWST of residents with and without Severe Mental Illness (SMI), by individual diagnostic categories and overall, with and without residents with severe dementia

SMI Item	Number SMI (percentage of sample)	Mean Nursing WWST		
		SMI	No SMI	Difference
Including residents with severe dementia				
Drug or Alcohol psychoses/dependence	139 (2.2%)	110.5	149.1	-38.6
Schizophrenia, hallucinations, delusions	461 (7.1%)	114.4	150.9	-36.5
Affective psychoses (bipolar)	241 (3.7%)	123.1	149.2	-26.1
Paranoid states	60 (0.9%)	126.9	148.5	-21.6
Other nonorganic psychoses	292 (4.5%)	139.4	148.7	-9.3
Personality disorders	19 (0.3%)	111.2	148.4	-37.2
Other organic psychotic conditions	426 (6.7%)	133.8	149.3	-15.5
Psychoses origin childhood	6 (0.1%)	214.3	148.2	66.1
Anorexia, sleep or eating disorder	16 (0.3%)	122.7	148.3	-25.6
Organic brain damage.	52 (0.8%)	131.7	148.4	-16.7
Any of 10 categories	1406 (22.0%)	128.6	153.7	-25.1
Excluding residents with severe dementia				
Drug or Alcohol psychoses/dependence	54 (0.8%)	105.2	148.6	-43.4
Schizophrenia, hallucinations, delusions	45 (0.7%)	105.8	148.6	-42.8
Affective psychoses (bipolar)	52 (0.8%)	104.3	148.6	-44.3
Paranoid states	4 (0.1%)	120.5	148.3	-27.8
Other nonorganic psychoses	22 (0.3%)	112.0	148.4	-36.4
Personality disorders	1 (0.02%)	56.6	148.3	-91.7
Other organic psychotic conditions	14 (0.2%)	119.7	148.3	-28.6
Psychoses origin childhood	1 (0.02%)	101.8	148.3	-46.5
Anorexia, sleep or eating disorder	5 (0.1%)	141.3	148.3	-7.0
Organic brain damage.	1 (0.02%)	43.6	148.3	-104.7

[88] This analysis included all observations in the derivation sample before exclusion of the 10 observations with missing RUG-III information.

Any of 10 categories	1,127 (17.6%)	119.7	154.3	-34.6

By either criterion, residents with severe mental illness had ***lower*** resource use than residents without severe mental illness. Using the first SMI criterion, the mean Nursing WWST for residents with SMI was 128.6, substantially lower than all others, with a WWST of 153.7; the comparable WWSTs for the second SMI criterion were 119.7 for those meeting the criterion, compared with 154.3 for all others. The same relationships – i.e., lower resource use for those with SMI) were also seen in each of the individual ten ICD-9-CM categories for the two SMI definitions with only one exception (psychoses with origin in childhood not excluding severe dementia, and based on only 6 observations).

However, these initial results do not adjust for the characteristics known to be associated with differences in resource use, i.e., for RUG-III group. We considered how much more resource intensive those residents with SMI were compared with all others in that RUG group. We computed the residual WWST for every observation after adjusting for the fitted mean WWST for each group,[89] a "residual analysis."

By the first definition, the mean residual Nursing WWST, after fitting RUG-III, was 4.5 WWST units lower for those with severe mental illness compared to all others; by the second SMI definition, WWST was 6.3 less for those with SMI compared to all others. There was no systematic source of difference in the residuals by specific categories or coherent clusters of RUG groups. Furthermore, in the SMI population identified by either definition, RUG groups were significantly predictive of Nursing WWST. Using the first SMI definition, RUG-III explained 32% of the variation in WWST for those with SMI compared with 30% for those without SMI. ; For the second SMI definition, RUG-III explained 34% of the variance for those with SMI and 30% for those without SMI.. These results indicate that SMI would not be useful in splitting RUG-III groups.

Finally, we examined the option of adding a full, new SMI category to RUG-III, and considered possible locations in the hierarchy. Based on the relatively low WWST of individuals with SMI, it was clear that if such a category were formed, it would need to be below the Clinically Complex category. Thus there were three options to insert a SMI category, as shown in Table 4-40, i.e., before either the Impaired Cognition (I), the Behavior Problem (B), or the Reduced Physical Function (P) categories (note that this RUG-III analysis was performed before combining the Impaired Cognition and Behavior Problem categories) in RUG-IV. Results concerning placement of an SMI category in RUG-III are presented in Table 4-40.

[89] Run as a regression of WWST = a+b*RUG, where WWST is nursing WWST and RUG was the categorical variable of all 53 RUG-III groups; the residual observed minus regression predicted WWST for each observation was saved and averaged separately for all residents with and without SMI.

Table 4-40. Characteristics of Potential RUG-III SMI category, for three options of category placement

No SMI Category			New SMI category above Impaired Cognition (I)		
Category*	**Number**	**Mean Nursing WWST**	**Category***	**Number**	**Mean Nursing WWST**
			SMI	444	72.2
I	633	84.5	I	398	86.4
B	39	71.2	B	12	56.6
P	2,044	112.1	P	1,862	117.2
New SMI category above Behavior Problem (B)			**New SMI category above Reduced Physical Function (P)**		
Category*	**Number**	**Mean Nursing WWST**	**Category***	**Number**	**Mean Nursing WWST**
I	633	84.5	I	633	84.5
SMI	209	62.0	B	39	71.2
B	12	56.6	SMI	182	59.6
P	1,862	117.2	P	1,862	117.2

* Category code: I=Impaired Cognition; B=Behavior Problem; P=Reduced Physical Function

For example, consider what would happen if we were to place a new SMI category above the Impaired Cognition (RUG-IV Version 1) category, i.e., between the Clinically Complex and Impaired Cognition categories. The new SMI category would consist of 444 individuals with a mean Nursing WWST of 72.2. These individuals are no longer in the three lower categories, with 235 individuals from I, 27 from B and 182 from P. After the formation of this new category, those remaining in the Impaired Cognition group would then consist of 398 individuals without SMI and a mean Nursing WWST of 86.4. This would therefore cause an inversion, with the SMI category, higher in the hierarchy, with lower mean WWST.

As demonstrated in Table 4-40, two of the three possible placements of a new SMI category (above Impaired Cognition and above Reduced Physical Function) would produce hierarchy category inversions – the new SMI category would have a mean Nursing WWST that was ***lower*** than the next hierarchy category. Placing the SMI category above Behavior Problems did not produce an inversion between SMI and Behavior Problems, but the WWST means for these two groups did not differ much (62.0 for SMI and 56.6 for Behavior Problems).

With regard to RUG-IV, we considered adding SMI as an additional variable after fitting RUG-IV. We only describe the results for the first SMI definition; the results for the second definition are similar. We compared regression equations with only RUG-IV as a (categorical) dependent variable with a model that included RUG-IV and SMI (as a

dichotomous variable). [90] The second model, with SMI, explains very minimally more variance (an increase in R^2 of only 0.05%). These results indicate that SMI would not be useful in splitting RUG-IV groups.

We evaluated the effect on overall fit (variance explanation) of the potential RUG-IV model if we created a new SMI category within the RUG-IV hierarchy. These analyses were performed before the decision to combine the cognition and behavior categories into the new RUG-IV Behavior Symptoms and Cognitive Performance category, and placement was test above an Impaired Cognition category, above a Behavior Problem category, and above the Reduced Physical Function category.. In particular, we ran a regression for each of these placements with and without the new SMI category, to see if the variance explanation increased.[91] The results are shown in Table 4-41. For each placement of the SMI category, the variance explanation remained essentially the same with and without the SMI category.

Table 4-41. Variance Explanation* For RUG-IV Models With Uncombined Cognition and Behavior Categories With and Without Additional SMI Category, for Three Options of SMI Category Placement

Placement of New SMI** Category Above Category:	Variance Explanation		
	With No SMI category	With SMI category	Difference
Impaired Cognition	23.6%	23.4%	0.2%
Behavior Problem	23.6%	23.5%	0.1%
Reduced Physical Function	23.6%	23.6%	0.0%

* Analysis based on 2,716 observations in the three categories listed.
** SMI definition does not exclude persons with dementia

As SMI is a subpopulation of nursing facility residents with special needs, we were surprised to find that they were consistently less resource-intensive then other residents with similar characteristics. It would be important to understand better whether the most resource intensive SMI residents are either discharged or diverted to other settings. In any case, we could find no rationale for incorporating an SMI category into the RUG-IV classification system.

[90] Two regressions were contrasted. The first was WWST = a+b*RUG, where WWST was nursing WWST and RUG was the categorical variable for the 66 RUG-IV, Version 1 groups. The second was WWST = a+b*RUG+c*SMI, where WWST and RUG were as before, and SMI was the 0-1 dichotomous variable indicating the presence of SMI without removing those with dementia (first definition).

[91] Run as two regressions of WWST=a+b*RUG, where WWST was nursing WWST. In the first model, RUG was a categorical variable with all of the 66 RUG-IV, Version 1 groups (including splits for ADL and restorative nursing), but limited to the 2,716 observations in the Impaired Cognition, Behavior Problems, and Reduced Physical Function categories. In the second regression, the same model was run, except that the RUG categorical variable now identified two additional SMI groups, split using the RUG-III ADL index (levels 4-5, and 6-10) and restorative nursing in the same manner as splits are performed in the Impaired Cognition category.

4.12.2 Younger Residents

As many nursing facilities have moved their focus increasingly to rehabilitation care, there now is a substantial population of younger residents seen in these facilities. Of the derivation sample of 6,464 residents (including the 10 observations unclassifiable in RUG-III but not yet eliminated from the sample), there were 778 (18%) who were less than 60 years old. Residents between the ages of 50 and 59 comprised 59% of this younger group.

In general, no matter what age was used as the threshold to define "younger," older and younger age groups had similar resource use. For example, the mean nursing WWST is 153.3 for those under age 60 and 147.6 for those age 60 and over (the equivalent numbers for mean nursing plus therapy WWST were 180.6 for younger residents and 185.4 for the older group).

These results do not, however, adjust for characteristics known to be associated with differences in resource use, i.e., for RUG-IV group. We considered the resource use between older and younger age groups using several alternative age thresholds. For each alternative threshold, we computed the residual Nursing WWST for every resident after adjusting for the fitted mean WWST for each RUG-IV group (see Table 4-42).[92] For example, using age 50 as the threshold for the younger group, 320 residents in the derivation sample would be classified as "young." Further, the mean residual Nursing WWST, after fitting RUG-IV,[93] was 12.8 more for those 49 years of age or younger, compared to those age 50 or older. As the threshold decreased, the number of individuals identified as "young" decreased but their residual mean Nursing WWST also increased. Examining these differences, we determined the best upper limit in defining the younger population for the purposes of case-mix analysis was at less than age 50. This threshold allowed a sufficient number of young residents to entertain age based end splits in some categories or a separate category for young residents.

Table 4-42. Number of Observations and Mean Nursing WWST Residuals, for Alternative Age Thresholds for "Young"

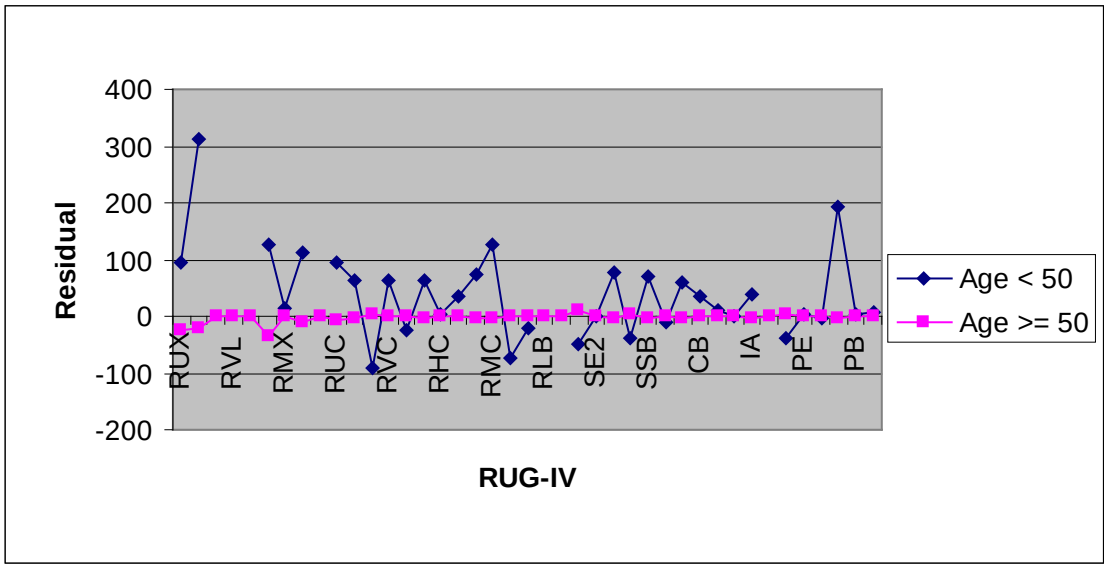
Age Threshold	Number < Age Threshold	Percent of Sample	Mean Nursing WWST Residuals		
			Resident < Age Threshold	Resident >= Age Threshold	Difference
40	87	1 %	29.5	-0.4	29.9
45	175	3 %	21.9	-0.6	22.5
50	320	5 %	12.2	-0.6	12.8
55	519	8 %	6.3	-0.5	6.8
60	778	12 %	4.6	-0.6	5.3

[92] Run as a regression of WWST = a+b*RUG, where WWST was the nursing WWST and RUG was the categorical variable of all 66 RUG groups of RUG-IV Version 2, using the derivation sample (N=6,454); the residual observed minus regression predicted WWST for each observation was saved and averaged separately for all residents within different age group.

[93] All analyses in this section used RUG-IV, Version 2.

To test how well the RUG system explains resource use, we examined the variance explanation of nursing WWST. As the sample sizes for some RUG groups within the under-50 population were small, we used the RUG-IV CMI as the independent variable[94] rather than RUG-IV group. The variance explanation of Nursing WWST is larger for those under age 50 than over age 50 (R^2=47.7% vs. 36.2%), indicating that RUG-IV is relatively more effective in explaining differences in resource use in the younger population than the older population. The same conclusion is reached using the WWST nursing plus therapy WWST. Also, there was no systematic source of difference in the residuals by specific RUG-IV categories or coherent clusters of RUG-IV groups (see Figure 4-18). This indicated that end splits based on age are not feasible within selected categories. Also a separate category for young residents below the Clinically Complex category was not feasible due to the low number of young residents below that category.

Figure 4-18. Mean Residuals for each RUG-IV Group, for Age <50 and Age >=50



, Although age could not be incorporated into the RUG system, we did perform additional analyses relevant to age. We found AIDS/HIV and Traumatic Brain Injury (TBI) to have high incidence rates (27% and 13%) for residents less than age 50 (see Table 4-43 for the prevalence by age range). We also considered AIDS and TBI as special populations of interest (see Section 4.12.3 and 4.12.5).

[94] Run as a regression of WWST = a+b*CMI, where CMI was the continuous variable of mean in each 53 RUG groups divided by overall mean. Analysis run on full derivation sample (N=6,454)

Table 4-43. Prevalence of Autoimmune Deficiency Syndrome (AIDS) and Traumatic Brain Injury (TBI) by Age Range

Age Range (years)	Number in Age Range	Mean Nursing plus Therapy WWST	AIDS		TBI	
			Number	Percentage	Number	Percentage
<20	1	867.85	0	0%	1	100%
20-24	8	335.89	0	0%	3	38%
25-29	18	250.85	2	11%	7	39%
30-34	23	223.82	5	22%	4	17%
35-39	37	220.89	13	35%	7	19%
40-44	88	173.97	32	36%	5	6%
45-49	145	178.79	34	23%	13	9%
50-54	199	164.86	41	21%	6	3%
55-59	259	174.04	25	10%	8	3%
60+	5686	185.38	25	0%	24	0%
TOTAL*	**6464**	**184.81**	**177**	**3%**	**78**	**1%**

* Derivation sample, prior to eliminating 10 observations with missing RUG-III information

4.12.3 AIDS/HIV

The current Medicare Skilled Nursing Facility PPS provides a 128% "add on" to the RUG-III base rate for all residents with AIDS/HIV. Recognizing that it would be better if this indicator could be incorporated into RUG-IV, we examined closely the 253 nursing facility residents with AIDS/HIV (hereafter referred to simply as "AIDS") representing 2.6% of our full sample. It should be noted that in several states[95] involved in STRIVE data collection, it is illegal to collect information about AIDS/HIV. In these states we cannot identify residents with this illness and, therefore, AIDS patients are included in the "No AIDS" group.

The prevalence of AIDS is 20% among residents below age 60 (222 of 1,135). The age group with the highest incidence is age 40 to 45, where AIDS comprises 38% of the full sample (52 of 137).

Overall, residents with AIDS have lower resource use (121.4) than all other residents (149.0). The same relationship holds considering the mean nursing plus therapy WWSTs: 131.4 for residents with AIDS and 186.6 for all others (see Table 4-44).

[95] Iowa, Illinois, Nevada, Texas, and Washington forbid collection of HIV. States where collection of HIV is allowed are DC, FL, KY, LA, MI, MT, NY, OH, SD, and VA. We collected HIV information in 10 of the 15 states.

Table 4-44. Mean WWST and Mean Residual WWST for those with AIDS and no AIDS

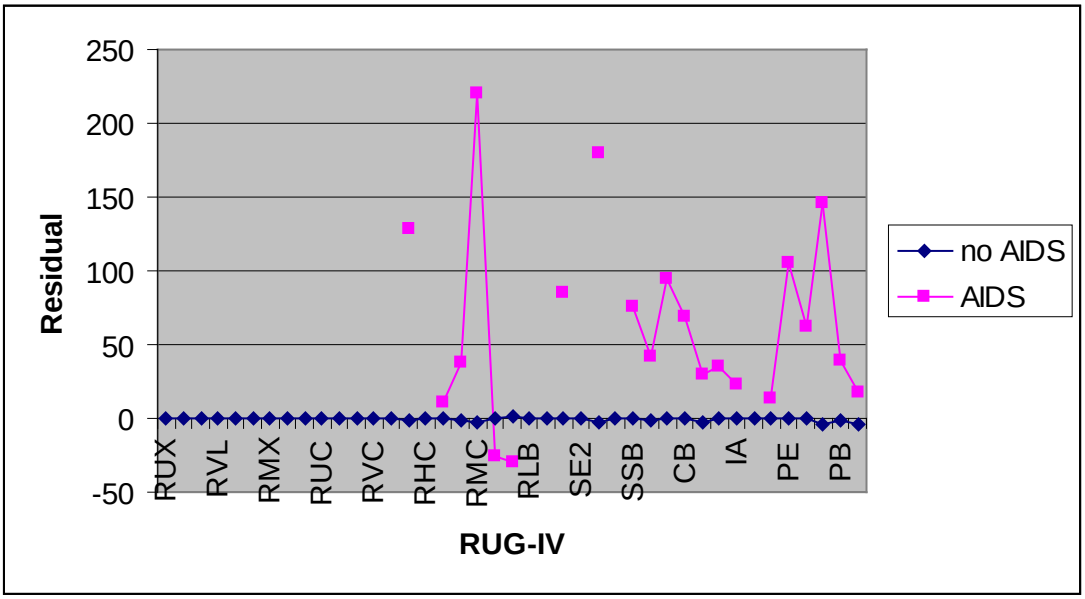
	AIDS	no AIDS	Difference	Variance Explanation by RUG-IV Case Mix in Subsample	
				AIDS	No AIDS
Number of Observations	253	9454			
Mean WWST					
Nursing	121.4	149.0	-27.6	35.0%	37.1%
Nursing + Therapy	131.4	186.6	-55.2	38.7%	58.5%
Mean Residual WWST					
Nursing	30.1	-0.8	30.9	NA	NA
Nursing + Therapy	30.6	-0.8	31.4	NA	NA

However, this does not adjust for already-known case-mix differences between the two populations. To test how well the RUG system explains resource use, we examined variance explanation. Because the sample sizes for AIDS groups are small and to avoid over-specification, we fit a model based on a direct numeric measure of relative resource use[96] rather than using RUG group as an independent variable. The variance explanation of Nursing WWST was 35.0% for AIDS, in contrast to 37.1% for all others. For the 253 out of 9,707 (2.6%) residents who have AIDS, the mean residual of Nursing WWST for residents who have AIDS, after fitting RUG-IV (version 2) is 30.1, while the mean residual of residents without AIDS is -0.8. This shows that AIDS residents do take systematically more care resource. Plots of the RUG-adjusted residuals by RUG group do not suggest any particular trend or pattern for RUG classification of AIDS residents (see Figure 4-19).[97]

[96] Run as a regression of WWST = a+b*CM, where CM was the continuous variable of mean nursing WWST in each 53 RUG-IV, Version 2 groups divided by overall mean. It should be noted that "CM" is different than "CMI" used elsewhere in this report.

[97] The reader should note that the mean residuals for the special population, here AIDS, are often based on very few – or even a single – individual, and can be highly unstable.

Figure 4-19. Residual Plot by AIDS Group after fitting RUG-IV



We also examined conditions associated with AIDS. Body Mass Index (BMI) was significantly related to AIDS. By examining the variance explanation (R^2) of alternate definitions of AIDs, we found that "AIDS and BMI less than 20" identified a resource-intense population with mean WWST in the range of the Special Care categories. However, performing an "in/out" analysis showed that this group was substantially diverse and forcing "AIDS and BMI less than 20" into a Special Care category would not improve the RUG-IV classification system.

We suggest that CMS continue the AIDS resident level "add on" in the payment system based upon RUG-IV, as is currently done in the existing RUG-III PPS. Further research should be done on appropriate payment for residents with AIDS.

4.12.4 Bariatric Residents

It has long been hypothesized that nursing residents who are extremely heavy take additional staff time, e.g. perhaps involving two aides to move them around in bed or to a chair, and therefore their care is more resource intensive. In RUG-III, there was no qualifier for high-weight residents. For RUG-IV, we considered whether high weight should be a qualifier.

For these analyses, we considered two alternative identifications of high-weight individuals: based upon a) weight and b) a commonly-used measure of obesity, the Body Mass Index (BMI) (a ratio of weight to squared height (kg/m^2). Regression of each on resource use (residual Nursing WWST after adjusting for the mean WWST in each RUG-III group) suggested that identifying the bariatric residents as those above a specific weight threshold was more related to resource use than was identifying those

above a specific BMI threshold.[98] Thus, further analyses focused on weight. These analyses were performed early in STRIVE and used RUG-III categories to develop the residual Nursing WWST.

We considered two weight thresholds. The weight threshold best in explaining differences in Nursing WWST (i.e., the largest variance explanation) was 250 lbs. A total of 137 (2.1%) residents exceeded this threshold.[99] We also examined a 300 lb. threshold which was suggested by members of the Technical Expert Panel. Only 52 (0.9%) residents weighed more than 300 lb.

With either threshold, high weight residents have higher resource use than other residents, with Nursing WWST of 171 for those exceeding 250 lbs., and 183 for those exceeding 300lb., both compared to 148 for all others. Likewise, for nursing plus therapy WWST, the 250+ lbs. residents were at 210 and the 300 lbs. residents at 228, compared to 184 for all others (see Tables 4-45 and 4-46). After adjustment for conditions captured by RUG group categorization ("Mean Residual WWST"), the differences between high weight individuals (both thresholds) and others were reduced to 20 WWST or less.

Table 4-45. Mean WWST for Weight Threshold at 250 pounds

	250 lbs. and above	Below 250 lbs.	WWST Difference	RUG-III CM Variance Explanation in Subsample	
				250 lbs. and above	Below 250 lbs.
Number of Observations	137	6306			
Mean WWST					
Nursing	170.9	147.6	23.3	34.1%	29.8%
Nursing + Therapy	210.2	184.1	26.1	57.0%	54.3%
Mean Residual WWST					
Nursing	16.3	-0.4	16.7	NA	NA
Nursing + Therapy	19.5	-0.4	19.9	NA	NA

[98] Run as regressions of RES= a+b*BAR, where RES was the nursing WWST RUG-III residual representing the difference between individual resident nursing WWST and average nursing WWST in each of the 53 categories of the RUG-III system and BAR represents in each model a dichotomous variable indicating heavy (bariatric) residents. In one model, BAR was defined by weight, in the second, BAR was defined by BMI. For either measure, multiple alternative thresholds were explored. Regressions were run on the derivation sample (N=6,443, as 11 residents lacked usable weight).
[99] Derivation sample, N= 6443; 11 residents lacked usable weight.

Table 4-46. Mean WWST for Weight Threshold at 300 pounds

	300 lbs. and above	Below 300 lbs.	WWST Difference	RUG-III CM Variance Explanation in Subsample[102]	
				300 lbs. and above	Below 300 lbs.
Number of Observations	52	6391			
Mean WWST					
Nursing	183.1	147.9	23.3	61.1%	54.3%
Nursing + Therapy	227.7	184.3	26.1	31.0%	29.9%
Mean Residual WWST					
Nursing	19.2	-0.2	19.4	NA	NA
Nursing + Therapy	19.6	-0.2	19.8	NA	NA

We also examined how well RUG-III explained resource use in the special population of bariatric residents. RUG-III variance explanation was reasonable for both high weight residents (identified by either the 250 or the 300 lbs. thresholds) and similar to the WWST variance explanation of the rest of the sample (see Figures 4-20 and 4-21).[100]

Qualitatively, RUG-adjusted residuals[101] for high weight residents appeared greater in the more resource intense groups, i.e., those to the left in Figures 4-20 and 4-21. Residuals for high weight residents are highly variable because they represent small numbers of individuals and are not statistically greater than the residuals for other residents.[102] Alternately stated, by this statistical measure, high weight individuals are typical of their RUG group.

[100] Run as a regression of WWST = a+b*CM, where WWST was either nursing WWST or nursing plus therapy WWST, and CM was the ratio of mean WWST for each of the 53 RUG-III groups to the overall mean WWST. It should be noted that "CM" is different than "CMI" used elsewhere in this report. See Section 6 for further explanation.

[101] Residuals from the regression WWST = a+b*RUG, where the model was run twice, once for WWST set to nursing WWST and once set to nursing plus therapy WWST; RUG is a categorical variables representing the 53 groups of RUG-III.

[102] Wilcoxon Rank Sign test (two-sided) on residual Nursing WWST, p = 0.1233 [250 lbs.] or p = 0.7600 [300 lbs.]. This statistic ranks individual differences from the RUG group average (residuals) and tests whether high weight individuals are more likely to be higher (or lower) ranking than other individuals.

Figure 4-20. Residuals for High Weight Residents [250 lbs.] by RUG-III (53 groups)

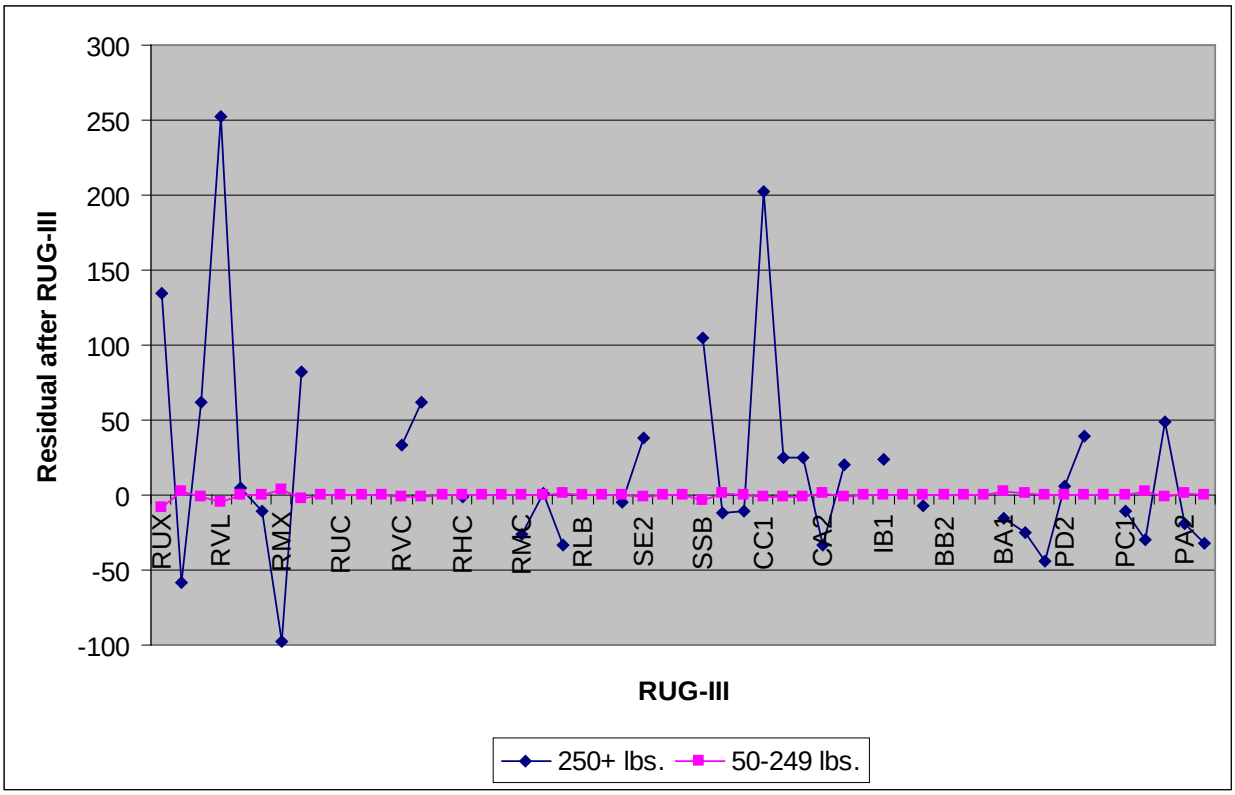
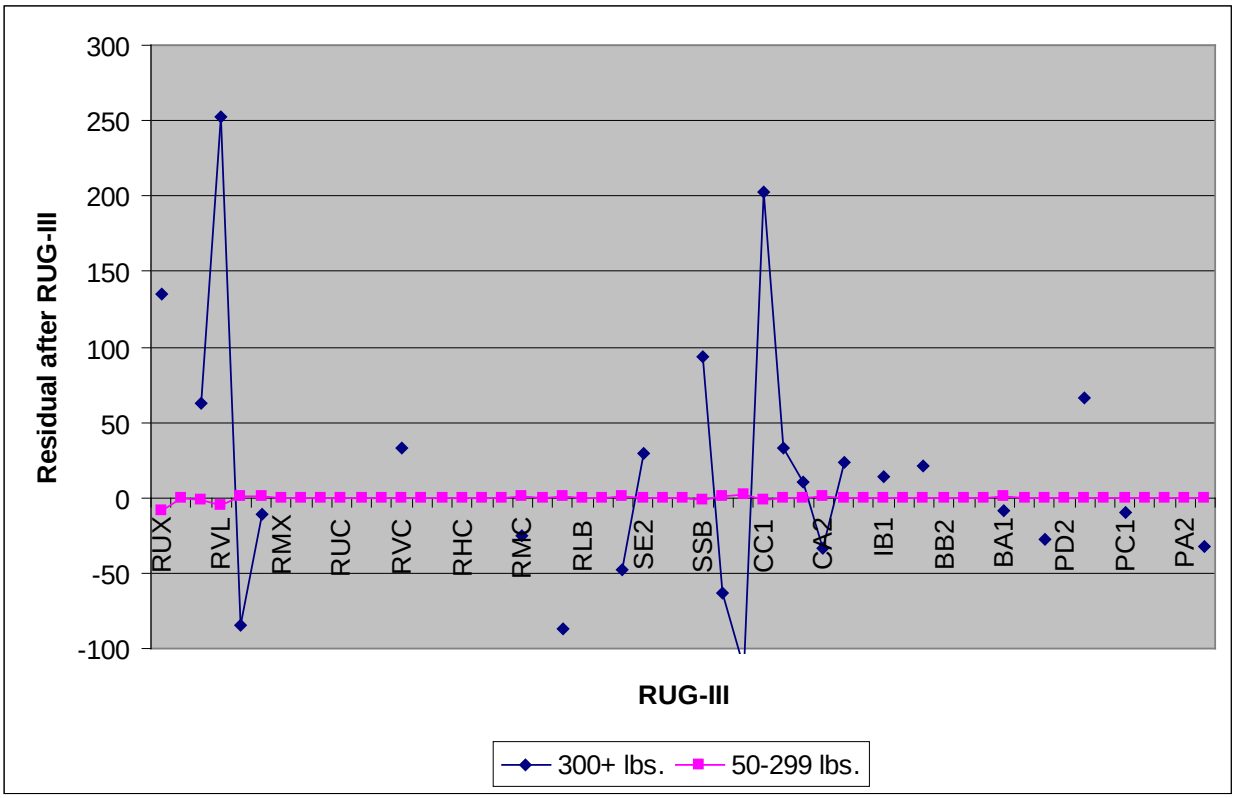


Figure 4-21. Residuals for High Weight Residents [300 lbs.] by RUG-III (53 groups)



As a final check for differences between high weight individuals and others, we obtained the increase in model fit if high weight were incorporated directly into the RUG-IV system. High weight provided negligible additional variance explanation of Nursing WWST (an improvement in R^2 of 0.05% (250 lbs.) or 0.03% (300 lbs.)) when added to a statistical model that already included the RUG-III category.[103] Using any of these approaches, model fit is reasonable for high weight individuals in RUG-III.

We also investigated whether high weight persons were more likely to receive two-person assists than others. First, we examined the use of 2-person transfer assistance. In RUG groups with at least one high-weight person, two-person transfer assists (identified by MDS 2.0 item G1bb = 3) were about equally common for residents under 250 lbs. (n = 1819; 32%) as residents at or above 250 lbs, (N= 49; 36%). Next, we compared RUG-III ADL Index values, since this index gives higher values to residents with two-person assist in bed mobility, transfer, or toilet use. High-weight residents did not have higher RUG-III ADL Index values than others.[104] Therefore, although residents that require two-person assist in mobility, transferring, and toilet use are more

[103] Run as regressions of nursing WWST = a+b*RUG+c*BAR, where RUG is the categorical variable describing all 53 categories of the RUG-III system and BAR represents in each model a dichotomous variable indicating heavy (bariatric) residents and run on the derivation sample (N=6,545).
[104] Chi Square = .20.

resource-intensive than other residents, assistance for high-weight persons is not distinguishably more frequent than for others. No alteration to the RUG-III ADL Index (or the ADL-IV Index, which also employs two-person assist in its coding – see Section 4.7) appears necessary to capture assistance for high-weight residents.

Accordingly, we concluded that resource utilization linked to high weight was already substantially present in the RUG-III system and that RUG-IV would not be improved by adding High Weight explicitly, either as a criterion or in the ADL-IV Index.

4.12.5 Traumatic Brain Injury

The presence of Traumatic Brain Injury (TBI) was examined as an additional possible RUG-IV qualifier. As TBI also can be comorbid with behavior problems, we considered their joint effect on resource use. Given the low prevalence of TBI, our analysis involved the full analytic sample.

TBI was seen in 118 (1.2%) of the full sample of 9,707. The prevalence rate was 7% for residents of age less than 60 (81 out of 1135) and is highest in age group 25 to 30 (37%, or 10 out of 27); there are few over the age of 60.[105]

Residents with TBI have higher resource use (Nursing WWST of 186) than residents without TBI (147.8). For nursing plus therapy WWST, the equivalent means are 232.2 for TBI and 184.5 for those without TBI (see Table 4-47).

Table 4-47. Mean WWST for those with TBI and no TBI

	TBI	no TBI	WWST Difference	RUG-IV CM Variance Explanation in Subsample	
				TBI	No TBI
Number of Observations	118	9589			
Mean WWST					
Nursing	186.0	147.8	-38.2	38.4%	36.7%
Nursing + Therapy	232.2	184.5	-47.7	62.5%	58.3%
Mean Residual WWST					
Nursing	5.6	-0.1	-5.7	NA	NA
Nursing + Therapy	10.3	-0.1	-10.4	NA	NA

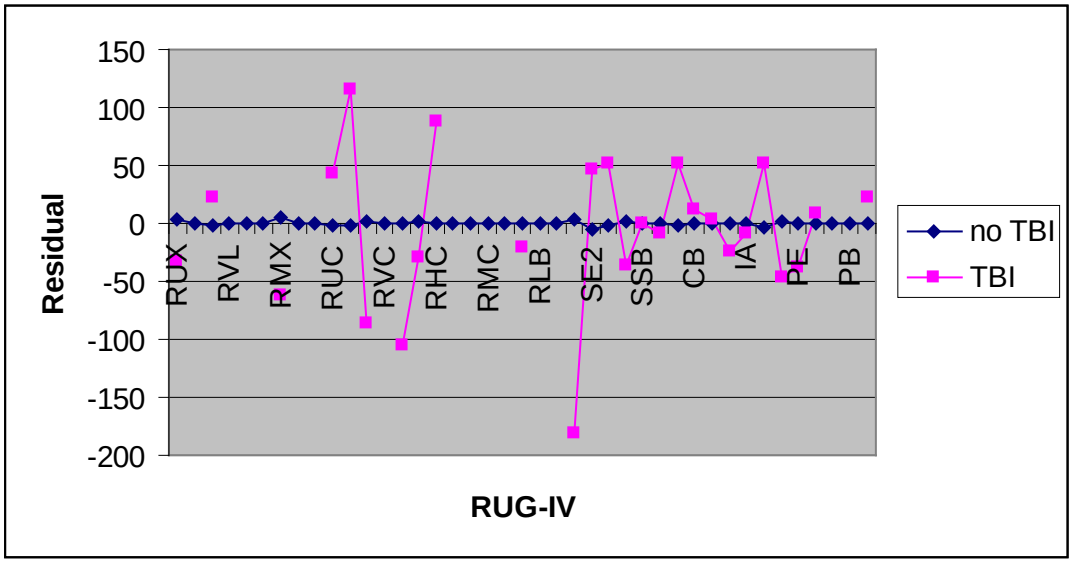
However, these initial results do not adjust for the characteristics known to be associated with differences in resource use, i.e., for RUGs. We thus considered how much more resource intense were those with TBI in a RUG group compared with all

[105] Similar proportions were found in the derivation sample.

others in that RUG group. We computed the residual WWST for every observation after adjusting for the fitted mean WWST for each group.[106]

The mean residual Nursing WWST, after fitting RUG-IV,[107] is 5.6 for those with TBI and -0.1 for those without TBI (see Table 4-47). But, the residual plot of Nursing WWST (Figure 4-22) does not suggest any particular trend or pattern for RUG classification of TBI residents. For example, TBI residents in lower RUG-IV groups do not routinely have higher WWST than their group's average WWST, thereby indicating that they could be classified in a higher RUG-IV category.[108] While we could not determine the reasons for this, it is possible that many state Medicaid agencies reimburse TBI patients outside their dedicated case mix systems. In some states TBI patients with behavior problems are referred to specialized facilities where not included in this time study.

Figure 4-22. Residual Plot by TBI group after fitting RUG-IV



Furthermore, in the TBI population, RUG groups were significantly predictive of Nursing WWST. The variance explanation[109] of Nursing WWST for TBI residents was 38.4%, comparing favorably to the 36.7% for all other residents in the full STRIVE

[106] Run as a regression of WWST = a+b*RUG, where RUG was the categorical variable of all 53 RUG groups; the residual WWST for each observation was saved and averaged separately for all residents with and without SMI.
[107] RUG-IV, Version 2
[108] See The reader should note that the mean residuals for the special population, here TBI, are often based on very few – or even a single – individual, and can be highly unstable.
[109] Run as a regression of nursing WWST = a+b*CMI, where CMI was the continuous variable of nursing WWST mean in each 53 RUG groups divided by overall mean. CMI was used here rather than the categorical variable of all RUG-IV groups because the sample size for TBI groups is small

sample (see Table 4-47). This higher variance explanation was also seen for therapy and Nursing WWST: 62.5% for TBI compared with 58.3% for all others. Even if residents with TBI have higher resource use than residents without TBI, the RUG-IV system explains resource use well for residents with TBI.

Another analytic approach we tried considered the relationship among TBI, behavior problems, and higher resource use. We defined the presence of behavior problems by any one of the following six conditions:

- Verbally abusive behavior occurred on 1 or more days in last week
- Physically abusive behavior occurred on 1 or more days in last week
- Socially inappropriate/disruptive behavior occurred on 1 or more days in last week
- Violent ideation occurred previously
- Intimidation of others or threatened violence occurred previously
- Violence to others occurred previously.

Using this definition, 39 of the 118 TBI residents in the full sample have behavior problems (see Table 4-48). Examining the 273 residents of age less than 45, the mean Nursing WWST is lower in TBI residents with behavior problems than in TBI residents with no behavior problems (mean Nursing WWST of 176.7 and 276.7, respectively). This finding is repeated in the 1,296 residents aged 45 to 64 (mean Nursing WWST 146.2 and 177.9, respectively). However, this finding reverses in those over age 65, where the residents with both TBI and behavior problems have mean Nursing WWST of 183.5, compared to 136.3 for those with TBI but no behavior problems (see Table 4-48). These relationships above and below age 65 remain after adjusting for RUG-IV.

Table 4-48. Number of Observations and Mean Nursing WWST for those with TBI and with or without Behavior Problems (BP)

Age	Number of Observations	Number of TBI Residents		Mean Nursing WWST for TBI Residents			Mean Residual Nursing WWST for TBI Residents		
		With BP	No BP	With BPs	No BP	Difference	With BP	No BP	Difference
<45	273	9	27	176.7	276.7	-100.0	14.1	35.5	-21.4
45-65	1,296	21	26	146.2	177.9	-31.7	-6.1	-3.9	-2.2
65+	8,138	9	26	183.5	136.3	47.2	47.4	-24.0	71.4
TOTAL	**9,707**	**39**	**79**	**161.8**	**198.0**	**-36.2**	**10.9**	**3.0**	**7.9**

On the basis of these results, and particularly the inversion that TBI residents with behavior problems take less time for the most prevalent age groups, it does not appear that including behavior in the identification of TBI helps in classifying these residents.

5 Characteristics of RUG-IV System

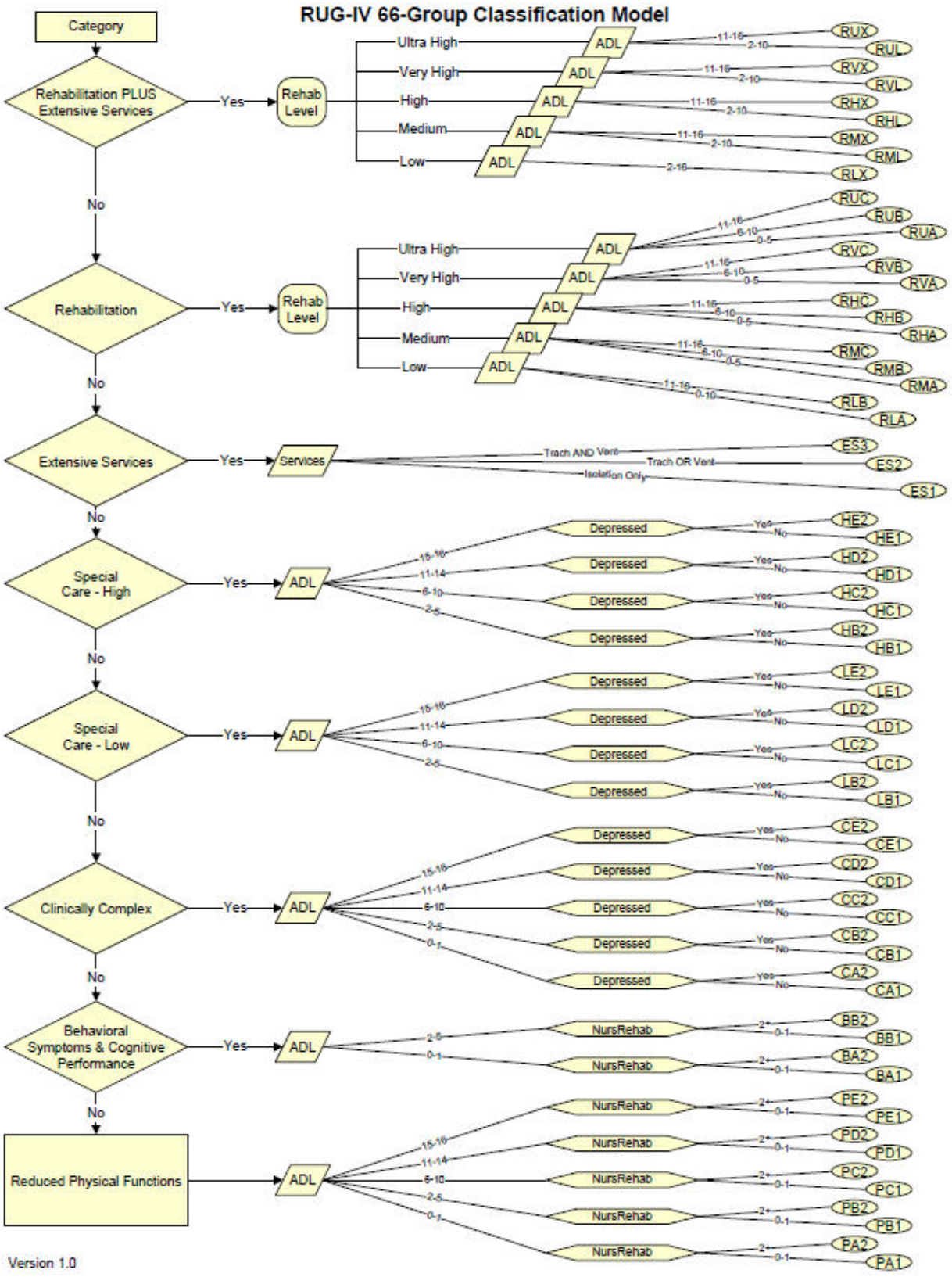
This section documents the final RUG-IV system derived from the analyses described in the previous section.[110]

5.1 Description of RUG-IV

The final RUG-IV system of 66 groups appears in Figure 5-1, displaying the hierarchy categories down the left side of the chart, as well as secondary and tertiary splits. The 66 RUG-IV groups are displayed as the ovals on the right surrounding the name of each group.

[110] The final version is RUG-IV Version 9, as described in Table 5.1. All discussion of RUG-IV in this section refers to this version.

Figure 5-1. RUG-IV 66-Group Classification Model



The details of the characteristics used for each of the splits, including the criteria and qualifiers for each of the hierarchy categories, are displayed in Table 5-1. Full specification of the logic to produce RUG-IV is provided in the SAS computer code (attached as Appendix A) based on MDS 2.0 and STRIVE Addendum items.

Table 5-1. RUG-III to RUG-IV Comparison

MAJOR RUG-III CLASSIFICATION CATEGORY REQUIREMENTS	MAJOR RUG-IV CLASSIFICATION CATEGORY REQUIREMENTS	RUG-III			RUG-IV		
ULTRA HIGH REHABILITATION PLUS EXTENSIVE SERVICES	**ULTRA HIGH REHABILITATION PLUS EXTENSIVE SERVICES**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents needing both extensive medical services and physical or occupational therapy or speech-language pathology services. Rehabilitation Rx 720 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week AND A second rehabilitation discipline at least 3 days/week AND IV feeding in last 7 days OR IV medications, suctioning, tracheostomy care, or, ventilator/respirator in the last 14 days AND ADL score of 7 or more	Residents needing both extensive medical services and physical or occupational therapy or speech-language pathology services. Rehabilitation Rx 720 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week AND A second rehabilitation discipline at least 3 days/week AND Tracheostomy care, ventilator/respirator, or isolation for active infectious disease while a resident AND ADL score >=2	16-18 7-15	RUX RUL	Not used Not used	11-16 2-10	RUX RUL	Not used Not used
VERY HIGH REHABILITATION PLUS EXTENSIVE SERVICES	**VERY HIGH REHABILITATION PLUS EXTENSIVE SERVICES**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents needing both extensive medical services and physical or occupational therapy or speech-language pathology services. Rehabilitation Rx 500 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week AND IV feeding in last 7 days OR IV medications, suctioning, tracheostomy care, or, ventilator/respirator in the last 14 days AND ADL score of 7 or more	Residents needing both extensive medical services and physical or occupational therapy or speech-language pathology services. Rehabilitation Rx 500 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week AND Tracheostomy care, ventilator/respirator, or isolation for active infectious disease while a resident AND ADL score >=2	16-18 7-15	RVX RVL	Not used Not used	11-16 2-10	RVX RVL	Not used Not used

MAJOR RUG-III CLASSIFICATION CATEGORY REQUIREMENTS	MAJOR RUG-IV CLASSIFICATION CATEGORY REQUIREMENTS	RUG-III			RUG-IV		
HIGH REHABILITATION PLUS EXTENSIVE SERVICES	**HIGH REHABILITATION PLUS EXTENSIVE SERVICES**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents needing both extensive medical services and physical or occupational therapy or speech-language pathology services. Rehabilitation Rx 325 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week AND IV feeding in last 7 days OR IV medications, suctioning, tracheostomy care, or, ventilator/respirator in the last 14 days AND ADL score of 7 or more	Residents needing both extensive medical services and physical or occupational therapy or speech-language pathology services. Rehabilitation Rx 325 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week AND Tracheostomy care, ventilator/respirator, or isolation for active infectious disease while a resident AND ADL score >=2	13-18 7-12	RHX RHL	Not used Not used	11-16 2-10	RHX RHL	Not used Not used
MEDIUM REHABILITATION PLUS EXTENSIVE SERVICES	**MEDIUM REHABILITATION PLUS EXTENSIVE SERVICES**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents needing both extensive medical services and physical or occupational therapy or speech-language pathology services. Rehabilitation Rx 150 minutes/week minimum AND 5 days any combination of 3 rehabilitation disciplines; AND IV feeding in last 7 days OR IV medications, suctioning, tracheostomy care, or, ventilator/respirator in the last 14 days AND ADL score of 7 or more	Residents needing both extensive medical services and physical or occupational therapy or speech-language pathology services. Rehabilitation Rx 150 minutes/week minimum AND 5 days any combination of 3 rehabilitation disciplines; AND Tracheostomy care, ventilator/respirator, or isolation for active infectious disease while a resident AND ADL score >=2	15-18 7-14	RMX RML	Not used Not used	11-16 2-10	RMX RML	Not used Not used

MAJOR RUG-III CLASSIFICATION CATEGORY REQUIREMENTS	MAJOR RUG-IV CLASSIFICATION CATEGORY REQUIREMENTS	RUG-III			RUG-IV		
LOW REHABILITATION PLUS EXTENSIVE SERVICES	**LOW REHABILITATION PLUS EXTENSIVE SERVICES**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents needing both extensive medical services and physical or occupational therapy or speech-language pathology services. Rehabilitation Rx 45 minutes/week minimum AND 3 days any combination of 3 rehabilitation disciplines; AND Nursing rehabilitation, 2 or more services, 6 or more days/week (see Reduced Physical Function for nursing rehab services count) AND IV feeding in last 7 days OR IV medications, suctioning, tracheostomy care, or, ventilator/respirator in the last 14 days AND ADL score of 7 or more	Residents needing both extensive medical services and physical or occupational therapy or speech-language pathology services. Rehabilitation Rx 45 minutes/week minimum AND 3 days any combination of 3 rehabilitation disciplines AND Restorative nursing, 2 or more services, 6 or more days/week (see Reduced Physical Function for restorative nursing services) AND Tracheostomy care, ventilator/respirator, or isolation for active infectious disease while a resident AND ADL score >=2	7-18	RLX	Not used	2-16	RLX	Not used
ULTRA HIGH REHABILITATION	**ULTRA HIGH REHABILITATION**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents receiving physical or occupational therapy, or speech-language pathology services Rehabilitation Rx 720 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week AND A second rehabilitation discipline at least 3 days/week	Residents receiving physical or occupational therapy, or speech-language pathology services Rehabilitation Rx 720 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week AND A second rehabilitation discipline at least 3 days/week	16-18 9-15 4-8	RUC RUB RUA	Not Used Not Used Not Used	11-16 6-10 0-5	RUC RUB RUA	Not Used Not Used Not Used
VERY HIGH REHABILITATION	**VERY HIGH REHABILITATION**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents receiving physical or occupational therapy, or speech-language pathology services Rehabilitation Rx 500 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week	Residents receiving physical or occupational therapy, or speech-language pathology services Rehabilitation Rx 500 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week	16-18 9-15 4-8	RVC RVB RVA	Not Used Not Used Not used	11-16 6-10 0-5	RVC RVB RVA	Not Used Not Used Not Used

MAJOR RUG-III CLASSIFICATION CATEGORY REQUIREMENTS	MAJOR RUG-IV CLASSIFICATION CATEGORY REQUIREMENTS	RUG-III			RUG-IV		
HIGH REHABILITATION	**HIGH REHABILITATION**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents receiving physical or occupational therapy, or speech-language pathology services Rehabilitation Rx 325 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week	Residents receiving physical or occupational therapy, or speech-language pathology services Rehabilitation Rx 325 minutes/week minimum AND At least 1 rehabilitation discipline 5 days/week	13-18 8-12 4-7	RHC RHB RHA	Not Used Not Used Not Used	11-16 6-10 0-5	RHC RHB RHA	Not Used Not Used Not Used
MEDIUM REHABILITATION	**MEDIUM REHABILITATION**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents receiving physical or occupational therapy, or speech-language pathology services Rehabilitation Rx 150 minutes/week minimum AND 5 days any combination of 3 rehabilitation disciplines	Residents receiving physical or occupational therapy, or speech-language pathology services Rehabilitation Rx 150 minutes/week minimum AND 5 days any combination of 3 rehabilitation disciplines	15-18 8-14 4-7	RMC RMB RMA	Not Used Not Used Not Used	11-16 6-10 0-5	RMC RMB RMA	Not Used Not Used Not Used
LOW REHABILITATION	**LOW REHABILITATION**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents receiving physical or occupational therapy, or speech-language pathology services Rehabilitation Rx 45 minutes/week minimum AND 3 days any combination of 3 rehabilitation disciplines AND Nursing rehabilitation, 2 or more services, 6 or more days/week (see Reduced Physical Function for nursing rehab services count)	Residents receiving physical or occupational therapy, or speech-language pathology services Rehabilitation Rx 45 minutes/week minimum AND 3 days any combination of 3 rehabilitation disciplines AND Restorative nursing, 2 or more services, 6 or more days/week (see Reduced Physical Function for restorative nursing services)	14-18 4-13	RLB RLA	Not Used Not Used	11-16 0-10	RLB RLA	Not Used Not Used

MAJOR RUG-III CLASSIFICATION CATEGORY REQUIREMENTS	MAJOR RUG-IV CLASSIFICATION CATEGORY REQUIREMENTS	RUG-III			RUG-IV		
EXTENSIVE SERVICES	**EXTENSIVE SERVICES**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents receiving the following complex clinical care: IV feeding in last 7 days OR IV medications, suctioning, tracheostomy care, or, ventilator/respirator in the last 14 days AND ADL score of 7 or more	Residents receiving the following complex clinical care: Tracheostomy care while a resident OR Ventilator or respirator while a resident OR Isolation for active infectious disease while a resident AND ADL score >=2	7-18	SE3	Count of other categories (special care, clinically complex, impaired cognition), plus IV medications, plus IV feeding. Extensive Count of 4 or 5	2-16	ES3	Tracheostomy care (while a resident) **AND** ventilator or respirator (while a resident)
		7-18	SE2	Count of other categories (special care, clinically complex, impaired cognition), plus IV medications, plus IV feeding. Extensive Count of 2 or 3	2-16	ES2	Tracheostomy care (while a resident) **OR** ventilator or respirator (while a resident)
		7-18	SE1	Count of other categories (special care, clinically complex, impaired cognition), plus IV medications, plus IV feeding. Extensive Count of 0 or 1	2-16	ES1	Isolation for active infectious disease (while a resident)
		Notes: Comorbidities count for end splits			Notes: Qualifiers count for end splits		

MAJOR RUG-III CLASSIFICATION CATEGORY REQUIREMENTS	MAJOR RUG-IV CLASSIFICATION CATEGORY REQUIREMENTS	RUG-III			RUG-IV		
SPECIAL CARE	**SPECIAL CARE**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Extensive Services qualifier AND ADL of 6 or less; OR Any one of the following Special Care Qualifiers: • cerebral palsy, multiple sclerosis or quadriplegia with and ADL sum > 10; • respiratory therapy for 7 days; • feeding tube (calories > 51%, or calories = 26-50% and fluid > 501 cc) and aphasia; • radiation therapy; • receiving therapy for surgical wounds/open lesions or ulcers (2 sites, any stage; or 1 site stage 3 or 4); • fever with dehydration, pneumonia, vomiting, weight loss, or feeding tube (calories > 51%, or calories = 26-50% and fluid > 501cc) AND ADL score of 7 or more	**SPECIAL CARE HIGH** Residents receiving the following complex clinical care or with a following medical condition: • Comatose and completely ADL dependent; • septicemia; • diabetes with daily injections requiring physician order changes on 2 or more days; • quadriplegia and ADL score >=5; • chronic obstructive pulmonary disease and shortness of breath when lying flat; • fever with pneumonia, or vomiting, or weight loss, or feeding tube(with calories > 51%, or calories = 26-50% and fluid > 501 cc); • parenteral/IV feedings; • respiratory therapy for 7 days AND ADL score >=2	17-18 15-16 4-14	SSC SSB SSA	Not Used Not Used Not Used	15-16 15-16 11-14 11-14 6-10 6-10 2-5 2-5	HE2 HE1 HD2 HD1 HC2 HC1 HB2 HB1	Signs of Depression No Signs of Depression Signs of Depression No Signs of Depression Signs of Depression No Signs of Depression Signs of Depression No Signs of Depression
					Notes: Signs of depression indicator used for end splits is the same as RUG-III signs of depression for the Clinically Complex category (see RUG-III "End Splits" column for Clinically Complex below). Notes: For conversion to MDS 3.0, signs of depression used for end splits; PHQ score in Section D for either the resident interview or staff assessment =>10.		
	SPECIAL CARE LOW Residents receiving the following complex clinical care or with a following medical condition: • Cerebral palsy and ADL score >=5; • multiple sclerosis and ADL score >=5; • Parkinson's disease and ADL score >=5; • respiratory failure and oxygen therapy while a resident; • feeding tube (calories > 51%, or calories = 26 -50% and fluid > 501 cc); • ulcers (2 or more stage II or 1 or more stage III or IV pressure ulcers; or 2 or more venous/arterial ulcers; or 1 stage II pressure		SSC SSB SSA	Not Used Not Used Not Used	15-16 15-16 11-14 11-14 6-10 6-10 2-5 2-5	LE2 LE1 LD2 LD1 LC2 LC1 LB2 LB1	Signs of Depression No Signs of Depression Signs of Depression No Signs of Depression Signs of Depression No Signs of Depression Signs of Depression No Signs of Depression

MAJOR RUG-III CLASSIFICATION CATEGORY REQUIREMENTS	MAJOR RUG-IV CLASSIFICATION CATEGORY REQUIREMENTS	RUG-III			RUG-IV		
SPECIAL CARE	**SPECIAL CARE**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
	ulcer and 1 venous/arterial ulcer) • with 2 or more skin treatments; • foot infection, diabetic foot ulcer, or open lesions on the foot with treatment; • radiation therapy while a resident; • dialysis while a resident AND ADL score >=2				Notes: Signs of depression indicator used for end splits is the same as RUG-III signs of depression for the Clinically Complex category (see RUG-III "End Splits" column for Clinically Complex below). Notes: For conversion to MDS 3.0, signs of depression used for end splits; PHQ score in Section D for either the resident interview or staff assessment =>10.		

MAJOR RUG-III CLASSIFICATION CATEGORY REQUIREMENTS	MAJOR RUG-IV CLASSIFICATION CATEGORY REQUIREMENTS	RUG-III			RUG-IV		
CLINICALLY COMPLEX	**CLINICALLY COMPLEX**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Special Care qualifier AND ADL score of 6 or less OR Any one of the following clinically complex qualifiers: • Burns; • coma and not awake and completely ADL dependent; • septicemia; • pneumonia, • foot infection/wound with treatment; • internal bleeding; • dehydration; • tube feeding (calories $\geq$ 51%, or calories = 26%-50% and fluid $\geq$ 501 cc); • oxygen therapy; • transfusions; • hemiplegia with ADL score > 10; • chemotherapy; • dialysis; • physician visits 1 or more days and order changes 2 or more days (last 14 days); • diabetes with injection 7 days/week requiring order change 2 days or more days (last 14 days);	Residents with Extensive Services, Special Care High, or Special Care Low qualifier AND ADL score = 0 or 1 OR Residents with any one of the following clinically complex qualifiers: • Pneumonia; • hemiplegia and ADL score >=5; • surgical wounds or open lesions with treatment; • burns; • chemotherapy while a resident; • oxygen therapy while a resident; • IV medications while a resident; • transfusions while a resident	17-18	CC2	Signs of Depression	15-16	CE2	Signs of Depression
		17-18	CC1	No Signs of Depression	15-16	CE1	No Signs of Depression
		12-16	CB2	Signs of Depression	11-14	CD2	Signs of Depression
		12-16	CB1	No Signs of Depression	11-14	CD1	No Signs of Depression
		4-11	CA2	Signs of Depression	6-10	CC2	Signs of Depression
		4-11	CA1	No Signs of Depression	6-10	CC1	No Signs of Depression
		Notes: Signs of depression used for end splits are indicated by three or more of any of the following 16 MDS 2.0 mood items (Items at E1) exhibited in the last 30 days: negative statements, repetitive questions, repetitive verbalizations, persistent anger, self-deprecation, unrealistic fears, recurrent statements that something terrible is going to happen, repetitive health complaints, repetitive non-health complaints/concerns, unpleasant mood in morning, insomnia/changes in usual sleep pattern, sad/pained/worried facial expression, crying/tearfulness, repetitive physical movements, withdrawal from activities of interest, and reduced social interaction.			2-5	CB2	Signs of Depression
					2-5	CB1	No Signs of Depression
					0-1	CA2	Signs of Depression
					0-1	CA1	No Signs of Depression
					Notes: Signs of depression indicator used for end splits is the same as RUG-III signs of depression for the Clinically Complex category (see column to the left). Notes: For conversion to MDS 3.0, signs of depression used for end splits; PHQ score in Section D for either the resident interview or staff assessment =>10.		

MAJOR RUG-III CLASSIFICATION CATEGORY REQUIREMENTS	MAJOR RUG-IV CLASSIFICATION CATEGORY REQUIREMENTS	RUG-III			RUG-IV		
IMPAIRED COGNITION	**BEHAVIORAL SYMPTOMS and COGNITIVE PERFORMANCE**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Score on MDS 2.0 Cognitive Performance Scale (CPS) $\geq$ 3 AND ADL score of 10 or less NOTES: No clinical variables used; CPS Score of "6" will be assigned Clinically Complex or PE2-PD1 See Reduced Physical Function for nursing rehab services count	Residents having cognitive impairment: BIMS score <=9 or CPS >= 3 OR hallucinations OR delusions OR Residents displaying any of the following on 4 or more days over last 7 days: physical behavior symptoms toward others, OR verbal behavioral symptoms toward others, OR other behavioral symptoms, OR rejection of care, OR wandering AND ADL score <=5	6-10	IB2	2 or more nursing rehab services on 6+ days/wk	2-5	BB2	2 or more restorative nursing, 6 or more days/wk
		6-10	IB1	Less nursing rehab	2-5	BB1	Less restorative nursing
		4-5	IA2	2 or more nursing rehab services on 6+ days/wk	0-1	BA2	2 or more restorative nursing, 6 or more days/wk
		4-5	IA1	Less nursing rehab	0-1	BA1	Less restorative nursing
BEHAVIOR PROBLEMS		**ADL**	**CODES**	**END SPLITS**			
Wandering, physical abuse, verbal abuse, inappropriate behavior or resisted care on 4+ days/week OR hallucination or delusions AND ADL score of 10 or less		6-10	BB2	2 or more nursing rehab services on 6+ days/wk			
		6-10	BB1	Less nursing rehab			
		4-5	BA2	2 or more nursing rehab services on 6+ days/wk			
		4-5	BA1	Less nursing rehab			
		Notes: Nursing rehab used for end splits See Reduced Physical Function for nursing rehab services count			Notes: Restorative nursing used for end splits See Reduced Physical Function for restorative nursing services count		

MAJOR RUG-III CLASSIFICATION CATEGORY REQUIREMENTS	MAJOR RUG-IV CLASSIFICATION CATEGORY REQUIREMENTS	RUG-III			RUG-IV		
REDUCED PHYSICAL FUNCTION	**REDUCED PHYSICAL FUNCTION**	**ADL**	**CODES**	**END SPLITS**	**ADL**	**CODES**	**END SPLITS**
Residents whose needs are primarily for activities of daily living and general supervision. Nursing Rehab service count: • passive and/or active ROM • amputation/prosthesis care training • splint or brace assistance • dressing or grooming training • eating or swallowing training • transfer training • bed mobility and/or walking training • communication training • scheduled toileting plan and/or bladder retraining program	Residents whose needs are primarily for activities of daily living and general supervision. Residents not qualifying for other categories Restorative Nursing services: • passive and/or active ROM; • amputation/prosthesis care training; • splint and/or brace assistance; • dressing and/or grooming training; • eating and/or swallowing training; • transfer training; • bed mobility and/or walking training; • communication training; • urinary and/or bowel training program	16-18	PE2	2 or more nursing rehab services on 6+ days/wk	15-16	PE2	2 or more restorative nursing, 6 or more days/wk
		16-18	PE1	Less nursing rehab	15-16	PE1	Less restorative nursing
		11-15	PD2	2 or more nursing rehab services on 6+ days/wk	11-14	PD2	2 or more restorative nursing, 6 or more days/wk
		11-15	PD1	Less nursing rehab	11-14	PD1	Less restorative nursing
		9-10	PC2	2 or more nursing rehab services on 6+ days/wk	6-10	PC2	2 or more restorative nursing, 6 or more days/wk
		9-10	PC1	Less nursing rehab	6-10	PC1	Less restorative nursing
		6-8	PB2	2 or more nursing rehab services on 6+ days/wk	2-5	PB2	2 or more restorative nursing, 6 or more days/wk
		6-8	PB1	Less nursing rehab	2-5	PB1	Less restorative nursing
		4-5	PA2	2 or more nursing rehab services on 6+ days/wk	0-1	PA2	2 or more restorative nursing, 6 or more days/wk
Notes: No clinical variables used	Notes: No clinical variables used	4-5	PA1	Less nursing rehab	0-1	PA1	Less restorative nursing

While RUG-IV retains the overall structure of RUG-III, it also incorporates several major changes, including the following:

- Concurrent therapy time allocated for identification of rehabilitation residents;
- Selected services (see list below) only considered if provided while a resident of the nursing facility;
- Additional category: Special Care Low;
- Impaired Cognition and Behavior categories joined;
- Multiple changes in specific category qualifiers;
  - New items: infection isolation, shortness of breath, Parkinson's disease, oxygen with respiratory failure;
- New ADL Index computations and consistent ADL breaks used across categories;
- Extensive Services category and group labels changed from "SE" used for RUG-III to "ES" used for RUG-IV.

Restorative Nursing, (formerly in RUG-III "Nursing Rehabilitation", was retained as a tertiary split for the lowest two RUG-IV categories.

The changes in RUG-IV require items to be added or modified in the MDS 2.0 to complete the algorithm. The form in which these items are provided in the instrument, including time frame, delimiters, exclusions, and examples, must be developed. These changes include the following:

- Services coded only in the last 7 days, but, if admission is within 7 days, the service must have been performed in the facility:
  - Tracheostomy care
  - Ventilator/respirator
  - Isolation for active infectious disease
  - Parenteral or IV feeding
  - IV medications
  - Transfusions
  - Oxygen therapy
  - Chemotherapy
  - Dialysis
  - Radiation therapy
- Concurrent therapy time provided both as unallocated (i.e., total therapy time provided) and allocated (i.e., time allocated to all individuals in group by the time slice method);

- Diabetes with daily injections;
- Chronic obstructive pulmonary disease (COPD);
- Shortness of breath when lying flat;
- Parkinson’s disease;
- Diabetic foot ulcer.

Finally, the frequency and mean Nursing WWST for each of the RUG-IV groups is shown in Table 5-2 separately for the derivation and validation samples, and in Table 5-3 for the full STRIVE sample.

Table 5-2. Number of Observations (N) and Mean Nursing WWST, Therapy WWST, and Nursing plus Therapy WWST. By RUG-IV Group – Derivation and Validation Samples

RUG-IV Group		DERIVATION SAMPLE						VALIDATION SAMPLE					
			SAMPLE WEIGHTS APPLIED						SAMPLE WEIGHTS APPLIED				
Name	Number	Actual N	WeightedN	%	Nursing WWST	Therapy WWST	Nursing + Therapy WWST	Actual N	Weighted N	%	Nursing WWST	Therapy WWST	Nursing + Therapy WWST
RUX	111	6	1.6	0.00%	346.29	283.88	630.17	5	1.4	0.00%	383.81	279.76	663.56
RUL	112	2	0.4	0.00%	449.76	309.11	758.87	--					
RVX	121	3	1	0.00%	341.48	217.9	559.38	4	2.7	0.10%	214.03	264.5	478.54
RVL	122	8	4.6	0.10%	320.47	181.3	501.76	3	1.7	0.10%	282.92	219.45	502.37
RHX	131	5	2.7	0.00%	426.58	140.25	566.83	6	3	0.10%	378.03	126.15	504.18
RHL	132	12	4.1	0.10%	275.78	152.01	427.79	7	3.6	0.10%	193.7	111.8	305.5
RMX	141	18	9.6	0.10%	360.96	92.81	453.77	3	1.9	0.10%	408.13	99.09	507.22
RML	142	9	9.4	0.10%	479.9	130.4	610.3	3	1	0.00%	357.39	116.06	473.45
RLX	151	--						--					
RUC	211	32	26.3	0.40%	223.81	267.58	491.39	17	16.7	0.50%	215.63	309.5	525.12
RUB	212	29	28.2	0.40%	269.46	309.13	578.59	19	7.8	0.20%	211.12	289.83	500.95
RUA	213	13	14.7	0.20%	161.59	276.27	437.86	12	6.1	0.20%	124.21	282.51	406.73
RVC	221	71	52.1	0.80%	233.64	200.81	434.45	36	14.9	0.50%	212.04	187.03	399.07
RVB	222	87	56.4	0.90%	156.34	200.36	356.71	46	21.6	0.70%	200.09	201.68	401.77
RVA	223	92	95.3	1.50%	168.28	198.14	366.42	49	18.7	0.60%	165.24	189.9	355.14
RHC	231	85	73	1.10%	224.49	143.36	367.85	34	16	0.50%	196.53	146.05	342.58
RHB	232	118	45	0.70%	192.15	139.29	331.45	79	40.7	1.20%	168.86	132.72	301.59
RHA	233	175	131.5	2.00%	144.63	125.32	269.94	102	76	2.30%	128.63	128.15	256.78
RMC	241	119	66.1	1.00%	214.9	83.81	298.71	56	29.6	0.90%	187.09	84.74	271.83
RMB	242	149	89.6	1.40%	184.92	86.72	271.64	88	56.8	1.70%	184.44	82.92	267.36
RMA	243	251	155.6	2.40%	130.14	78.79	208.94	90	76.8	2.40%	122.86	89.75	212.61
RLB	251	13	13.6	0.20%	261.96	31.37	293.33	5	7.7	0.20%	166.35	52.84	219.18

		DERIVATION SAMPLE						VALIDATION SAMPLE					
RUG-IV Group			SAMPLE WEIGHTS APPLIED						SAMPLE WEIGHTS APPLIED				
Name	Number	Actual N	WeightedN	%	Nursing WWST	Therapy WWST	Nursing + Therapy WWST	Actual N	Weighted N	%	Nursing WWST	Therapy WWST	Nursing + Therapy WWST
RLA	**252**	24	21.2	0.30%	113.49	49.57	163.06	4	3.1	0.10%	68.69	38.68	107.37
ES3	**311**	124	68.1	1.00%	406.92	4.57	411.5	76	33.5	1.00%	400.28	4.04	404.32
ES2	**312**	64	39.8	0.60%	309.11	2.28	311.38	37	17.5	0.50%	285.86	3.54	289.4
ES1	**313**	30	21.5	0.30%	256.58	2.95	259.53	10	6.8	0.20%	278.87	14.93	293.8
HE2	**411**	15	13.5	0.20%	239.89	1.47	241.36	6	7.3	0.20%	211.18	0.24	211.42
HE1	**412**	70	85.2	1.30%	197.46	1.41	198.86	33	29.4	0.90%	225.39	11.78	237.17
HD2	**421**	34	33.2	0.50%	263.58	21	284.58	15	14.8	0.50%	205.08	8.21	213.29
HD1	**422**	88	83.6	1.30%	186.48	4.2	190.68	53	80.8	2.50%	165.84	3.54	169.38
HC2	**431**	31	27.7	0.40%	176.29	3.3	179.59	14	14.5	0.40%	220.24	1.07	221.31
HC1	**432**	94	86.2	1.30%	182.54	3.67	186.21	53	38.8	1.20%	157.39	14.5	171.89
HB2	**441**	26	25.5	0.40%	306.96	4.88	311.84	14	9.8	0.30%	245.97	10.95	256.92
HB1	**442**	85	79.4	1.20%	147.32	6.88	154.2	38	26.7	0.80%	113.19	5.35	118.54
LE2	**511**	31	35.2	0.50%	219.57	0.41	219.98	17	26	0.80%	180.54	0.65	181.19
LE1	**512**	135	141.2	2.20%	184.45	5.69	190.14	71	114.3	3.50%	164.95	2.77	167.71
LD2	**521**	50	52.1	0.80%	198.41	1.59	200	25	32	1.00%	180.68	9.56	190.24
LD1	**522**	167	200.4	3.10%	139.65	1.72	141.37	88	158.5	4.90%	154.04	1.16	155.2
LC2	**531**	46	32.3	0.50%	183.67	0.95	184.62	23	21.5	0.70%	124.46	8.35	132.82
LC1	**532**	134	160	2.50%	136.08	3.96	140.04	68	60.7	1.90%	158.4	0.71	159.11
LB2	**541**	15	20.7	0.30%	181.57	0.82	182.39	4	5.6	0.20%	147.12	0.24	147.36
LB1	**542**	59	58.9	0.90%	124.89	8.11	132.99	26	23.4	0.70%	128.48	7.36	135.84
CE2	**611**	23	13.9	0.20%	156.79	0.62	157.41	10	10.4	0.30%	203.36	3.11	206.48
CE1	**612**	45	39.8	0.60%	173.59	1.65	175.24	26	38.3	1.20%	147.5	0.54	148.03
CD2	**621**	45	57.7	0.90%	198.44	8.16	206.6	29	26.2	0.80%	175.54	6.78	182.33
CD1	**622**	124	130.5	2.00%	141.29	3.43	144.73	63	52.6	1.60%	167.89	2.73	170.62
CC2	**631**	66	56.7	0.90%	134.85	3.8	138.65	33	36.7	1.10%	122.11	1.31	123.41
CC1	**632**	186	170.6	2.60%	115.37	2.2	117.57	96	126.2	3.90%	131.11	1.87	132.98

RUG-IV Group		DERIVATION SAMPLE						VALIDATION SAMPLE					
			SAMPLE WEIGHTS APPLIED						SAMPLE WEIGHTS APPLIED				
Name	Number	Actual N	WeightedN	%	Nursing WWST	Therapy WWST	Nursing + Therapy WWST	Actual N	Weighted N	%	Nursing WWST	Therapy WWST	Nursing + Therapy WWST
CB2	641	44	21.4	0.30%	135.22	16.89	152.12	19	27.3	0.80%	122.45	3.38	125.83
CB1	642	113	102.9	1.60%	104.48	3.45	107.93	56	81.7	2.50%	129.62	3.3	132.93
CA2	651	63	68.3	1.10%	88.95	3.98	92.93	22	7.9	0.20%	121.8	2.37	124.17
CA1	652	241	158.1	2.40%	94.54	2.81	97.35	128	123.2	3.80%	84.12	6.6	90.72
BB2	711	62	145.4	2.20%	120.57	2.19	122.76	40	98.6	3.00%	80.69	0.82	81.52
BB1	712	359	537.6	8.30%	104.04	1.53	105.57	169	298.4	9.20%	101.69	3.38	105.07
BA2	721	18	21.7	0.30%	89.68	0.34	90.02	16	25.7	0.80%	96.07	0.57	96.64
BA1	722	401	589.1	9.10%	72.6	0.83	73.43	194	140.7	4.30%	67.89	0.96	68.86
PE2	811	24	34.5	0.50%	171.91	1.63	173.54	12	28.8	0.90%	149.39	0	149.39
PE1	812	153	160.5	2.50%	161.92	3.28	165.19	70	86.9	2.70%	159.12	0.86	159.99
PD2	821	59	121.5	1.90%	155.94	0.45	156.39	35	57.9	1.80%	146.92	1.42	148.34
PD1	822	300	395.6	6.10%	145.65	2.65	148.3	166	159	4.90%	146.04	2.17	148.21
PC2	831	105	222.2	3.40%	108.49	0.3	108.79	55	85.4	2.60%	111.9	0.48	112.38
PC1	832	503	500.7	7.70%	121.7	3.02	124.72	248	317.6	9.70%	118.75	2.34	121.09
PB2	841	46	70.1	1.10%	121.59	3.07	124.66	21	20.8	0.60%	86.1	0.14	86.24
PB1	842	263	277.8	4.30%	81.48	4.07	85.55	124	119.1	3.70%	88.67	5.58	94.25
PA2	851	32	50.6	0.80%	50.07	0.02	50.09	19	22.4	0.70%	39.15	1.86	41.01
PA1	852	560	377.6	5.80%	65.19	2.23	67.43	262	138.7	4.30%	62.21	3.01	65.21
TOTAL		6454	6491.4					3252	3260.4				

Table 5-3. Number of Observations (N); Mean Nursing WWST, Therapy WWST, and Nursing plus Therapy WWST; and Coefficient of Variation (CV). By RUG-IV Group– Full STRIVE Sample

RUG-IV Group			FULL SAMPLE WITH SAMPLE WEIGHTS APPLIED					
Name	Number	Actual N	Estimated N	%	Nursing WWST	Therapy WWST	Nursing + Therapy WWST	Nursing WWST CV
RUX	111	11	3	0.00%	364.14	281.92	646.06	21.22
RUL	112	2	0.4	0.00%	449.76	309.11	758.87	17.45
RVX	121	7	3.7	0.00%	247.72	252.19	499.91	20.36
RVL	122	11	6.3	0.10%	310.56	191.37	501.92	41.94
RHX	131	11	5.7	0.10%	401.33	132.92	534.25	17.79
RHL	132	19	7.7	0.10%	237.69	133.36	371.05	31.47
RMX	141	21	11.6	0.10%	368.88	93.87	462.75	21.23
RML	142	12	10.3	0.10%	468.28	129.04	597.31	43.42
RLX	151	--						
RUC	211	49	43.1	0.40%	220.63	283.88	504.5	37.22
RUB	212	48	36.1	0.40%	256.76	304.93	561.69	48.99
RUA	213	25	20.8	0.20%	150.65	278.1	428.75	66.92
RVC	221	107	67	0.70%	228.84	197.74	426.58	31.06
RVB	222	133	78	0.80%	168.47	200.73	369.2	37.27
RVA	223	141	114	1.20%	167.78	196.79	364.58	42.76
RHC	231	119	88.9	0.90%	219.47	143.84	363.31	33.33
RHB	232	197	85.7	0.90%	181.1	136.17	317.27	34.61
RHA	233	277	207.5	2.10%	138.77	126.35	265.12	46.97
RMC	241	175	95.7	1.00%	206.29	84.09	290.39	37.36
RMB	242	237	146.4	1.50%	184.74	85.24	269.98	43.43
RMA	243	341	232.4	2.40%	127.74	82.41	210.15	51.96
RLB	251	18	21.3	0.20%	227.37	39.14	266.51	51.49
RLA	252	28	24.3	0.20%	107.73	48.17	155.9	25.58
ES3	311	200	101.6	1.00%	404.73	4.4	409.13	20.48
ES2	312	101	57.3	0.60%	302	2.66	304.66	28.16
ES1	313	40	28.4	0.30%	261.94	5.83	267.77	40.09
HE2	411	21	20.8	0.20%	229.87	1.04	230.92	34.3
HE1	412	103	114.6	1.20%	204.61	4.06	208.68	43.29
HD2	421	49	48.1	0.50%	245.51	17.05	262.56	37.82
HD1	422	141	164.4	1.70%	176.33	3.88	180.21	52.97
HC2	431	45	42.2	0.40%	191.4	2.53	193.93	42.85
HC1	432	147	125	1.30%	174.73	7.03	181.76	48.31
HB2	441	40	35.4	0.40%	290.03	6.56	296.59	31.93
HB1	442	123	106.1	1.10%	138.73	6.49	145.23	47.56

RUG-IV Group		Actual N	FULL SAMPLE WITH SAMPLE WEIGHTS APPLIED					
Name	Number		Estimated N	%	Nursing WWST	Therapy WWST	Nursing + Therapy WWST	Nursing WWST CV
LE2	511	48	61.2	0.60%	202.97	0.51	203.48	40.61
LE1	512	206	255.4	2.60%	175.73	4.38	180.11	33.77
LD2	521	75	84.1	0.90%	191.66	4.63	196.29	46.3
LD1	522	255	358.9	3.70%	146.01	1.47	147.48	46.79
LC2	531	69	53.7	0.60%	160.03	3.9	163.93	39.96
LC1	532	202	220.8	2.30%	142.22	3.07	145.28	44.35
LB2	541	19	26.3	0.30%	174.18	0.7	174.88	65.46
LB1	542	85	82.3	0.80%	125.91	7.89	133.8	49.95
CE2	611	33	24.3	0.20%	176.65	1.68	178.33	34.63
CE1	612	71	78.1	0.80%	160.79	1.1	161.89	35.17
CD2	621	74	84	0.90%	191.29	7.73	199.02	30.14
CD1	622	187	183	1.90%	148.93	3.23	152.16	39.72
CC2	631	99	93.4	1.00%	129.84	2.82	132.66	46.75
CC1	632	282	296.9	3.00%	122.06	2.06	124.12	45.02
CB2	641	63	48.7	0.50%	128.05	9.31	137.37	35.29
CB1	642	169	184.5	1.90%	115.61	3.39	118.99	45.99
CA2	651	85	76.2	0.80%	92.35	3.81	96.17	71.75
CA1	652	369	281.3	2.90%	89.97	4.47	94.45	68.86
BB2	711	102	244	2.50%	104.46	1.64	106.09	94.41
BB1	712	528	836	8.60%	103.2	2.19	105.39	66.97
BA2	721	34	47.4	0.50%	93.15	0.46	93.61	92.66
BA1	722	595	729.8	7.50%	71.7	0.85	72.55	78.4
PE2	811	36	63.3	0.60%	161.67	0.89	162.56	49.8
PE1	812	223	247.4	2.50%	160.94	2.43	163.37	44.02
PD2	821	94	179.4	1.80%	153.03	0.76	153.79	46.54
PD1	822	466	554.6	5.70%	145.76	2.51	148.27	43.84
PC2	831	160	307.6	3.20%	109.44	0.35	109.79	59.56
PC1	832	751	818.3	8.40%	120.56	2.76	123.31	44.58
PB2	841	67	90.9	0.90%	113.47	2.4	115.87	58.89
PB1	842	387	396.9	4.10%	83.64	4.52	88.16	56.23
PA2	851	51	73	0.70%	46.72	0.58	47.3	64.35
PA1	852	822	516.3	5.30%	64.39	2.44	66.83	58.5
TOTAL		9706	9751.7	100.00%	134.64	20.91	155.55	63.92

5.2 Adjusted Mean Staff Time by RUG-IV Group

STRIVE used measured staff minutes in two ways: a) for classification, i.e., therapy minutes were used for classifying residents in the rehabilitation and rehabilitation plus extensive RUG-IV groups, and b) as the basis for dependent variables of relative staff resource use (WWST) for both nursing and therapy staff resources. Both uses of staff minutes were employed to evaluate the fit of the RUG-IV system and to estimate the relative resource use (case-mix index) of each RUG-IV group. This section addresses a specific issue regarding the use of the measured staff minutes as a dependent variable.

For the dependent variables representing resource use – nursing and therapy – generally only the computed WWSTs were needed. However, states, providers, and others also may need information on the average number of minutes by each type of staff for residents classified into each RUG-IV group, as they may wish to apply wage rates different than those used in STRIVE (e.g., state-specific wage rates) to calculate a state-specific WWST for each RUG-IV group. To get these numbers, additional computations were needed.

The situations are different for nursing times and for therapy times. For nursing times, in creating the Nursing WWST several adjustments (e.g., truncating the total time of all nursing roles) were made on an aggregate level ("RN", "LPN", and "Aide") that had to be reflected back to the "raw" minutes for each role (e.g., "Restorative Aide"). In creating the therapy WWST, we made overall adjustments for underreporting of PT, OT, and SLP time that had to be reflected back to the "raw" minutes for each role (e.g., "PT Assistant").

In this section, we describe the methodology used to create nursing and therapy time measures.

5.2.1 Nursing Time

Nursing times were subject to two truncations in the data cleaning. First was an upper truncation. Nursing times were summed by nursing job category. These included the following groupings: "RN" (Registered Nurses [RN] and Respiratory Therapists [RT]); "LPN" (Licensed Practical Nurse [LPN] and Licensed Vocational Nurse [LVN]); and "Aide" (Certified Nurse Assistant, Feeding Aide, Bath Aide, etc.). These nursing category sums were each truncated for each observation to a maximum value ("ceiling"), set as the 99th percentile of the nursing category observations (see Section 4.3.3). Second was a lower truncation. For each observation, times were accumulated across all nursing staff (i.e., total nursing including the RN, LPN, and Aide groupings), and truncated to a minimum value ("floor") of at least 10 minutes. As both of these truncations were performed after aggregating staff roles, truncated role-specific measures had to be computed. We wanted the loss due to truncation to be distributed among staff roles within a job category (e.g. RNs and RTs for RN job category) in proportion to each role's contribution to WWST.

The steps to accomplish this, with an example of the computations, are provided in Exhibit 5-1 for the two job roles involved in the RN grouping. As a result of the computations for this hypothetical resident, the raw per diem minutes for RNs were reduced from 160.0 to 109.31 minutes and the raw minutes for RT were reduced from 59.5 to 40.65 minutes. Note that wage weighting the final numbers results in the truncated total Nursing WWST of 369 (=109.31*2.58+40.65*2.14). Note that adjustments were only performed for observations subject to truncation.

Exhibit 5-1: Adjustment of Raw Per Diem Nursing Role Times

	Registered Nurse (RN)	Respiratory Therapist (RT)	Total
Example Data			
Relative Wage Rate (from Table 4-11)	2.58*	2.14*	
Raw Number of Minutes	160.0*	59.5*	
Truncated Total WWST			369*
Calculations			
WWST (raw minutes*rate)	412.8 (=160.0*2.58)	127.33	540.13
Percent of WWST	76.4% (=412.8/540.13)	23.6%	100.0%
Reduction in WWST, due to truncation, to be allocated			171.13
WWST reduction allocation by Percent of WWST	130.79 (=76.4%*171.13)	40.34	
Minutes associated with WWST reduction (using Relative Wage Rate)	50.69 (=130.79/2.58)	18.85	
Adjusted time	109.31 (=160.0-50.69)	40.65	

* Original data – other numbers represent the described computations. Note that some computations have small differences due to rounding and that formulas used are only shown in the RN column, as examples.

We also performed a lower truncation ("floor") on the total of all nursing time across all staff roles; we assumed that at least 10 minutes of nursing staff time per day was provided to everyone. From the STRIVE database (N=9,706), we calculated the overall mean WWST (150.5) and nursing minutes (108.5) *before truncation*. The ratio between these was 1.387; this is the blended overall nursing wage rate. It follows, then, that applying this same ratio to the floor per-diem number of nursing minutes (10) would be associated with a Nursing WWST of 13.87. The computations needed here identify observations with less than 10 minutes of total nursing staff time, and adjust for these

observations the times for individual nursing staff roles so that, when wage-weighted, would result in a Nursing WWST of 13.87. As we did not have any guide of how to allocate these small numbers of minutes across staff roles, we assumed, *a priori,* that each role would get an equal number of minutes. It is easy to calculate that assigning 0.85 minutes to each of the staff roles would, when wage-weighted, result in a WWST of 13.87.

5.2.2 Therapy Time

Raw therapy minutes had to be adjusted for the underreporting of therapy time. These calculations were complex, because we went through a series of steps (see Section 4.3.4) to derive STRIVE therapy times, including adjusting for paper instrument use and capping group time ("25% limit"). We treated the adjustments to the weekly time used for classification as a "black box" that transformed "raw minutes" into "adjusted minutes," and computed for each discipline the ratio between these two. We then applied this ratio to allocated time of each staff role contributing to the original minutes. The resultant values, when wage-weighted, resulted in therapy WWST.

We describe these steps in Exhibit 5-2, which provides the calculations used to adjust the occupational therapy for a example (hypothetical) observation. In this case, the unallocated measured time for all OT staff was 470 minutes, while the adjusted unallocated time was 600 minutes. Applying this ratio to the occupational therapist and the COTA-allocated time results in 344.7 and 127.7 weekly minutes, or 49.2 and 18.2 per diem minutes, respectively. As with nursing time, multiplying each of these adjusted staff minutes by the standardized wage rate for that role provided the therapy WWST measures used in STRIVE. The calculation for other therapy roles was the same.

Exhibit 5-2: Adjustment of Raw Therapy Times

	Unallocated	Allocated
Example Data		
Occupational Therapist (weekly)		
Group time	50	10
Concurrent time	120	60
Individual time (note: same whether or not allocated)	200	200
Total occupational therapist time	370	270
Certified Occupation Therapist Aide (weekly)		
Group time	0	0
Concurrent time	0	0
Individual time (note: same whether or not allocated)	100	100
Total COTA time	100	100
Total All Occupational Therapy time	470	370
Total: Adjusted Occupational Therapy time	600	
Calculations		
Ratio of Unadjusted to Adjusted Occupational Therapy time	1.276595745 (=600/470)	
Apply ratio to Allocated Occupational Therapist Time (weekly)		344.68 (=270*1.28)

Apply ratio to Allocated COTA Time (weekly)		127.66 (=100*1.28)

There was a very small amount of time measured for “therapy aides” and “therapy transport” that was not associated with a specific discipline (i.e., neither PT therapy aides nor OT therapy aides). On average there were 0.76 unallocated, unadjusted raw minutes of therapy aide time per resident and 0.82 minutes for therapy transports. In the absence of other information, we assigned these individuals a wage rate of 1.06 that was the average of the PT therapy aide wage rate (0.99) and OT therapy aide wage weights (1.13). The raw therapy times for these two roles also had to be estimated. We used the same approach as described earlier in this section to adjust other therapy times – the calculation of a ratio between unadjusted and adjusted time (Exhibit 5-2 above shows the calculation for one observation of this ratio for OT time to be 1.28). For each observation, we calculated the PT and OT ratios, as described above and in Exhibit 5-2, and averaged them. This ratio was then applied to the therapy aides and therapy transport times in the same manner as Exhibit 5-2.

Once we computed for each observation the reconciled minutes for each staff role, we were able to accumulate them to a mean within each RUG-IV group. The results are displayed in Table 5-4 for the nursing roles, and in Table 5-5 for the therapy roles. With this information, states or other users can employ other wage weights (e.g., state-specific wage rates) to compute alternative case-mix indexes.

Table 5-4. Mean Adjusted Per Diem Minutes for RUG-IV Groups, by Nursing Staff Role*

RUG-IV 66 Group	Weighted N	RN	Respiratory Therapist	LPN	CNA, GNA, RCT	Certified Medication Aide	Restorative Aide	Bath Aide	Feeding Aide	Psych Aide	Non Certified Care Tech	Clinical Associate	Transportation	Respiratory Therapy Assistant
Wage Weights		2.58	2.14	1.65	1.00	1.00	1.20	0.85	0.85	1.08	0.85	1.00	0.85	1.76
RUX	3.02	26.79	28.35	84.11	86.14	1.01	6.48	0.00	0.00	0.00	1.17	0.00	0.00	0.00
RUL	0.44	45.55	45.54	47.99	155.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RVX	3.68	10.08	1.72	75.27	76.83	0.07	13.94	0.00	0.55	0.00	0.00	0.00	0.00	0.00
RVL	6.31	38.22	8.20	71.18	74.84	1.38	0.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHX	5.69	43.56	53.23	35.50	116.54	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHL	7.66	32.73	12.16	30.74	69.46	2.95	1.19	0.00	0.38	0.00	0.00	0.00	2.89	0.00
RMX	11.60	47.28	23.32	53.84	102.19	0.15	3.02	0.03	0.00	0.00	0.00	0.00	0.00	1.35
RML	10.33	103.14	3.30	57.50	98.98	0.49	0.00	0.00	0.00	0.00	0.63	0.00	0.38	0.00
RUC	43.09	19.15	0.00	44.46	91.67	2.71	0.30	0.00	0.23	0.00	3.27	0.12	0.18	0.00
RUB	36.06	735.95	0.00	48.52	80.75	0.49	2.10	0.00	0.00	0.00	0.39	0.00	0.00	0.00
RUA	20.80	20.37	0.00	32.53	41.19	0.56	2.01	0.08	0.00	0.00	0.35	0.00	0.00	0.00
RVC	66.98	20.61	0.01	45.67	91.58	3.38	3.16	0.00	0.41	0.00	1.52	0.00	0.05	0.00
RVB	78.03	17.18	0.02	34.71	63.66	0.41	1.95	0.00	0.22	0.00	0.08	0.24	0.00	0.00
RVA	113.99	21.37	0.00	34.80	49.49	2.30	1.97	0.01	0.14	0.00	0.74	0.17	0.26	0.00
RHC	88.94	22.20	0.06	34.61	95.66	3.99	1.42	0.03	0.05	0.00	3.49	0.00	0.65	0.06
RHB	85.67	23.21	0.01	29.48	67.93	0.50	2.83	0.04	0.14	0.00	0.37	0.07	0.31	0.00
RHA	207.52	15.45	0.11	30.10	46.32	1.37	0.96	0.01	0.10	0.00	0.09	0.04	0.06	0.00
RMC	95.75	18.51	0.19	35.78	83.05	3.73	6.19	0.02	0.25	0.00	4.83	0.00	0.75	0.00
RMB	146.38	18.20	0.27	33.17	77.66	0.57	1.97	0.03	0.11	0.00	2.03	0.06	0.09	0.00
RMA	232.41	13.17	0.06	27.52	45.63	1.05	0.58	0.01	0.09	0.00	0.66	0.03	0.29	0.00
RLB	21.25	20.96	0.04	24.31	130.57	0.06	1.93	0.00	0.00	0.00	0.27	0.00	0.00	0.00
RLA	24.28	3.38	0.00	21.18	60.34	0.38	2.83	0.00	0.00	0.00	0.01	0.00	0.00	0.00
ES3	101.57	42.88	55.13	39.35	108.65	0.00	0.88	0.00	0.00	0.00	0.00	0.00	0.00	1.00
ES2	57.33	26.71	14.55	58.01	92.20	0.07	2.16	0.00	0.19	0.00	13.41	0.00	0.00	0.00

RUG-IV 66 Group	Weighted N	RN	Respiratory Therapist	LPN	CNA, GNA, RCT	Certified Medication Aide	Restorative Aide	Bath Aide	Feeding Aide	Psych Aide	Non Certified Care Tech	Clinical Associate	Transportation	Respiratory Therapy Assistant
Wage Weights		2.58	2.14	1.65	1.00	1.00	1.20	0.85	0.85	1.08	0.85	1.00	0.85	1.76
ES1	28.37	47.97	0.21	33.22	66.15	7.61	6.15	0.00	0.00	0.00	2.17	0.00	0.00	0.00
HE2	20.79	7.67	0.72	43.72	132.45	0.50	2.29	0.00	0.01	0.00	0.25	0.00	0.76	0.00
HE1	114.60	11.06	0.13	46.51	89.00	0.47	4.26	0.00	0.08	0.00	5.35	0.00	0.00	0.00
HD2	48.07	22.71	0.43	49.60	101.24	0.07	1.10	0.00	0.90	0.00	1.07	0.00	0.01	0.00
HD1	164.37	10.19	0.01	34.83	86.19	0.44	2.10	0.03	0.53	0.00	3.51	0.02	0.00	0.00
HC2	42.16	19.58	0.00	29.54	79.42	0.12	0.27	0.00	0.00	0.00	14.49	0.00	0.00	0.00
HC1	125.00	13.39	0.04	33.15	81.72	0.23	2.61	0.02	0.02	0.00	0.44	0.00	0.00	0.00
HB2	35.37	46.87	0.00	53.42	77.31	0.71	2.35	0.00	0.24	0.00	0.04	0.00	0.00	0.00
HB1	106.08	11.50	0.01	31.12	53.63	1.49	1.98	0.17	0.00	0.05	0.13	0.00	0.00	0.00
LE2	61.18	10.30	0.00	34.83	110.20	0.05	0.99	0.05	0.02	0.00	8.85	0.00	0.00	0.00
LE1	255.44	11.05	0.14	32.76	81.22	1.24	1.98	0.00	0.00	0.01	9.21	0.05	0.24	0.00
LD2	84.11	11.22	0.02	36.62	90.19	0.29	1.70	0.16	0.02	0.00	11.02	0.00	0.35	0.00
LD1	358.88	5.93	0.06	26.46	79.08	0.27	1.34	0.09	0.01	0.00	6.89	0.00	0.08	0.00
LC2	53.72	16.04	0.00	30.36	64.02	1.28	2.70	0.09	0.00	0.00	0.02	0.00	0.00	0.00
LC1	220.76	8.06	0.00	27.44	66.64	1.40	2.34	0.03	0.02	0.00	6.12	0.00	0.13	0.00
LB2	26.33	17.66	0.00	30.62	63.01	6.43	6.39	0.06	1.15	0.00	0.06	0.00	0.00	0.00
LB1	82.30	9.86	0.01	27.76	52.86	0.33	0.80	0.08	0.15	0.00	0.46	0.00	0.00	0.00
CE2	24.28	12.99	0.00	21.12	103.37	0.70	1.55	0.25	0.06	0.00	2.50	0.00	0.01	0.00
CE1	78.09	11.01	0.11	14.75	96.31	3.47	1.71	0.00	0.02	0.00	6.71	0.00	0.25	0.04
CD2	83.98	10.47	0.00	25.46	110.78	0.60	1.42	0.00	0.05	0.00	10.81	0.00	0.00	0.00
CD1	183.02	5.36	0.03	21.67	86.21	3.90	3.06	0.00	0.02	0.00	6.38	0.00	0.12	0.01
CC2	93.39	8.04	0.00	17.53	74.57	0.34	1.01	0.03	0.00	0.00	4.80	0.00	0.00	0.00
CC1	296.86	7.95	0.03	15.03	66.88	2.80	2.35	0.00	0.51	0.00	4.46	0.01	0.00	0.00
CB2	48.74	12.14	0.00	20.25	53.74	0.87	0.55	0.09	0.00	0.00	9.43	0.00	0.00	0.00
CB1	184.54	7.13	0.00	17.51	58.96	2.76	2.58	0.07	0.00	0.00	4.00	0.03	0.06	0.00
CA2	76.23	9.09	0.55	24.04	25.69	0.29	0.39	0.10	0.02	0.02	1.32	0.02	0.47	0.02
CA1	281.34	10.97	0.47	21.70	20.82	0.78	0.71	0.08	0.03	0.02	2.73	0.01	0.01	0.00

RUG-IV 66 Group	Weighted N	RN	Respiratory Therapist	LPN	CNA, GNA, RCT	Certified Medication Aide	Restorative Aide	Bath Aide	Feeding Aide	Psych Aide	Non Certified Care Tech	Clinical Associate	Transportation	Respiratory Therapy Assistant
Wage Weights		2.58	2.14	1.65	1.00	1.00	1.20	0.85	0.85	1.08	0.85	1.00	0.85	1.76
BB2	244.04	4.00	0.00	13.87	69.71	0.54	0.75	0.01	0.00	0.00	0.01	0.00	0.15	0.00
BB1	836.00	5.34	0.02	15.57	57.02	2.80	1.51	0.04	0.02	0.02	2.34	0.01	0.02	0.02
BA2	47.39	5.78	0.00	17.98	46.96	0.56	0.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BA1	729.84	4.82	0.02	13.94	33.05	1.63	0.72	0.07	0.03	0.05	0.61	0.01	0.01	0.01
PE2	63.26	6.06	0.00	18.54	109.26	1.09	3.56	0.00	0.06	0.00	0.00	0.60	0.26	0.00
PE1	247.40	8.37	0.03	17.34	100.34	3.27	1.87	0.07	0.23	0.00	5.09	0.15	0.12	0.01
PD2	179.40	4.49	0.00	17.23	109.66	0.79	1.83	0.10	0.00	0.00	0.03	0.11	0.23	0.00
PD1	554.61	8.26	0.01	15.85	90.93	1.48	2.27	0.02	0.01	0.02	3.62	0.01	0.03	0.01
PC2	307.60	2.58	0.00	14.41	76.54	0.33	1.65	0.08	0.00	0.00	0.03	0.00	0.10	0.00
PC1	818.26	6.07	0.00	18.11	68.48	2.23	1.77	0.01	0.26	0.01	2.29	0.00	0.05	0.00
PB2	90.93	6.25	0.00	18.29	62.47	0.45	3.03	0.00	0.00	0.00	0.06	0.00	0.71	0.00
PB1	396.94	4.43	0.01	13.55	40.82	3.52	2.33	0.04	0.01	0.04	2.94	0.01	0.13	0.01
PA2	73.00	1.62	0.00	14.09	17.69	0.07	1.28	0.00	0.00	0.00	0.04	0.00	0.00	0.00
PA1	516.32	6.23	0.11	13.77	21.74	0.71	1.08	0.11	0.05	0.07	1.40	0.05	0.12	0.05

* RN=Registered Nurse; LPN= Licensed Practical Nurse; CNA= Certified Nursing Assistant; GNA= Geriatric Nurse Assistant; RCT= Resident Care Technician; CMA= Certified Medication Aide.

Table 5-5. Mean Adjusted Per Diem Minutes for RUG-IV Groups, by Therapy Staff Role*

RUG-IV 66 Group	Weighted N	Physical Therapist	PT Assistant	PT Aide	Occupational Therapist	COTA	OT Aide	Speech/ Language Pathologist	Audiologist	Therapy Aide	Therapy Transport
Wage Weight		2.98	1.86	0.99	2.72	1.90	1.13	2.60	2.57	1.06	1.06
RUX	3.02	21.91	17.18	8.44	26.66	6.78	5.40	32.54	0.00	0.00	0.00
RUL	0.44	11.69	21.98	0.00	13.31	19.16	0.00	61.82	0.00	0.00	0.00
RVX	3.68	34.26	5.38	0.03	27.66	10.60	0.12	17.05	0.00	0.00	0.00
RVL	6.31	20.40	13.21	11.26	20.12	8.03	2.23	8.39	0.00	0.00	0.31
RHX	5.69	23.14	3.37	0.00	5.94	13.13	2.37	5.31	0.00	0.00	0.08
RHL	7.66	12.89	11.74	1.66	14.24	13.78	0.00	2.45	0.00	0.00	0.09
RMX	11.60	6.70	6.05	0.33	13.65	4.32	0.00	6.46	0.00	0.00	0.12
RML	10.33	23.61	8.02	1.54	4.53	15.41	0.00	0.22	0.00	0.00	0.00
RUC	43.09	24.30	20.74	7.10	30.29	17.35	1.35	18.68	0.00	0.00	0.19
RUB	36.06	37.27	10.26	1.49	32.51	19.08	0.76	18.25	0.01	0.06	0.05
RUA	20.80	31.92	14.08	1.03	29.30	26.37	0.32	8.61	0.00	0.26	2.63
RVC	66.98	14.32	17.74	3.00	15.05	19.38	0.19	15.40	0.00	0.09	0.87
RVB	78.03	17.30	22.47	1.73	19.75	10.91	1.06	11.38	0.00	0.25	0.02
RVA	113.99	15.69	20.73	2.61	19.54	15.74	0.65	9.52	0.00	0.13	0.06
RHC	88.94	19.75	9.94	1.51	12.42	10.31	0.64	4.09	0.00	0.15	0.01
RHB	85.67	12.82	14.15	0.88	12.98	10.94	0.34	5.30	0.00	0.28	0.11
RHA	207.52	12.90	13.86	1.64	9.50	15.26	0.12	1.99	0.00	0.16	0.11
RMC	95.75	8.88	5.83	0.86	7.12	7.46	1.05	4.02	0.00	0.09	0.59
RMB	146.38	7.94	9.96	0.64	7.30	8.09	0.36	2.48	0.00	0.16	0.10
RMA	232.41	8.57	8.70	0.45	5.77	10.49	0.17	1.62	0.00	0.07	0.09
RLB	21.25	3.03	1.68	0.23	6.04	0.46	0.15	3.55	0.00	0.00	0.02
RLA	24.28	3.72	7.19	0.06	3.10	3.44	0.00	3.34	0.00	0.00	0.00
ES3	101.57	0.60	0.08	0.00	0.17	0.41	0.00	0.47	0.00	0.00	0.00
ES2	57.33	0.05	0.14	0.00	0.17	0.20	0.02	0.53	0.00	0.00	0.01
ES1	28.37	0.70	0.36	0.01	0.94	0.12	0.00	0.09	0.00	0.01	0.00
HE2	20.79	0.12	0.01	0.00	0.00	0.26	0.00	0.06	0.00	0.00	0.00

RUG-IV 66 Group	Weighted N	Physical Therapist	PT Assistant	PT Aide	Occupational Therapist	COTA	OT Aide	Speech/ Language Pathologist	Audiologist	Therapy Aide	Therapy Transport
Wage Weight		2.98	1.86	0.99	2.72	1.90	1.13	2.60	2.57	1.06	1.06
HE1	114.60	0.05	0.06	0.00	0.56	0.27	0.00	0.65	0.00	0.05	0.00
HD2	48.07	0.85	0.12	0.02	1.78	0.05	0.00	3.59	0.00	0.00	0.01
HD1	164.37	0.46	0.22	0.08	0.42	0.17	0.04	0.18	0.00	0.01	0.00
HC2	42.16	0.07	0.81	0.00	0.24	0.03	0.00	0.00	0.00	0.07	0.00
HC1	125.00	0.62	0.88	0.03	0.57	0.81	0.08	0.12	0.00	0.01	0.00
HB2	35.37	1.05	0.98	0.00	0.47	0.01	0.00	0.01	0.00	0.01	0.26
HB1	106.08	0.97	0.89	0.05	0.56	0.16	0.00	0.00	0.00	0.06	0.00
LE2	61.18	0.10	0.02	0.00	0.00	0.05	0.00	0.03	0.00	0.00	0.00
LE1	255.44	0.15	0.07	0.00	0.45	0.96	0.00	0.29	0.00	0.00	0.00
LD2	84.11	0.24	0.26	0.01	1.14	0.11	0.00	0.04	0.00	0.00	0.00
LD1	358.88	0.06	0.13	0.00	0.06	0.18	0.16	0.13	0.00	0.00	0.01
LC2	53.72	0.81	0.57	0.03	0.13	0.02	0.00	0.00	0.00	0.00	0.00
LC1	220.76	0.66	0.17	0.02	0.09	0.15	0.00	0.09	0.00	0.00	0.00
LB2	26.33	0.09	0.06	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.06
LB1	82.30	1.61	0.70	0.00	0.52	0.17	0.00	0.02	0.00	0.00	0.02
CE2	24.28	0.05	0.34	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00
CE1	78.09	0.10	0.00	0.03	0.17	0.06	0.00	0.08	0.00	0.00	0.00
CD2	83.98	0.77	0.15	0.00	1.83	0.02	0.00	0.05	0.00	0.00	0.00
CD1	183.02	0.28	0.07	0.06	0.11	0.43	0.00	0.42	0.00	0.00	0.00
CC2	93.39	0.19	0.52	0.00	0.08	0.36	0.12	0.08	0.01	0.00	0.00
CC1	296.86	0.16	0.30	0.01	0.08	0.18	0.00	0.17	0.00	0.00	0.00
CB2	48.74	0.57	1.01	0.04	1.59	0.42	0.00	0.20	0.00	0.04	0.00
CB1	184.54	0.46	0.38	0.03	0.33	0.12	0.02	0.06	0.00	0.00	0.00
CA2	76.23	0.53	0.32	0.02	0.09	0.72	0.00	0.00	0.00	0.02	0.00
CA1	281.34	0.46	0.18	0.01	0.46	0.74	0.03	0.02	0.00	0.00	0.01
BB2	244.04	0.04	0.10	0.00	0.12	0.30	0.00	0.17	0.00	0.00	0.00
BB1	836.00	0.32	0.18	0.00	0.20	0.01	0.00	0.13	0.00	0.00	0.01
BA2	47.39	0.05	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.00

RUG-IV 66 Group	Weighted N	Physical Therapist	PT Assistant	PT Aide	Occupational Therapist	COTA	OT Aide	Speech/ Language Pathologist	Audiologist	Therapy Aide	Therapy Transport
Wage Weight		2.98	1.86	0.99	2.72	1.90	1.13	2.60	2.57	1.06	1.06
BA1	729.84	0.07	0.13	0.00	0.02	0.01	0.00	0.08	0.03	0.00	0.00
PE2	63.26	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00
PE1	247.40	0.08	0.20	0.00	0.14	0.46	0.00	0.21	0.00	0.00	0.00
PD2	179.40	0.01	0.00	0.00	0.17	0.01	0.00	0.10	0.00	0.00	0.00
PD1	554.61	0.21	0.18	0.01	0.39	0.03	0.01	0.11	0.03	0.04	0.00
PC2	307.60	0.01	0.05	0.00	0.03	0.06	0.00	0.01	0.00	0.00	0.00
PC1	818.26	0.32	0.30	0.03	0.12	0.09	0.00	0.27	0.00	0.00	0.00
PB2	90.93	0.01	0.00	0.00	0.84	0.03	0.00	0.00	0.00	0.00	0.00
PB1	396.94	0.35	0.60	0.04	0.20	0.29	0.02	0.45	0.00	0.00	0.00
PA2	73.00	0.00	0.29	0.02	0.00	0.01	0.00	0.01	0.00	0.00	0.00
PA1	516.32	0.32	0.42	0.03	0.14	0.02	0.00	0.09	0.00	0.00	0.00

***** PT=Physical Therapy, PTA= Physical Therapy Assistant; OT= Occupational Therapist; COTA= Certified Occupational Therapist Assistant; SLP= Speech, Language Pathologist

** Audiologist time was not collected over the entire 7 days and was not adjusted. Over the entire 9706 residents, there was a total of 97 minutes collected during the 48-hour period.

5.3 Changes in classification between RUG-III and RUG-IV

The changes made from RUG-III to RUG-IV, described in Section 4, made substantial alterations in the assignment of a RUG group to a resident. Table 5-6 displays these changes at the level of RUG category for all observations while Table 5-7 displays these same changes, but only within the Medicare sample. In both Tables, weighted numbers of observations are reported (see discussion of weighting in Section 4.1.2).

In large part due to the change in RUG-IV to using services such as IV medications only when provided during the nursing facility stay, the percent of cases in the RE category dropped substantially from the percentage in the equivalent category in RUG-III: from 4.4% in RUG-III to 0.5% in RUG-IV, with most of the observations moving into the associated Rehabilitation category (see Table 5-6). Similarly, the Extensive Services category dropped from 6.9% of all weighted observations in RUG-III to 1.9% in RUG-IV, with those dropped from the RUG-III Extensive category going rather evenly into the H, L, and C categories. On the other end of the spectrum of case mix, there were relatively few changes in the Behavior/Impaired and Reduced Physical Function categories. These findings are mirrored in the statistics for the Medicare population only (see Table 5-7).

Table 5-6. Frequency Comparisons of RUG-III and RUG-IV* Categories, for all Residents (weighted)

RUG-III	RUG-IV									
	RE	R	E	H	L	C	B	P	Total	Percent
Rehabilitation Plus Extensive	41.9	322.8	3.9	29.6	4.8	23.4	0.0	0.4	426.8	4.4%
Rehabilitation	6.8	933.2	0.0	12.4	13.7	47.2	12.2	57.6	1083.2	11.1%
Extensive	0.0	0.7	164.6	185.1	152.2	161.1	0.0	11.2	674.9	6.9%
Special Care	0.0	1.3	17.1	150.9	408.0	163.4	1.3	17.4	759.4	7.8%
Clinically Complex	0.0	0.5	0.6	260.3	354.0	950.9	77.2	175.3	1818.7	18.6%
Impaired Cognition	0.0	0.0	0.8	1.7	3.0	1.6	1454.9	3.1	1465.1	15.0%
Behavior Problems	0.0	0.0	0.0	0.0	0.0	0.0	65.7	0.0	65.7	0.7%
Reduced Physical Function	0.0	2.7	0.3	16.5	207.1	3.0	245.9	2982.6	3457.9	35.5%
Total	48.7	1261.2	187.3	656.4	1142.7	1350.5	1857.3	3247.7	9751.8	100.0%
Percent	0.5%	12.9%	1.9%	6.7%	11.7%	13.8%	19.0%	33.3%	100.0%	

* Version 9

Table 5-7. Frequency Comparisons of RUG-III and RUG-IV Categories, for Medicare Residents (weighted)

RUG-III	RUG-IV									
	RE	R	E	H	L	C	B	P	Total	Percent
Rehabilitation Plus Extensive	29.8	262.8	3.0	20.4	3.8	14.1	0.0	0.4	334.4	24.5%
Rehabilitation	5.8	642.7	0.0	6.9	7.8	28.1	2.7	31.7	725.8	53.2%
Extensive	0.0	0.3	18.0	54.1	21.9	10.8	0.0	3.9	109.1	8.0%
Special Care	0.0	1.1	0.0	21.0	13.4	15.4	0.9	0.9	52.8	3.9%
Clinically Complex	0.0	0.0	0.4	23.2	17.4	40.4	1.2	6.4	88.9	6.5%
Impaired Cognition	0.0	0.0	0.0	0.0	0.0	0.0	22.5	0.0	22.5	1.6%
Behavior Problems	0.0	0.0	0.0	0.0	6.1	0.0	6.3	19.0	31.3	2.3%
Reduced Physical Function	35.6	907.0	21.4	125.7	70.3	108.8	33.6	62.4	1364.8	100.0%
Percent	2.6%	66.5%	1.6%	9.2%	5.2%	8.0%	2.5%	4.6%	100.0%	

It follows from these findings that the percent of all nursing facility residents meeting the presumptive Medicare threshold (viz., assignment to a RUG group Clinically Complex or higher) did not change greatly from RUG-III (48.8%) to RUG-IV (47.7%) (see Table 5-8). This 1.1% decrease was largely due to a decrease in Rehabilitation classifications due to the allocation of concurrent rehabilitation therapy time and a resulting drop to a classification below Clinically Complex.

Table 5-8. Medicare Eligibility under RUG-III and RUG-IV (weighted)

	Eligible RUG-IV Categories (RE through C)	Not Eligible RUG-IV Categories (below C)	Total	Percent
Eligible RUG-III Categories (RE through C)	4410.3	352.7	4763.0	48.8%
Not Eligible RUG-III Categories (below C)	236.5	4752.2	4988.7	51.2%
Total	4646.8	5104.9	9751.7	100.0%
Percent	47.7%	52.3%	100.0%	

5.4 Statistical properties of RUG-IV

Some of the most important statistical properties of RUG-IV are displayed in Table 5-9, which compares these criteria for RUG-IV and RUG-III, both calculated using the sample-weighted STRIVE data.[111]

Table 5-9. Statistical Criteria Achieved by RUG-III and RUG-IV Systems*

Criterion	RUG-III	RUG-IV**
Number of groups	53	66
Variance explanation of Nursing WWST		
Full sample	30.0%	41.5%
Validation sample		41.2%
Variance Explanation of Nursing and Therapy WWST		
Full Sample	53.0%	62.0%
Validation sample		61.0%
Ratio of high/low Nursing WWST	9.1	10.0
Number of Groups less homogeneous than average***	9 of 53	8 of 66

* STRIVE data, N=9706, sample weighted
** RUG-IV, Version 9
*** Lower coefficient of variation in Nursing WWST than overall (63.9)

The full RUG-IV system was substantially superior to the RUG-III system in explaining our measures of resource use, as measured by wage-weighted staff time (WWST). The variance explanation (R^2) when fitting RUG-IV to Nursing WWST was

[111] Note that the full sample size with weighting (9751.7) is slightly different than the actual sample size (9,706) due to rounding.

41.5%, compared to 30.0% for RUG-III[112] (see Table 5-9). If we add in therapy WWST, then the variance explanation increases to 62.0% (compared to 53.0% for RUG-III),[113] although much of this additional variance explanation for the joint measure is the result of using therapy time as part of the classification. The variance explanations with sample weighting were routinely higher than those computed without sample weighting (for example, the Nursing WWST in the total sample had a 41.5% variance explanation when sample weighted and a 38.6% when not sample weighted) (see Table 5-10).

Table 5-10. Variance Explanation of RUG-IV System (sample weighted)

Sample	Nursing WWST	Therapy WWST	Nursing plus Therapy WWST
Sample Weighted			
Total	41.5%	92.8%	62.0%
Derivation	43.2%	93.1%	63.4%
Validation	41.2%	92.9%	61.0%
Unweighted			
Total	38.6%	91.3%	58.2%
Derivation	38.8%	91.3%	57.5%
Validation	39.4%	91.5%	59.6%

With any study such as this, with many analyses and some performed on relatively small subsamples, there is a concern about over-fitting the data. As described earlier, one-third of the STRIVE sample was reserved for validation. When the RUG-IV system was tested independently on the validation sample, virtually the same Nursing WWST variance explanation (41.2%) was obtained (see Table 5-9).

Two other statistical measures of importance were also considered. As a measure of heterogeneity, the Nursing WWST coefficient of variation (CV)[114] of each RUG-IV and RUG-III group was computed and compared to the overall coefficient of variation (63.9) for the sample (all sample weighted) (see Tables 5.3 and 5.11). A good system has many groups with low CV. In RUG-III, 9 out of the 53 groups (17%) had CVs higher than the overall sample; this number decreased to 8 out of 66 (12%) for RUG-IV. The CVs for all RUG-IV groups is presented in Table 5-11. Finally, the ratio of the mean of the most resource intense group and the mean of the least resource intense group provides an indication of how well a case-mix system classifies the rare, high-cost residents. The RUG-III system already achieved a high ratio (9.1 to 1); RUG-IV does even slightly better (10.0 to 1) (see Table 5-9).

[112] Run as a regression of nursing WWST = a+b*RUG, where RUG was the categorical variable denoting the 66 RUG-IV groups or the 53 RUG-III groups, respectively. This model was run with sample weighting.
[113] Run as same model as above, but with nursing plus therapy WWST as the dependent variable.
[114] The coefficient of variation (CV) of any sample is defined as 100 times the standard deviation divided by the mean.

Table 5-11. Nursing WWST Coefficients of Variation (CV) for RUG-IV Groups (sample weighted, full sample)

RUG-IV Group	CV
RUX	21.22
RUL	17.45
RVX	20.36
RVL	41.94
RHX	17.79
RHL	31.47
RMX	21.23
RML	43.42
RLX	Na
RUC	37.22
RUB	48.99
RUA	66.92*
RVC	31.06
RVB	37.27
RVA	42.76
RHC	33.33
RHB	34.61
RHA	46.97
RMC	37.36
RMB	43.43
RMA	51.96
RLB	51.49
RLA	25.58
ES3	20.48
ES2	28.16
ES1	40.09
HE2	34.30
HE1	43.29
HD2	37.82
HD1	52.97
HC2	42.85
HC1	48.31
HB2	31.93
HB1	47.56
LE2	40.61
LE1	33.77
LD2	46.30
LD1	46.79
LC2	39.96
LC1	44.35
LB2	65.46*
LB1	49.95
CE2	34.63
CE1	35.17
CD2	30.14
CD1	39.72

RUG-IV Group	CV
CC2	46.75
CC1	45.02
CB2	35.29
CB1	45.99
CA2	71.75*
CA1	68.86*
BB2	94.41*
BB1	66.97*
BA2	92.66*
BA1	78.40*
PE2	49.80
PE1	44.02
PD2	46.54
PD1	43.84
PC2	59.56
PC1	44.58
PB2	58.89
PB1	56.23
PA2	64.35
PA1	58.50
TOTAL	63.92

***** RUG-IV group CV greater than overall CV (63.92)

5.5 Hospital-based Facilities

The primary aim of this analysis was to compare hospital-based (HB) nursing facilities with other nursing facilities to determine whether:

- staffing is higher, after adjusting for resident resource needs (i.e., nursing CMI)
- RUG-IV distribution is different

For this analysis, we used the full sample (dropping one outlier as was done for the rest of the final analyses, so N=9706) and sample weights.

First, we performed an unadjusted comparison of mean Nursing WWST and raw minutes, as well as by job category, in HB facilities vs. all other facilities. We used a t-test[115] to test for statistical significance, assuming unequal variances. In comparison to all other facilities, on average, HB facilities had significantly higher mean Nursing WWST and raw minutes overall (overall WWST means: 182.4 vs. 132.6) as well as for registered nurses (RN WWST means: 79.6 vs. 24.1), but slightly lower mean WWST for

[115] The t-test is a statistical test of the difference in means (here mean nursing WWST) between two samples, when the number of observations (i.e., facilities) are relatively small (e.g., less than 30 in a group).

licensed practical nurses (LPN WWST means: 27.9 vs. 37.0). There were no significant or substantial differences in aide raw minutes or WWST.

These differences appeared to hold when we adjusted for RUG-IV distribution in the two types of facilities. The statistical significance for CMI-adjusted staffing means was determined based on the significance of the indicator variable in the following two regression models:

Regression model 1: Nursing WWST= a+b*RUG+c*HB, where RUG is a categorical variable of the 66 RUG-IV, Version 8 groups and HB is a dichotomous Indicator (i.e., 0-1) variable equal to 1 for a hospital-based facility.

Regression model 2: Nursing WWST= a+b*CMI+c*HB, where CMI is the single continuous variable representing the average nursing WWST for each RUG-IV, Version 8 group, and HB is as in Model 1.

In both regression models, the indicator variable HB was significant and positive, which supports the hypothesis that HB facilities have higher staffing (i.e., higher Nursing WWST) after adjusting for case-mix (or how resource-intensive the residents are).

RUG-IV distribution differences between HB facilities and other facilities were examined by comparing distributions of RUG-IV categories and by testing the mean Nursing WWST. The chi-square test[116] indicated that the RUG-IV distribution was different between the two types of SNFs (chi-square= 178.5; p-value<0.0001). While the statistical significance may have been primarily the result of the large sample size, there were substantive differences seen in the distributions (see Table 5-12). Relative to all other facilities, HB facilities had much higher proportions of all the categories from Clinically Complex and above (with the exception of Rehabilitation), while other facilities had substantially more individuals in the Rehabilitation and the lowest two categories.

Table 5-12. RUG-IV category distribution: Hospital-based vs. Other (sample weighted)

RUG-IV Category	Hospital-Based	Other
Rehabilitation + Extensive	1.4%	0.5%
Rehabilitation	10.8%	12.8%
Extensive	6.9%	1.7%
Special Care High	8.3%	6.6%
Special Care Low	16.9%	15.1%
Clinically Complex	12.5%	10.4%
Behavior Symptoms and Cognitive Performance	12.7%	19.5%
Reduced Physical Function	30.4%	33.5%
TOTAL	100.0%	100.0%

[116] The chi-squared statistic is used to compare the similarity of two discrete frequency distributions, such as of RUG-IV groups.

When we compared the mean Nursing WWSTs, hospital-based nursing facilities had a substantially and significantly higher mean Nursing WWST (t-test, unequal variances: t-value=5.4; p-value <0.0001) than that of the other nursing facilities.

6 Calculating CMIs

A critical step in determining appropriate case-mix adjusted PPS rates was to use the STRIVE data to develop Case Mix Indices (CMIs) for the each of the 66 RUG-IV groups, and then to standardize these CMIs across the population to which they will be applied. The STRIVE CMIs represent relative indices of the staff costs for the groups and are based upon the WWST staff cost means for the groups. Separate STRIVE CMI sets were calculated for nursing staff cost and for rehabilitation therapy staff cost, allowing separate rate components to be established for nursing and rehabilitation therapy staff costs.

There were two major steps in deriving CMIs for payment:

1) Calculate the STRIVE CMIs for the STRIVE sample. The STRIVE CMI for a RUG-IV group is the survey-weighted WWST for the group in the full STRIVE survey divided by the average survey-weighted WWST across the full STRIVE sample (see Table 5.3). For example, the overall average Nursing WWST is 134.64 and the mean Nursing WWST for the RUX group is 364.14, so the STRIVE CMI for that group is 2.70. To lead to CMIs resulting in appropriate and equitable payment rates, the STRIVE RUG-IV group WWST means must be good estimates of the true staff costs. However, there are practical limitations in obtaining large numbers of residents in the rarer groups in a staff time study. As a result, some group sizes are small and the Nursing WWST means have relatively large margins of error. For example, there are only 94 residents in the nine Rehabilitation plus Extensive Services (RE) category groups (groups RUX through RLX, with group sizes ranging from a low of no residents in the RLX group to a high of 21 for the RMX group). Imprecision in the means for the RE category groups and also for lower groups is evidenced by “inversions” among groups.

2) Calculate payment CMIs for a target population. To set payment rates for a target population (e.g., Medicare Part A residents), the STRIVE CMIs were standardized to that population so that the average CMI for the target population was 1.00. For example, Medicare residents were concentrated in higher STRIVE CMI groups, while Medicaid and other residents were concentrated in low STRIVE CMI groups. As a result, the average Medicare STRIVE CMI was expected to be substantially larger than 1.00 and therefore needed to be “standardized” to this value. In particular, the procedure to calculate the national Medicare Part A CMIs began with the RUG-IV distribution for all Medicare Part A residents in the nation, obtained from payment claims or MDS data. This was applied to the RUG-IV groups’ STRIVE CMIs and the average STRIVE CMI for all Medicare Part A residents was determined. Each group’s STRIVE CMI was then divided by this average to yield a Medicare Part A CMI for each group which, by construction, averaged 1.00 for the Medicare Part A population. To obtain a Medicare Part A staff cost rate component for each group, the group Medicare CMI was multiplied by the average staff cost, obtained from Medicare cost report data. An analogous method could be used to determine CMIs and staff cost rate components for other populations (e.g., Medicaid residents, Depart of Veteran Affairs residents, etc.).

6.1 Inversions and Imprecision in Nursing Time Means

RUG-IV, like its predecessor RUG-III, was developed as a hierarchical classification model that classifies residents using a step-by-step, top-down algorithm. First the characteristics of the most resource intensive residents – provision of both rehabilitation therapy services and extensive care services – are used to identify those to be classified into the top category (RE). The remaining unclassified residents are then considered for classification into the next most resource intensive clinical category (Rehabilitation), based on the provision of rehabilitation therapy services. This process continues classifying the remaining residents sequentially into the subsequent hierarchy categories—Extensive Services (E), Special Care High (H), Special Care Low (L), Clinically Complex (C), and Behavior Symptoms and Cognitive Performance (B). Finally, all remaining residents not already classified into a higher hierarchy category are assigned to the Reduced Physical Function (P) category.

When such a hierarchical classification model is used for payment, rate "inversions" can occur. One type of rate inversion occurs when a resident who qualifies for a higher category also qualifies for a lower category associated with a higher rate. An example of such a hierarchy inversion for RUG-IV is as follows. The total Medicare Part A urban payment rates for the RUG-IV groups are established as the sum of a case mix adjusted nursing staff component, a case mix adjusted therapy staff component, and other fixed (not case mix adjusted) components. The resulting total urban rates were $465.67 for RUG-IV group RHL and $690.61 for RML. With this inversion, a much higher rate would be paid for providing less therapy (medium rehabilitation with group RML) than in the high rehabilitation group RHL, even for residents with the same level of functional dependency (ADL-IV range of 2 – 10). This inversion is due to a Nursing WWST mean of 468.28 for RML that is almost twice as high as the 237.69 value for RHL (see Table 5-3), a difference caused by imprecision in the Nursing WWST means for these two groups.

A second type of inversion involves the ADL-IV Index. The Ultra-High Rehabilitation plus Extensive Services - high ADL (RUX) group has a lower Nursing WWST (364.14, in Table 5-3) than the Ultra-High Rehabilitation plus Extensive Services - low ADL (RUL) group with a value of 449.76. However, a lower ADL (less dependent) resident should be less resource intensive in terms of nursing staff time than the corresponding higher ADL resident. This ADL inversion again occurs because the two groups are relatively small with imprecise estimates of the nursing staff time means (there are only 11 residents in RUX and 2 in RUL).

Other inversions in RUG-IV occurred in groups that were created by tertiary splits of subcategories based on the provision of restorative nursing (the B and P category groups). (Note that in the following, we denote groups of residents defined by a hierarchy category and an ADL-IV range "subcategories"; when subcategories as divided by tertiary splits, they form RUG-IV groups.) As discussed in Section 4.10.1, it was deemed important to have restorative nursing splits (as in RUG-III) to provide an incentive to provide these services in nursing homes. However, the measured resource impact of restorative nursing was modest and splits based on these services resulted in inversions. An example in Table 5-3 is a Nursing WWST value of 109.44 for group PC2

(with restorative nursing) but a value of 120.56 for PC1 (without restorative nursing). It does not make sense that provision of additional nursing services should result in lower nursing staff resource use. We recommend that additional research be performed in this area.

A final set of potential nursing staff rate problems with RUG-IV occurred in groups that were created by splits based on depression (the H, L, and C category groups). These depression splits can create some inversions with lower groups. Further, the impact of depression varied widely across groups, as evidenced by widely varying differences in Nursing WWST between the paired group with depression and without depression. The greatest difference in Table 5-3 was 151.29, seen between the HB2 group with depression (value of 290.03) and HB1 without depression (value of 138.73); the smallest difference was 2.34 for CA2 with depression (value of 92.35) and CA1 without depression (value of 89.97). Such wide variance in the impact of depression again represents imprecision in the estimates.

A solution to all of these precision problems is presented in the next section.

6.2 Mathematical Smoothing of Nursing WWST Means

With RUG-IV, our approach was to develop several different smoothing techniques to minimize inversions and improve precision of the Nursing WWST group means. Methods for smoothing RUG-IV Nursing WWST group means were as follows:

1. **Weighted average**. One pair of R groups with adjacent ADL-IV ranges within the same hierarchy category was smoothed by assigning the weighted average Nursing WWST for the two groups to each group.

2. **Ratio smoothing**. While evaluating the H, L, C, B, and P categories, a few inverted subcategories were identified. These inversions occurred when the mean Nursing WWST of a subcategory was lower than that of the corresponding ADL-IV group in the next lower category. Smoothing of the inverted mean was performed by interpolating a new mean value for the inverted subcategory based on the general ratios between the subcategory above, the subcategory with the inversion, and the subcategory below for ADL-IV ranges where inversions did not occur.

3. **Depression offsets**. The H, L, and C categories have tertiary splits on the basis of depression. The depression effect was much larger for the H and L as compared to the C categories. For that reason, a larger offset was used for the former. The average percent increase in Nursing WWST attributable to depression was determined across all of the Special Care subcategories (after they were split by ADL-IV ranges). The means for a Special Care depression group and the corresponding non-depression group were then “offset” using this percentage, such that the difference between the two groups corresponded to that percentage, while the weighted average of the two groups did not change. A similar, but smaller, depression offset was applied to the C category groups.

4. **Restorative nursing offset**. The B and P categories have tertiary splits on the basis of restorative nursing services. The average Nursing WWST increase attributable to restorative nursing (calculated as a percent of the Nursing WWST for the combined groups) was determined across all of these groups. The means for a restorative nursing group and the corresponding group without restorative nursing were then "offset" using this percentage, such that the difference between the two groups corresponded to that percentage, while the weighted average of the two groups did not change.

5. **Extensive add-on**. The RE and R category groups were smoothed by determining the average Nursing WWST added when a Rehabilitation resident also received extensive services. This average Extensive "add-on" was then added to the Nursing WWST mean of a Rehabilitation group to obtain the smoothed Nursing WWST mean of the corresponding Rehabilitation plus Extensive group.

Each of these RUG-IV smoothing methods is discussed in detail in the order that the methods were applied.

Weighted Average Smoothing

One pair of groups was smoothed with the weighted average method before any other smoothing. The Nursing WWST mean for RUB was 256.76, while that for higher ADL-IV RUC was 220.63 (see Table 5-3). The Nursing WWST means for both RUB and RUC were set to 237.09, the weighted average of the two groups.

Ratio Smoothing

This method was used three times before other smoothing. Table 6-1 illustrates the problems addressed with ratio smoothing.

Table 6-1. Subcategory Nursing WWST Means* for the Lower RUG-IV Categories, before Tertiary Splits (using Depression or Restorative Nursing) and Before Smoothing

	ADL-IV Range				
	A	B	C	D	E
Category	0-1	2-5	6-10	11-14	15-16
Special Care High (H)		176.56	178.93	191.98	208.49
Special Care Low (L)		137.61	145.70	154.67	180.99
Clinically Complex (C)	90.48	118.21	123.92	162.25	164.55
Behavior/Cognition (B)	73.00	103.49			
Physical Function (P)	62.20	89.20	117.52	147.54	161.08

*Note that each entry is the mean for a subcategory defined by a combination of the Hierarchy Category and ADL-IV range. These numbers can be obtained from Table 5-3 by recombining tertiary splits using a weighted average that employs the sample-weighted frequencies. For example, the Table 5-3 PA1 and PA2 values can be used to compute the Table 6-1 PA value of 62.20=73.00*46.72+516.3*64.39. For other entries in Table 6-1, there are small differences from values recalculated from Table 5-3 due to rounding,

since the Table 6-1 entries are based on unrounded values but the Table 5-3 entries are rounded to two decimals.

Table 6-1 shows that there is an inversion for ADL-IV range D, with the mean Nursing WWST for the L subcategory (154.67) lower than that for the C subcategory (162.25). For the other ADL-IV ranges (B, C, and E) there is an orderly decrease from H to L to C. To fix this problem in the D ADL-IV range, and produce a similar orderly decrease from H to L to C, the mean for the L category for the D ADL-IV range was adjusted upward using ratio smoothing. To accomplish this, the ratio of the Nursing WWST difference from L to C to the difference from H to C subcategories was calculated for the B, C, and E ADL-IV ranges. The formula of these ratios is:

(L – C) / (H – C).

For example, for the B ADL-IV range, the ratio was 0.332 (=137.61-118.21)/(176.56-118.21); the values for the C and E ADL-IV ranges were 0.396 and 0.374, respectively. Using the values in Table 5-3, the weighted average of these ratios was calculated across the B, C, and E ADL-IV ranges, weighting each ratio by the sample weighted number of observations involved in the ratio's denominator. The B ADL-IV range ratio was weighted by the sum of observations in the HB2, HB1, CB2, and CB1 groups; the C ADL-IV range ratio by the sum of the observations in the HC2, HC1, CC2, and CC1 groups; and the E ADL-IV range ratio weighted by the sum of observations in the HE2, HE1, CE2, and CE1 groups. The resulting weighted average ratio across the B, C, and E ADL-IV ranges was 0.371, this value being calculated as follows:

0.371= [(0.332*(35.4+106.1+48.7+184.5)+0396*(42.2+125.0+93.4+296.9)

+0.374*(20.8+114.6+24.3+78.1)]

/

(35.4+106.1+48.7+184.5+42.2+125.0+93.4+296.9

+20.8+114.6+24.3+78.1)

Thus, the L subcategory value is located about one-third (viz., 0.371) of the WWST difference from C to H. For the D ADL-IV range, the smoothed WWST value for the L subcategory was therefore estimated as:

C + 0.371 * (H – C).

This yielded a smoothed WWST value of 173.29 (=162.25+0.371*[191.98 -162.25] for the L category and D ADL-IV range.

Two other problems are apparent in Table 6-1. For the C ADL-IV range, the difference is quite small for the C category (123.92) and P category (117.52). For the E ADL-IV range, a small difference is also found for the C category (164.55) and P category (161.08). When offset smoothing methods were applied to split the C groups into subgroups with and without depression, and the P groups into subgroups with and without restorative nursing, inversions resulted. To avoid these inversions, ratio

smoothing (using methods parallel to those above for smoothing the L category group for the D ADL-IV range) was also applied to the C category groups for ADL-IV ranges C and E. For the B and D ADL-IV ranges, the weighted average ratio involving the placement of the C group between the L and P groups was calculated as 0.581. This ratio was taken as a standard to apply to the C and E ADL-IV ranges as follows:

P + 0.581 * (L – P).

This yielded a smoothed WWST value of 133.89 for the C category and the C ADL-IV range, and a value of 172.65 for the C category and the E ADL-IV range.

Depression Offset

In RUG-IV, the H, L, and C subcategories are all split based on the presence or absence of depression. The impact of depression in different subcategories is quite variable as illustrated in Table 6-2. This table presents the percent increase in the Nursing WWST for a final group with depression above the corresponding final group without depression.

Table 6-2. Differences (as Percentages) of Special Care High (H), Special Care Low (L), and Clinically Complex (C) Subcategories, with and without Depression

Category and ADL-IV Range	Depression Percent Increase
HE	12.3%
HD	39.2%
HC	9.5%
HB	109.1%
LE	15.5%
LD	31.3%
LC	12.5%
LB	38.3%
CE	9.9%
CD	14.7%
CC	6.4%
CB	10.8%
CA	2.6%

Examination of Table 6-2 shows that that the effect of depression on Nursing WWST is generally smaller for the C subcategories than the Special Care (H and L) subcategories. Also, there is great variability in the impact of depression, ranging from 9.5% to 109.1% for the H subcategories and 2.6% to 14.7% for the C subcategories. This demonstrates that the precision of the depression estimates within the individual subcategories is not very good, in part due to the small sample sizes involved in several of the computations. To improve precision, the overall impact of depression across

subcategories was determined and then this overall impact was used to establish standard depression differences (offsets) within the individual subcategories.

Given the different impact of depression for the Special subcategories versus the C subcategories, separate depression offsets were established for these two sets of subcategories. Separate resident-level linear regression analyses were performed for the two sets of subcategories. The regression model predicted Nursing WWST from the presence or absence (0, 1) of depression controlling with five indicator (dichotomous) variables representing membership (0, 1) in the different ADL-IV ranges (0-1, 2-5, 6-10, 11-14, and 15-16). The resulting regression coefficients for depression were 47.76 for the Special Care residents and 15.91 for C residents. These coefficients represent the average depression impact on Nursing WWST. To establish a depression percentage offset for the Special subcategories, the coefficient of 47.76 was divided by the average WWST across all Special residents (173.92), yielding a 27.46% offset. To establish a depression percentage offset for the C subcategories, the coefficient of 15.91 was divided by the average WWST across all C residents (124.74), yielding a 12.75% offset.

These percentage offsets were then used to estimate the Nursing WWST for a non-depression subgroup with the following formula:

$WWST_n = M * N / (n_n + (1 + p_d) * n_d)$

where:

$WWST_n$ = Nursing WWST for the RUG-IV group without depression

M = weighted average WWST for the combined depression and non-depression groups before splitting

N = total number of observations for the combined depression and non-depression groups before splitting (= $n_d + n_n$)

n_n = number of observations in the non-depression subgroup

n_d = number of observations in the depression subgroup

p_d = appropriate depression percentage offset

The estimate of Nursing WWST for a depression subgroup is obtained with the following formula:

$WWST_d = WWST_n * (1 + p)$

where:

$WWST_d$ = Nursing WWST for the RUG-IV group with depression

With these formulas, the offset smoothing methodology resulted in WWST values for a pair of depression and non-depression final groups that meet two needed criteria: a) they differ by the selected percentage offset value, and b) the weighted mean across

the two subgroups will be the same as the overall group prior to splitting. The smoothed WWST values based on the depression offset are presented in Table 6-5 in the next section (6.3).

Restorative Nursing Offset

In RUG-IV, the B and P subcategories all have tertiary splits based on the presence or absence of restorative nursing services. Similar to the depression tertiary splits for the H, L, and C subcategories, the impact of restorative nursing in different groups is quite variable as illustrated in Table 6-3. This table presents the percentage increase in the Nursing WWST for a subgroup with restorative nursing above the corresponding subgroup without these services.

Table 6-3. Differences (as Percentages) of Behavior Symptoms and Cognitive Performance (B) and Reduced Physical Function (P) Subcategories, with and without Restorative Nursing Services

Category and ADL-IV Range	Restorative Nursing Percent Increase
BB	1.2%
BA	29.9%
PE	0.5%
PD	33.2%
PC	-9.2%
PB	35.7%
PA	-27.4%

The impact of restorative nursing on Nursing WWST was positive in 3 subcategories (the impact was 29.9% for BA, 33.2% for PD, and 35.7% for PB), minimal in 2 cases (1.2% for BB and 0.5% for PE), and negative for 2 subcategories (-9.2% for PC and -27.4% for PA). While the impact of restorative nursing was quite variable, we decided to include tertiary splits in B and P categories to provide an incentive for their provision, as in RUG-III.

To achieve a consistent, modest increase for restorative nursing, the WWST means for the B and P subcategories were smoothed using a restorative nursing offset. A resident-level regression analysis was performed to determine the overall impact of restorative nursing across all B and P subcategories. The regression model predicted Nursing WWST from the presence or absence (0, 1) of restorative nursing controlling with indicator (dichotomous) variables indicating membership (0, 1) in the Behavior/Cognitive category and the Physical Function category. The resulting regression coefficient for restorative nursing was 8.71. To establish a restorative nursing percentage offset for the B and P subcategories, the coefficient of 8.71 was divided by the average WWST across all Behavior/Cognition and Physical Function residents (105.65), yielding a 8.24% offset.

This percent offset was then used (in the same manner as with the depression splits) to estimate the Nursing WWST for a subgroup without restorative nursing, using the following formula:

$WWST_n = M * N / (n_n + (1 + p_r) * n_r)$

where:

$WWST_n$ = Nursing WWST for the RUG-IV group without restorative nursing

M = weighted average WWST for the combined restorative and non-restorative groups before splitting

N = total number of observations for the combined restorative and non-restorative groups before splitting (= $n_n + n_r$)

n_n = number of observations in the non-restorative subgroup

n_r = number of observations in the restorative subgroup

p_r = appropriate restorative nursing percentage offset

The estimate of Nursing WWST for a restorative nursing group is obtained with the following formula:

$WWST_r = WWST_n * (1 + p_r)$

where:

$WWST_r$ = Nursing WWST for the RUG-IV group with restorative nursing

With these formulas, the offset smoothing methodology resulted in WWST values for a pair of final groups with and without restorative nursing that meet two needed criteria: a) they differ by the selected percentage offset value, and b) the weighted mean across the final groups will be the same as the overall subcategory prior to splitting. The smoothed WWST values based on the restorative nursing offset are presented in Table 6-5 in the next section (6.3).

Additional Weighted Average Smoothing

After the depression and restorative nursing offsets were applied, the Nursing WWST means for all H, L, C, B, and P groups were examined for inversions.

Relatively small inversions were found for two pairs of groups. The first inversion involved the CD1 group with a mean Nursing WWST of 156.00 and the PD2 group with a mean of 156.54. The Nursing WWST means for both CD1 and PD2 were set to 156.27, the weighted average of the two groups. The second inversion involved the CE1 group with a mean Nursing WWST of 167.58 and the PE2 group with a mean of 171.49. The Nursing WWST means for both CE1 and PE2 were set to 169.33, the weighted average of the two groups. After these adjustments, there were no remaining Nursing WWST inversions in these categories.

Extensive Add-On

Two large inversions in Nursing WWST occurred among the Rehabilitation plus Extensive Services (RE) category groups. First, group RVX has a WWST mean of 247.72 (3.7 residents[117]) but RHX has a much higher mean of 401.33 (5.7 residents) and RMX has a higher mean of 368.88 (11.6 residents) (see Table 5-3). Second, group RHL has a mean of 237.69 (7.7 residents) and Group RML has a mean of 468.28 (10.3 residents). Other small inversion can be found among the RE category groups. All of these inversions are the result of imprecise estimates due to the small number of observations in the groups.

In general, when the addition of extensive services criterion moves a resident from an R group to an equivalent RE group, the difference in the mean Nursing WWSTs is substantial, but both are highly variable and do not follow any systematic pattern (for example, the difference is not related to the Nursing WWST of the Rehabilitation-only group or the level of rehabilitation). This is illustrated in Table 6-4, which presents the average percentage by which the Nursing WWST for a RE group exceeded that of the corresponding Rehabilitation group (with the Rehabilitation groups combined and restricted to specific ADL-IV ranges to allow comparability to the RE groups – see Figure 4-17).

Table 6-4. Increase in Nursing WWST for Rehabilitation Plus Extensive (RE) Groups over ADL-IV Equivalent Rehabilitation (R) Groups

R Group(s) (with Restricted ADL-IV Range for Comparability)	Equivalent RE Group	Percent Increase in Nursing WWST for Equivalent RE Subcategory
RUC	RUX	53.6%
RUA (ADL-IV=2-10) + RUB	RUL	103.1%
RVC	RVX	8.3%
RVA (ADL-IV=2-10)+RVB	RVL	88.1%
RHC	RMX	82.9%
RHA (ADL-IV=2-10)+RHB	RML	49.8%
RMC	RMX	78.8%
RMA (ADL-IV=2-10)+RMB	RML	183.7%
RLA (ADL-IV= 2-16)+RLB	RLX	n/a[118]

The increase in Nursing WWST associated with the addition of Extensive Services ranges from 8.3% to 183.7%. Again, these findings are the result of imprecise estimates due to the small number of observations in the groups.

[117] Resident counts are not whole numbers due to survey weighting.
[118] No value can be determined because there were no low rehabilitation residents also receiving extensive services.

To achieve a consistent, more stable increase for extensive services and to prevent inversions, we used a smoothing method similar that employed for depression and restorative nursing. However, the "percentage offset" method employed for depression and restorative nursing could not be used for extensive services, because it requires corresponding pairs of groups, with and without extensive services, and there are no low rehabilitation residents with extensive services.

Thus, an alternative "add-on" smoothing method was used here. A resident-level regression analysis was performed to determine the average increase in Nursing WWST for a RE resident versus an R resident with the same level of rehabilitation and the same ADL-IV range. The regression model predicted Nursing WWST from the presence or absence (0, 1) of extensive services, with control indicator (0, 1) variables indicating membership in the nine Rehabilitation Level and ADL-IV range combinations shown in the left column in Table 6-4.

The resulting regression coefficient for extensive services was 168.16. This is the average impact on Nursing WWST of adding extensive services for a resident receiving rehabilitation. The WWST mean for residents without extensive services was calculated for each of the nine Rehabilitation and ADL-IV range combinations (see Table 6-4) and then 168.16 was added to that mean to estimate the mean for the corresponding RE group.

For the Rehabilitation plus Extensive Services groups with the higher ADL-IV range (11-16), the estimate was quite simple. For example, the mean for Very High Rehabilitation and ADL-IV 11-16 without extensive services (the RVC group) was 228.84 (see Table 5-3). The estimated mean for the Very High Rehabilitation plus Extensive Services with ADL-IV 11-16 (RVX) group was 228.84 + 168.16 for a total of 397.00 (see Table 6-5). A similar calculation was used for throughout the RE category for group codes ending in "X" except RLX.

For the RLX group, the estimate was slightly more complicated, as it corresponds to two Rehabilitation groups. A weighted mean was calculated across all RLB residents and those RLA residents with an ADL-IV range of 2 – 5. The resulting weighted mean of 174.98 was based on Low Rehabilitation with an ADL-IV range of 2-16 (the ADL-IV range for RLX).[119] The estimated mean for the RLX group was 174.98 + 168.16 for a total of 343.14.

As an example of the calculation for Rehabilitation plus Extensive Services groups ending in "L" (ADL-IV range 2-5), consider RVL. A weighted mean was calculated across all RVB residents and those RVA residents with an ADL-IV range of 2-5. The resulting weighted mean of 165.06 was based on Very High Rehabilitation with an ADL-IV range of 2-10 (the ADL-IV range for RVL). The estimated mean for the RVL group was 165.06 + 168.16 for a total of 333.22. A similar calculation was used for all Rehabilitation plus Extensive Services groups with group code ending in "L".

---

[119] When apply either an offset or add-on for a pair groups with and without a characteristic, it is appropriate that the pair of groups have a common ADL range.

The smoothed WWST values based on the extensive services add-on are presented in Table 6-5 in the next section (6.3).

6.3 Smoothed Means and CMIs for Nursing Time

The final smoothed Nursing WWST mean for each of the 66 RUG-IV groups is presented in Table 6-5. STRIVE full sample nursing CMIs are also present for each group. The STRIVE nursing CMI for a group is the Nursing WWST mean for the group divided by the weighted average Nursing WWST for the full STRIVE sample. This full sample weighted average is 135.98 (see the last row of Table 6-5). The STRIVE nursing CMIs are relative indices with a weighted average of 1.00 for the STRIVE sample. Since the CMIs are based upon group means with the survey weights applied, these values are estimates for the national population of all nursing home residents.

Table 6-5. Smoothed Nursing WWST and STRIVE Nursing CMI, by RUG-IV Group for the Full STRIVE Sample

RUG-IV Group		Number of Observations		Survey Weighted Nursing WWST	STRIVE Nursing CMI
Name	Number	Actual N	Survey Weighted N		
RUX	111	11	3.0	405.25	2.98
RUL	112	2	0.4	389.61	2.87
RVX	121	7	3.7	397.00	2.92
RVL	122	11	6.3	333.22	2.45
RHX	131	11	5.7	387.63	2.85
RHL	132	19	7.7	326.84	2.40
RMX	141	21	11.6	374.45	2.75
RML	142	12	10.3	333.20	2.45
RLX	151	0	0	343.14	2.52
RUC	211	49	43.1	237.09	1.74
RUB	212	48	36.1	237.09	1.74
RUA	213	25	20.8	150.65	1.11
RVC	221	107	67.0	228.84	1.68
RVB	222	133	78.0	168.47	1.24
RVA	223	141	114.0	167.78	1.23
RHC	231	119	88.9	219.47	1.61
RHB	232	197	85.7	181.10	1.33
RHA	233	277	207.5	138.77	1.02
RMC	241	175	95.7	206.29	1.52
RMB	242	237	146.4	184.74	1.36
RMA	243	341	232.4	127.74	0.94
RLB	251	18	21.3	227.37	1.67
RLA	252	28	24.3	107.72	0.79
ES3	311	200	101.6	404.73	2.98
ES2	312	101	57.3	302.00	2.22
ES1	313	40	28.4	261.94	1.93

RUG-IV Group		Number of Observations		Survey Weighted Nursing WWST	STRIVE Nursing CMI
Name	Number	Actual N	Survey Weighted N		
HE2	411	21	20.8	254.99	1.88
HE1	412	103	114.6	200.05	1.47
HD2	421	49	48.1	230.39	1.69
HD1	422	141	164.4	180.75	1.33
HC2	431	45	42.2	213.30	1.57
HC1	432	147	125.0	167.34	1.23
HB2	441	40	35.4	210.59	1.55
HB1	442	123	106.1	165.22	1.22
LE2	511	48	61.2	219.07	1.61
LE1	512	206	255.4	171.87	1.26
LD2	521	75	84.1	209.93	1.54
LD1	522	255	358.9	164.70	1.21
LC2	531	69	53.7	176.24	1.30
LC1	532	202	220.8	138.27	1.02
LB2	541	19	26.3	164.45	1.21
LB1	542	85	82.3	129.02	0.95
CE2	611	33	24.3	188.95	1.39
CE1	612	71	78.1	169.33	1.25
CD2	621	74	84.0	175.89	1.29
CD1	622	187	183.0	156.27	1.15
CC2	631	99	93.4	146.49	1.08
CC1	632	282	296.9	129.92	0.96
CB2	641	63	48.7	129.82	0.95
CB1	642	169	184.5	115.14	0.85
CA2	651	85	76.2	99.32	0.73
CA1	652	369	281.3	88.09	0.65
BB2	711	102	244.0	109.97	0.81
BB1	712	528	836.0	101.59	0.75
BA2	721	34	47.4	78.63	0.58
BA1	722	595	729.8	72.64	0.53
PE2	811	36	63.3	169.33	1.25
PE1	812	223	247.4	158.43	1.17
PD2	821	94	179.4	156.27	1.15
PD1	822	466	554.6	144.62	1.06
PC2	831	160	307.6	124.40	0.91
PC1	832	751	818.3	114.93	0.85
PB2	841	67	90.9	95.09	0.70
PB1	842	387	396.9	87.85	0.65
PA2	851	51	73.0	66.65	0.49
PA1	852	822	516.3	61.57	0.45
TOTAL		9706	9751.7		
Weighted Average				135.98	

6.4 Rehabilitation Category Means and CMIs for Therapy Time

There were no inversions or precision problems with the WWST rehabilitation therapy WWST means used for CMI calculation and determination of a rehabilitation therapy staff time rate component. This is because the CMIs were established for the five levels of rehabilitation (Ultra-High, Very-High, High, Medium, and Low) rather than for the individual RUG-IV groups. The Ultra-High level combines the groups RUX, RUL, RUC, RUB, and RUA; the Very-High level combines the groups RVX, RVL, RVC, RVB, and RVA; etc. The STRIVE therapy CMIs were computed in the same manner as the STRIVE nursing CMIs: the mean Therapy WWST was divided by the overall sample weighted average therapy WWST (137.45). The weighted numbers of cases, therapy WWST means, and STRIVE therapy CMIs are given in Table 6-6.

It can be seen in Table 6-6 that four of the five therapy levels had at least 100 residents. The Low Rehabilitation level involved only 45 residents. In general, the rehabilitation category WWST means should have good precision. There were no hierarchical inversions in the WWST means and the means aligned as expected.

Table 6-6. Therapy WWSTs and CMIs by Level of Rehabilitation Therapy form all Rehabilitation Plus Extensive Residents, Survey Weighted*

Rehab Level	RUG-IV Groups	Weighted Number of Observations	Mean Therapy WWST	Therapy CMI
Ultra-High	RUX, RUL, RUC, RUB, RUA	103.4	290.11	2.11
Very-High	RVX, RVL, RVC, RVB, RVA	269.0	198.80	1.45
High	RHX, RHL, RHC, RHB, RHA	395.5	132.64	0.97
Medium	RMX, RML, RMC, RMB, RMA	496.5	84.81	0.62
Low	RLX, RLB, RLA	45.5	43.95	0.32
TOTAL		1,309.9		
Weighted Average			137.45	

* Analysis performed on the RE and R categories in the full STRIVE sample (N=9,706).

6.5 Medicare Part A Rate Inversions after Smoothing

CMS calculated RUG-IV SNF Medicare Part A urban rates using the same methodology used to set rates every year and documented in the annual SNF rule in the Federal Register. When the calculated rates were analyzed, the following rate inversions were found:

- RHL, RML, RLX versus ES3.

These inversions will have little impact because RHL, RML, and RHX are rare groups in Part A (1.1% of the survey weighted Part A residents), and many residents in these groups will not qualify for ES3.

- RHA with ADL-IV range of 2 – 5 versus HB2.

This inversion will have moderate impact since RHA comprises 12.4% of the survey weighted Part A residents. However the impact will be more limited since many of the RHA residents will not have an ADL-IV range of 2 – 5 or will not qualify for HB2.

- RMA with ADL-IV range of 2 – 5 versus HB2, HB1, LB2, LB1, and CB2.

These inversions will have moderate impact since RMA comprises 11.0% of the survey weighted Part A residents. However the impact will be more limited since many of the RMA residents will not have an ADL-IV range of 2 – 5 or will not qualify for the lower groups listed.

- RLA with ADL-IV range of 2 – 5 versus HB2, HB1, LB2, LB1, and CB2.

These inversions will have little impact because RLA is a rare group (0.4% of the survey weighted Part A residents), and many RLA residents will not have an ADL-IV range of 2 – 5 or will not qualify for the lower groups listed.

- RLA with ADL-IV range of 6 – 10 versus HC2, HC1, LC2, LC1, CC2, and PC2.

These inversions will have little impact because RLA is a rare group (0.4% of the survey weighted Part A residents), and many RLA residents will not have an ADL-IV range of 6 – 10 or will not qualify for the lower groups listed.

- RLB with ADL-IV range of 15 – 16 versus HE2.

This inversion will have little impact because RLB is a rare group (0.3% of the survey weighted Part A residents), and many RLB residents will not have an ADL-IV range of 15 – 16 or will not qualify for HE2.

In the SNF Medicare Part A payment system, these inversions will be handled by index maximizing RUG-IV classification. With index maximizing, if a resident qualifies for more than one group, then classification is made to the group with the highest payment rate. In other words, a resident is assigned the highest possible payment rate. Index maximizing for the rate inversions will increase Medicare Part A payments. However, the overall fiscal impact of all of these inversions on Medicare SNF Part A payments has been estimated as an increase well below 1%.

7 Other Studies

7.1 Cost of drugs

One of the initial goals of STRIVE was to collect, simultaneously with information about the staffing cost of care, information that would permit estimating the daily cost of prescription drugs. This would permit investigation into predictors of the cost of drugs and possible incorporation of drug cost predictors into RUG-IV. As described in the following, problems obtaining drug information seriously undermined the usefulness of this analysis. We describe in this section the data collected and the analyses that were performed.

7.1.1 Collection and Coding of Drug information

We envisioned three possible sources of information on the prescription drugs used by STRIVE nursing facility residents. First, we asked facilities if they could provide us with a full listing of each sampled resident's drug profile. Most facilities indicated that these data were not easily accessible by them, but suggested that they could be obtained from their local pharmacies. Therefore, we abandoned this first approach and CMS worked with a few of the major pharmaceutical suppliers to develop a System of Records that would permit these data to be provided to the project. Late in the data collection, the Department of Health and Human Services ruled that provision of these data would violate personal privacy regulations. The facilities participating in STRIVE were asked to provide Medication Administration Records (MARs) for Medicare residents in the study. As STM collection usually started on a Monday and ended the following Sunday, we asked for MARs covering the calendar month(s) in which the time study occurred.

Fifty-six facilities in 14 states plus the District of Columbia were able to provide MARs printed forms including data on a total of 428 Medicare residents. These forms were hand-entered into Microsoft ACCESS (Version 2002) data bases. After data cleaning and deleting duplicate entries, there were a total of 7237 drug entries for the 428 residents, or 16.9 drugs per resident. Drugs include all prescription medications, but also other medications or materials, such as vitamins, minerals, and TB tests.

From the MARS data for each medication, we extracted the drug National Drug Code (NDC), drug the drug name, including strength and, if appropriate, volume, the amount given at each administration, drug strength unit, volume unit, quantity, frequency, route of administration, whether provided as needed (PRN), on what days during the study week the drug was provided, the dates (if any during the period) when the drug was started or stopped. We also noted monthly medications on the record but not administered during the study week.

The primary source of information about individual drug cost was obtained from the December 2007 version of First Data Bank, which links the NDC with drug name,

dose, and cost measures; however, the NDC was often missing on the STRIVE facility MARs. A project pharmacist provided the Generic Sequence Number (GSN) for each drug in the MARs data; this number was used instead of NDCs to link with First Data Bank. When a GSN could not be determined, the strength or quantity was missing, or the drug was a duplicate, the drug was dropped (30 MARs drug entries).

The First Data Bank can include several GSN entries for each drug name. Each entry includes 3 average wholesale price (AWP) variables: AWP, AWP Brand Name, and AWP Generic. Using SAS statistical software (Version 9.1), the lowest price among the 3 average wholesale prices, ignoring missing prices, was determined. The lowest price across all GSN entries for each drug name was selected and attached to each the 7,237 MARs drugs. The drug name includes the strength and volume, which helps define the cost for each drug. For example, the cost for drug name "LANOXICAPS 0.05 MG CAPSULE" is $0.27 per pill.

We then sought a measure of the daily cost of drugs for a resident. Of the 7,237 drug entries, 92.2% were administered daily or weekly. This also includes drugs given part of a week or PRN. Only 0.1% of all drugs were provided once a month, 2.8% were given once or twice a month, 1.4% were defined as sliding scale (e.g., the amount of insulin injected is based on the results of a blood test), 0.6% were drugs without a matched price, and 2.9% had missing drug frequency. Due to problems defining daily dose for non daily/weekly drugs, only daily and weekly drugs were used to define each resident's daily drug cost; still, we were able to develop a daily cost covering almost all drugs listed in the MARS.

The average number of times each drug was administered daily during the week (number of administrations divided by 7) was also calculated from the MARS, and then multiplied by the average daily drug quantity and by the lowest AWP to get the average daily cost for each drug. Finally, the average daily cost across all drugs for each resident was summed: this is the total daily drug cost for each resident. Two residents had drugs with a very high outlier drug costs: $1,230.44 for a Copaxone 20 MG Injection Kit and $133,089.62 for Calcitonin Powder. These two daily costs were defined as outliers and excluded from analysis.

7.1.2 Relationship of RUG-IV to Drug Costs

Of the 428 residents with MARs data, 395 had all the required information for analysis: MDS data without missing data for any of the items used in RUG-IV classification, full STM (48 hours) data, and non-outlier daily drug cost[120]. For daily and weekly drugs, the 395 residents had a mean prescription count of 15.7 (median 15 and standard deviation of 6.6). The maximum number of drugs was 40 with a minimum of 1. The mean daily drug cost was $34.16, (median $14.00 and standard deviation $56.45).

[120] Of the remaining 33, 30 did not have a full STM, 1 was missing data need to classify in RUG-IV, and 2 were the drug cost outliers discussed earlier.

The most expensive daily cost (after dropping the two outliers described earlier) was $421.97 and the least was less than a penny.

Given the relatively small sample size, the test of whether RUG-IV[121] would explain drug costs was based on the prediction of drug cost by a CMI based on the mean WWST for each RUG-IV group[122]. The CMI was calculated as the mean WWST for the group divided by the overall mean WWST. Two predictions of drug cost were made, one for CMI based on nursing WWST and the other for CMI based on nursing plus therapy WWST. The WWST values used were from RUG-IV Version 6.

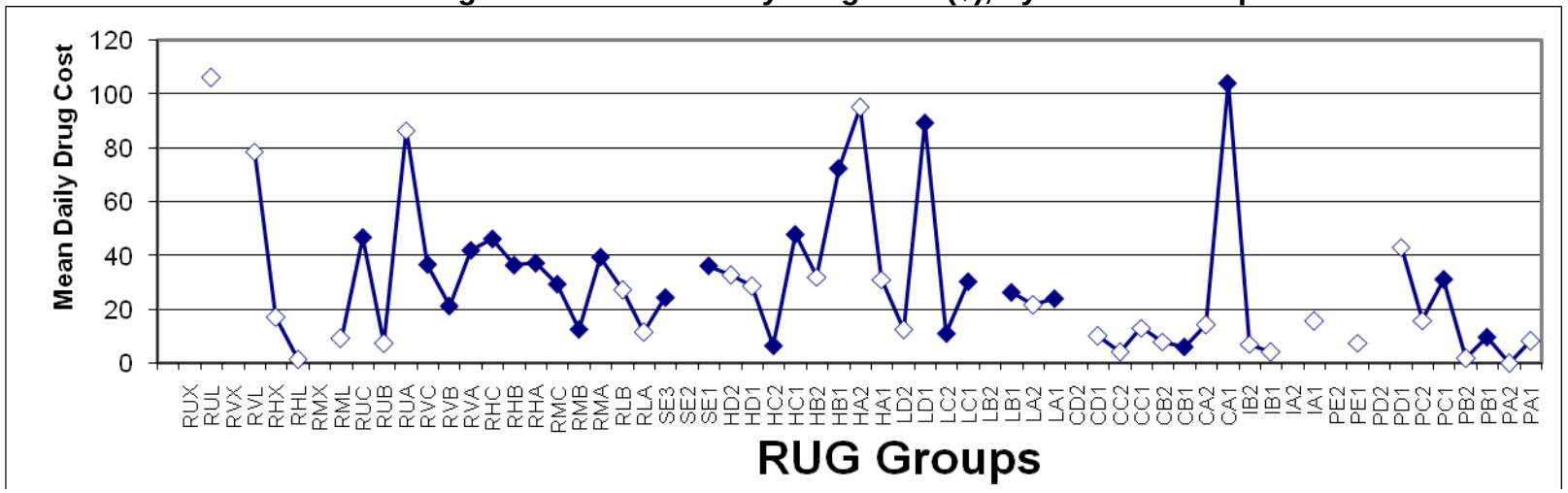
Daily drug cost for the 395 residents with drug data was explained neither by group CMI based on nursing WWST nor nursing plus therapy WWST, with $R^2 = 0.01\%$, $p = 0.88$ and $R^2 = 0.00\%$, $p = 0.93$, respectively). In addition, when analyzed in the same way, daily drug cost was explained neither by the CPS nor the ADL-IV index for the 395 residents, $R^2 = 0.11\%$, $p = 0.52$ for CMI based on nursing WWST and $R^2 = 0.25\%$, $p = 0.32$ for CMI based on nursing plus therapy WWST.

[121] RUG-IV, Version 6 was tested.

[122] Run as a regression of Drug Cost = a+b*CMI, where CMI was the mean WWST for each of the 66 RUG-IV groups divided by overall mean WWST. The model was run twice: once each for nursing WWST and nursing plus therapy WWST.

Figure 7-1 shows the mean daily drug cost by the RUG-IV Version 6 groups for the 395 residents with daily drug cost available. There is no pattern to indicate that particular categories or set of groups have high drug costs.

Figure 7-1. Mean Daily Drug Cost ($), by RUG-IV Group



Open symbols represent less than 5 residents.

We also looked at results for particularly expensive drugs. It was the cost of particular expensive drugs rather than multiplicity of drugs that drives the total daily cost. Figure 7-2 shows the relationship of the highest daily drug cost with the total daily drug cost. High daily cost is closely associated with the most expensive drugs. Figure 7-3 displays the relationship between the count of daily and weekly drugs by daily drug cost, in which we see little relationship: high daily cost is not associated with the number of drugs.

Figure 7-2. Highest Daily Drug Cost ($) by Total Daily Drug Cost

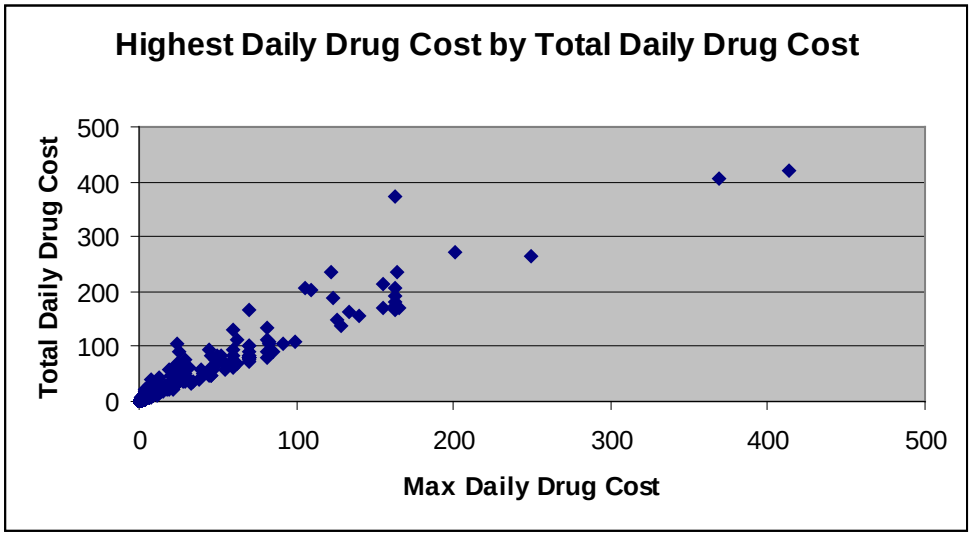
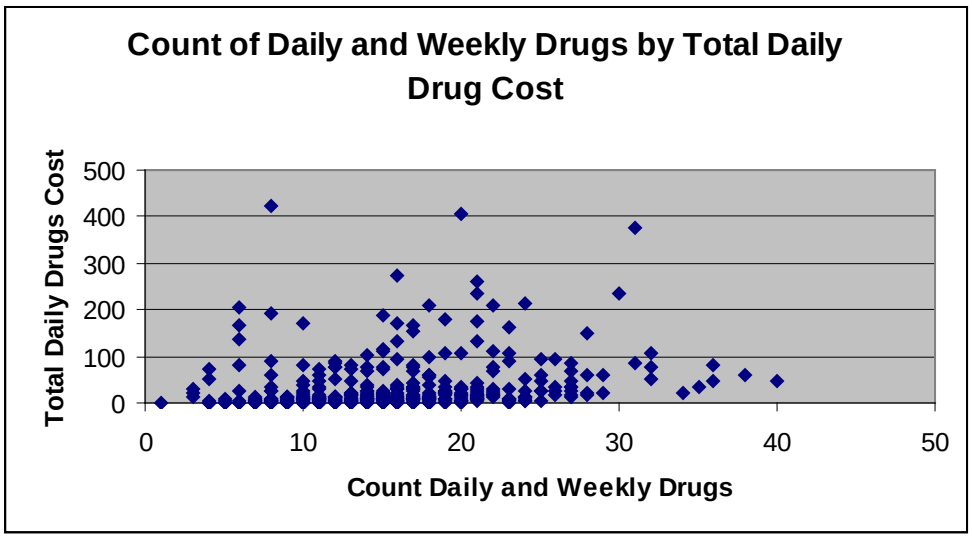


Figure 7-3. Total Daily Drug Cost ($) by Count of Daily and Weekly Drugs



7.2 Inter-Rater Reliability

Several assessment items used in the STRIVE Addendum (see Appendix B) were new and untested. As described in the Phase I Report, pg. 16 and its Appendix A-17, Addendum items were from the MDS 3.0 development effort or interRAI instruments with additional items specifically created to address RUG-III problems (such as post-admission use of services). To assure that any items potentially usable for RUG-IV had appropriate reliability, we tested any items for which there had been no prior reliability test. We were also asked by CMS to test some new items being considered for MDS 3.0, such as Addendum Item XQ1: Resident expresses interest in returning to community. We thus describe here an inter-rater reliability test performed on new, untested Addendum items.

7.2.1 Background

There are a variety of measures that can inform the reliability of assessment items. For the purposes here, the most appropriate approach was agreed to be inter-rater reliability (IRR). In IRR, a single resident is assessed twice, independently, by two assessors, within a relatively short period of time (so that the assessed characteristics of the resident can be expected not to have changed). It is critical to the test that the two assessments be performed completely independently, without communication between the assessors. Inter-rater reliability is measured by the congruity of these two assessments. It should be noted that formal validity tests of these items were not performed, these being beyond the scope of this project. If any of these measures are eventually found to be predictors of case mix, this will be a strong measure of their validity. In addition, if an unreliable item is measuring a concept that is truly related to case mix, it is likely that the unreliability will cause it to be found as an insignificant case-mix predictor. As discussed previously, some of the new items tested for reliability were not expected to be of interest for measuring case mix.

In addition to IRR, we examined the distributional properties of new variables, to assure that responses were obtained across the full spectrum of the responses.

We report here only on inter-rater reliability of the selected items. Other item characteristics, discussed in Section 3.3, were considered in determining whether an assessment item could be considered useful for case mix measurement.

7.2.2 Methods

Data collection

IRR Data collection occurred as part of the standard data collection by facilities, beginning in August 2006. A special IRR assessment was completed and linked with the normal STRIVE assessment for inter-rater reliability purposes. A small number of IRR assessments, performed by facility staff, were completed in one of the two pilot facilities, during May testing, and these assessments were included in the IRR analysis.

As part of the regular data collection, skilled nursing facility staff (usually the MDS Coordinator) performed the regular STRIVE MDS Addendum on all study residents in the facility with a reference date of the Thursday of the Staff Time Measurement. These assessments were usually performed during the week following the STM.

IRR assessments were performed by either a facility staff ("staff") or project staff ("project") member. When project assessors were licensed within that state and had state authority to perform the assessment, they perform the IRR; however, they were not always available, so the IRR design incorporated both staff-staff and staff-project regular/IRR assessment pairs.

For staff-staff IRR assessment pairs, a facility assessor was an RN or LPN who completed the normal assessment (STRIVE MDS and STRIVE MDS Addendum) and a second staff assessor completed an IRR assessment of the same resident using the STRIVE MDS Addendum only. These assessment pairs were performed the week after the time study but all had the reference date of the Thursday of the STM week.

For staff-project IRR assessments, the state licensed project staff member performed the IRR assessment (Addendum only) on Thursday or Friday of the staff time measurement week, and a facility staff assessor completed the normal assessment (STRIVE MDS and STRIVE MDS Addendum) of these selected residents

Whenever possible, and especially for the measurement of pressure ulcers, the IRR assessment was performed within 24 hours of the normal assessment.

Given the additional burden on facility staff, IRR assessments were done on only a subsample of the residents on selected units. Every tenth resident was chosen, by selecting residents for whom the last digit of their STRIVE number was a "1." As residents with Stage 2 through 4 pressure ulcers occur relatively rarely and such conditions were a special IRR focus, whenever possible, two additional time study residents with these conditions were selected for inclusion in the IRR study. During the time study, facility staff provided a list of active stage 2-4 pressure ulcer residents on the unit selected for special IRR study. If there were more than two available, those with the lowest last digit in their STRIVE number were chosen.

Early detection phase

As we wished to identify as soon as possible any significant problems in item reliability, an Early Detection phase was implemented. At least five IRR assessments were performed in each of the first six facilities involved in the data collection. We monitored the reliability for new items after accumulating a sample of 30 pairs of assessments so that early "mid-course" corrections could be made, if necessary.

In this Early Detection phase, in addition to IRR testing, we also sought the opinion of the Data Collection Monitors about the Addendum items, including the following:

- Informal face validity: did assessors feel that this assessment item measures something of value, and measures it appropriately?

- Clarity: duplication with other items, such as others already collected on the MDS 2.0
- Burden: complexity of obtaining the item value (with burden rated on a 1 to 5 scale), and whether it was especially difficult for particular types of residents

This feedback was collected once, after the monitors completed 5 assessments in the first six facilities. When appropriate, this information was used to improve the instructions for individual items.

Documentation phase

After the Early Detection phase, five IRR assessments were performed for time study residents in the first three facilities in each state (we skipped the first facility in each state if the IRR assessments were being performed by project data monitors who were new to the process).

The IRR assessment was performed by the project data monitor in states where the data monitor was licensed (i.e., Texas, Ohio). Otherwise, this IRR assessment was performed by a second facility staff assessor.

Up to two additional time study residents in the facility were selected for IRR because of the presence of a stage 2-4 pressure ulcer, as described earlier. Other than these changes, the procedures were the same as in the "early detection" phase.

An IRR sample size of 150 was expected to be obtained from the five "at random" residents in each facility. This was achieved for most items. These numbers were well above the minimum of 103 determined by statistical power analysis. Items where a sample size of 150 was not achieved were items that had multiple versions (e.g. pain items), or were extremely rare (e.g. recent admissions).

Approximately 103 "stage 2-4 pressure ulcer" residents were anticipated. However, this was not achieved. There were approximately 82 residents with stage 2-4 pressure ulcers and completed pairs of assessments.

IRR Addendum forms were sent to IFMC for keying. They were paired with the STRIVE MDS Addendum for the same individual and an IRR data set was developed for analysis. Analysis was done on items with adequate sample sizes for all IRR STRIVE Addendum items.

Data cleaning

Out-of-range values were re-coded to missing. Furthermore, only pairs of values both with non-missing values were analyzed for each item. For this reason, and due to the multiple item versions, the number of pairs analyzed for each item varies. For example, some items (e.g., distance walked) had many more paired assessments than items only included in a later version of the Addendum (e.g., facility pain indicator items). However, for most items, there were greater than 150 pairs available for analysis.

Although there were multiple versions of items, a few changed items were considered similar enough to treat as the same item. The primary instance of this was a change in the look back period for the pain items from 5 days in the first version of the Addendum to 3 days for all other versions.

Statistical methods

Inter-rater reliability was evaluated by w the Kappa statistics. With these pairs of assessments, the evaluation of item reliability traditionally applies the Kappa statistic for dichotomous response items. For items with more than 2 responses, the weighted Kappa is preferred, using Fleiss-Cohen weights[123]. The Kappa statistic is a measure of the congruity of two measurements – here two independent assessments of the same item – but improves upon simple agreement measures by accounting for the distribution of the responses. Kappa statistics above 0.4 are considered minimally acceptable and those above 0.7 are considered good[124].

In addition, the distributions of the values were examined, since uneven distributions can skew the Kappa statistic. The Kappa statistic is considered inappropriate to be computed on very skewed distributions, for example, when one response occurs over 90% of the time or more.

Some variables were on a continuous scale (e.g. pressure ulcer measurements) where a Pearson Correlation Coefficient was used to measure congruity, and the relationship was also investigated graphically.

7.2.3 Results

A sample size of approximately 30 pairs was produced for the Early Detection phase by December 2006. The reliabilities were acceptable for most items (Kappa >0.40), but were low for the majority of the pain items (<0.30). Additionally, there were some very low prevalence items (e.g. oral/nutritional status items), which could not be evaluated. Based on the Addendum Feedback forms, the wording for the special treatments (Section XP of the Addendum) was unclear and also resulted in assessors not indicating post-admission special treatments and procedures for *all* residents. The wording of these items was revised.

Table 7-1 displays the final Inter-rater Reliability Results, based on a minimum of 19 (and up to 235) paired assessments (again, note that not all pairs of could be used for all items). Items in Table 7-1 are listed in the order they appear in the Addendum (see Appendix B). Some of the important columns in this table are as follows:

Kappa > 0.4: The first column indicates whether the Kappa value is acceptable. An acceptable Kappa is a value greater than 0.40, and is indicated with a “y” in this column.

---

[123] Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley and Sons, 1981: 218.
[124] Fleiss JL. 1981. *Op cit.*

An “r” in this column indicates that Kappa was not the best indication of reliability, due to the nature of the item. In these cases, a correlation coefficient (r) was used to evaluate reliability (e.g., pressure ulcer measurements).

Name: The form designation for the STRIVE Addendum item (e.g., “xg1” for the “Distance walked” item).

Label: A brief verbal description of the item.

Kappa: The Kappa value. A shaded entry indicates that the Kappa is not interpretable because the distribution of the values were unbalanced (frequency >95%), which makes the Kappa unstable. If the entry in the “Kappa > 0.4” column is an “r”, then this column contains a correlation coefficient rather than a Kappa.

N res: The number of residents who had non-missing and valid values on the item for the pair of assessments.

Min: The minimum item value on either of the two assessments.

Max: The maximum item value on either of the two assessments.

Coding Frequency: The distribution of responses on this item across both of the two assessments.

Notes: Brief comments related to inter-rater reliability.

IRR on versions: Lists the addendum versions included in the Kappa calculation.

The results demonstrated which items were reliable.

Physical Function and Structural Problems: Inter-rater reliability was acceptable for most of the items in this section (distance walked, distance wheeled, knee replacement, etc.). However, the number of prior independent ADLs (item xg3b) did not achieve good inter-rater reliability. This may be due to the slightly uneven distribution of the values, where approximately 86% of the values were “9” (no fracture/replacement).

Disease Diagnoses: In the subsample where IRR was performed, none of the assessors indicated that a resident had Post-Traumatic Stress Disorder. As a result, the Kappa statistic could not be calculated. For coding of sleep apnea, there was perfect agreement though there were also very few codings of sleep apnea and the Kappa was unstable (shaded).

Health Conditions (Pain): Most of the items in this section were not reliable. Only a few items exceeded the minimum acceptable Kappa threshold of 0.40: whether the resident was on a scheduled pain medication regimen, whether the resident was comfortable most of the time during the past week, whether the resident was able report pain, and whether the resident was in any pain for the past 3 days. Despite the change in versions, the pain items continued to have low inter-rater reliability.

Oral/Nutritional Status: Assessors were reliable when coding whether residents showed coughing or choking during meals, as well as coding for the primary mode of

nutritional intake. However, all of the items in this section had extremely uneven distributions (>90% of values were 0). As a result, the Kappas are unstable and are not good indicators of the reliability of these items.

Skin Condition: Assessments of pressure ulcer status had good inter-rater reliability (Kappa=0.52). Most of the items that asked assessors to indicate the reason(s) why an ulcer was the most problematic were not reliable. This is a very subjective question, because nurses may find an ulcer to be problematic for various reasons.

Overall, paired pressure ulcer measurements were somewhat well correlated (>0.5). However, paired measurements of stage 3 ulcer width (correlation of 0.21) and stage 2 ulcer lengths (0.39) were not very well-correlated. Also, there were many fewer "valid" measurements of pressure ulcer depth, relative to pressure ulcer length and width. This is due to several assessors indicating that they were unable to measure depth because of wound condition or the presence of a wound dressing. A paired t-test of the measurements indicated no statistically significant differences between assessors' measurements of pressure ulcers. Figures 7-4 through 7-8 show a majority of the pressure ulcer measurements were near the line of "perfect" agreement. However, there are some marked differences in a few measurements, which cannot be explained by difference in time between measurements. Since the pressure ulcer measurement items asked the assessors to identify the largest stage 2, and/or largest stage 3 or 4 pressure ulcer, the agreement of the measurements depended upon the assessors' choice of the "largest" pressure ulcer. As a result, it is not certain that the two assessors were always measuring the same pressure ulcer.

This issue – the uncertainty that the pair of measurements were performed evaluating the same pressure ulcer – was realized during data collection when it was too late to revise the Addendum items and procedures. Thus, an additional analysis was performed using the pressure ulcer count items on the MDS 2.0 (Items M1: Number of ulcers of any type at stages 1 thru 4). We looked at the correlation between pressure ulcer measurements for those residents where the MDS indicated only 1 ulcer. Limiting consideration to residents where it is known that both assessors measured the same ulcer resulted in higher correlation coefficients. The correlations increased more when consideration was further limited to paired measurements made within 7 days of each other, with correlation coefficients varying from 0.674 to 0.926 (see Figures 7-9 through 7-12). Note in each of Figures 7-9 through 7-12, that is one resident with set of paired measurements where both values are 0. This indicates that the single ulcer for this resident was not a Stage 2 or higher pressure ulcer (was another type ulcer or a Stage 1 pressure ulcer) and the Addendum measurements could not be made.

Special Treatments and Procedures: For all but three of these items, the response distributions were extremely uneven with the prevalence of the treatment or procedure being less than 5%. These items do not met the Kappa criterion of no value with more than 95% of the responses and Kappa is not a good measure of reliability. The three items with acceptable distribution for Kappa all showed good reliability. These three items were IV medication prior to admission, oxygen therapy prior to

admission, and oxygen therapy post admission, with Kappa values of 0.77, 0.74, and 0.56, respectively.

Discharge Potential and Overall Status: The four items in this section that had acceptable reliability were interest in discharge to community (Kappa of 0.59), prefer discharge to own home (0.60), prefer discharge to relative/friend home (.44), discuss return to community (0.60). Of the remaining items, two had distributions precluding the use of Kappa (single value with more than 95% of responses), and the others had low reliability (Kappa < 0.40).

Table 7-1. Inter-Rater Reliability for New Addendum Items

Kappa >0.4	Use for case mix	Variable		Kappa [1]	N res	Min	Max	Coding Frequency (%)									Notes	IRR on versions
		Name	Label					0	1	2	3	4	5	6	7	9		
Sec XG: Physical Function and Structural Problems																		
y	y	xg1	Distance walked	0.83	189	0	3	42.3	20.9	12.7	24.1							Versions 042706 to 102506
y	y	xg2	Distance wheeled	0.47	179	0	5	27.4	36.9	3.6	10.1	7.5	14.5				0.57 for wheelchair use (n=107); 0.31 for distance wheeled (n=43)	Versions 060906 to 102506
y	y	xg3aa	Knee replacement	0.53	194	0	2	56.4	0.5	43.0								Versions 042706 to 102506
y	y	xg3ab	Hip replacement	0.52	193	0	2	56.0	0.8	43.3								
y	y	xg3ac	Hip fracture	0.54	194	0	2	53.6	3.6	42.8								
	y	xg3b	Prior independent ADLs	0.38	132	0	9	6.1	0.4	1.1	0.4	6.1				86.0		
Sec XI: Disease Diagnoses																		
	y	xi1a	Post-Traumatic Stress Disorder	--	171	0	1	100.0									All ratings were 0	Versions 090506 to 102506
y	y	xi1b	Sleep apnea	1.00	197	0	1	98.0	2.0									Versions 042706 to 102506
Sec XJ: Health Conditions																		
y		xj3a	Sched pain medication	0.42	99	0	1	62.1	37.9									Version 020107
		xj3b	Received PRN pain med	0.21	96	0	1	43.2	56.8									
		xj3c	Received non-med intervention	0.28	97	0	1	71.6	28.4									
		xj4	Report pain (with option of "unclear")	0.31	98	0	2	8.7	87.8	3.6								
		xj5a	Comfortable	0.27	96	0	9	15.1	75.5							9.4		
		xj5b	In pain/hurting	0.21	97	0	9	64.4	26.3							9.3		
		xj5c	How bad is pain	0.21	84	0	9	63.1	18.5	12.5	5.4	0.6					9 (unable to answer) was not used	
		xj5d	Pain make unable to sleep	0.33	96	0	9	50.0	13.0	26.6						10.4		
		xj5e	Limited activities	0.25	96	0	9	49.5	14.1	25.5						10.9		
		xj5f	Done enough to make comfortable	0.04	96	0	9	10.9	48.4	28.1						12.5		

Shading indicates rating prevalence >90%

1: If number of rating categories = 2 then a simple kappa is reported. If categories >2 AND are ordered, then the weighted kappa is reported (using Fleiss-Cohen weights)

Kappa >0.4	Use for case mix	Variable		Kappa 1	N res	Min	Max	Coding Frequency (%)									Notes	IRR on versions
		Name	Label					0	1	2	3	4	5	6	7	9		
y		xj5g	Comfortable most of the time	0.52	95	0	9	8.4	78.9							12.6		
		x6a1	Crying, whining. In response to movement	0.18	99	0	1	85.9	14.1									
		xj6a2	Facial expressions	0.22	98	0	1	70.4	29.6									
		xj6a3	Protective body movements	0.37	99	0	1	81.3	18.7									
		xj6a4	Vocal complaints	0.25	99	0	1	67.2	32.8									
		xj6a5	Reduced interest	0.08	99	0	1	92.9	7.1									
		xj6b	Treatment to relieve pain behavior	0.11	99	0	9	32.3	4.5	28.8						34.3		
		xj6c	Care plan addressing pain prevention or relief	0.07	85	1	4	34.7	58.2	0.6	6.5							
y		xj3	Report pain (no/yes)	0.50	196	0	1	20.4	79.6									Versions 042706 to 102506
y		xj4a	Pain	0.45	182	0	2	39.8	49.2	11.0								Versions 060906 to 102506
		xj4b	Pain Most/ All Time	0.29	182	0	2	66.8	20.6	12.6								Versions 060906 to 102506
		xj4c	Worst Pain Intensity	0.30	180	0	5	36.7	11.7	26.4	6.4	3.9	15.0				0.40 if restrict to ratings <5 (unable to answer)	Versions 060906 to 102506
		xj4d	Pain treatment working	0.26	179	0	2	21.8	62.6	15.6								Versions 060906 to 102506
		xj5a	Non-verbal pain sounds	0.27	188	0	1	82.4	17.6									Versions 042706 to 102506
		xj5b	Painful expressions	0.23	187	0	1	68.4	31.6									
		xj5c	Bracing	0.13	185	0	1	78.4	21.6									
		xj5d	Complaints of pain	0.30	184	0	1	63.3	36.7									
		xj6a	Pain medication	0.40	197	0	1	35.8	64.2									
		xj6b	Pain non-med	0.07	193	0	1	67.6	32.4									
Sec XK: Oral Nutritional Status																		
y	y	xk1a_v3 (xk1c)	Swallowing: Coughing	0.45	167	0	1	90.4	9.6								1 rater used only 1 value	Versions 060906 to 102506

Shading indicates rating prevalence >90%

1: If number of rating categories = 2 then a simple kappa is reported. If categories >2 AND are ordered, then the weighted kappa is reported (using Fleiss-Cohen weights)

Kappa >0.4	Use for case mix	Variable		Kappa [1]	N res	Min	Max	Coding Frequency (%)									Notes	IRR on versions
		Name	Label					0	1	2	3	4	5	6	7	9		
	y	xk1b	Swallowing Food: Change in Voice	0.14	166	0	1	96.4	3.6								1 rater used only 1 value	Versions 090506 to 102506
	y	xk1c_v3	Swallowing Food: Difficulty eating	0.27	180	0	1	90.3	9.7									Versions 090506 to 102506
	y	xk1d	Swallowing: Complaints	0.25	180	0	1	92.5	7.5									Versions 060906 to 102506
y	y	xk2	Primary Nutritional intake	0.75	170	0	3	91.1	0.6	0.3	7.9							Versions 090506 to 102506
	y	xk3a	# Parenteral feeding	0.21	176	0	7	96.0	0.6	0.3					3.1		correlation coefficient	Versions 042706 to 102506
Sec XM: Skin Condition																		
y		xm1a	Pressure Ulcer Status	0.52	202	0	4	57.4	6.2	9.2	16.6	10.6						Versions 090506 to 020107
y		xm1b1	Ulcer Problematic - Largest	0.44	203	0	1	73.2	29.9									Versions 090506 to 020107
		xm1b2	Ulcer Problematic -- Stage	0.35	203	0	1	76.1	23.9									Versions 090506 to 020107
		xm1b3	Ulcer Problematic - Treatment	0.21	203	0	1	81.8	18.2									Versions 090506 to 020107
y		xm1b4	Ulcer Problematic - Pressure	0.58	202	0	1	70.3	29.7									Versions 090506 to 020107
		xm1b5	Ulcer Problematic - Location	0.35	203	0	1	74.1	25.9									Versions 090506 to 020107
		xm1b6	Ulcer Problematic - Other	0.08	203	0	1	92.9	7.1									Versions 090506 to 020107
		xm1b7	Ulcer Problematic - Reason	--													write-in variable	Versions 090506 to 020107
r		xm2a	Ulcer Length - Stage 3 Or 4	0.78	34	0.4	11										correlation between nonzero measurements	Versions 090506 to 020107
r		xm2b	Ulcer Width - Stage 3 or 4	0.21	34	0.1	20										correlation between nonzero measurements= .69 (n=33; 1 outlier removed)	Versions 090506 to 020107

Shading indicates rating prevalence >90%

1: If number of rating categories = 2 then a simple kappa is reported. If categories >2 AND are ordered, then the weighted kappa is reported (using Fleiss-Cohen weights)

Kappa >0.4	Use for case mix	Variable		Kappa 1	N res	Min	Max	Coding Frequency (%)									Notes	IRR on versions
		Name	Label					0	1	2	3	4	5	6	7	9		
r		xm2c	Ulcer Depth - Stage 3 or 4	0.54	19	0.1	3.8										correlation between nonzero measurements	Versions 090506 to 020107
r		xm3a	Largest Ulcer Length -Stage 2	0.39	48	0.2	15										correlation between nonzero measurements	Versions 090506 to 020107
r		xm3b	Largest Ulcer Width -Stage 2	0.65	48	0.1	8.7										correlation between nonzero measurements	Versions 090506 to 020107
r		xm4	# Unhealed Stage 2 Ulcers	0.29	203	0	7	80.1	13.3	5.4	0.7			0.3	0.3		correlation	Versions 090506 to 020107
Sec XP: Special Treatments & Procedures																		
y	y	xp1ab	Barium Swallows -Post Admit	0.66	182	0	1	99.2	0.8									Combined: v042706-090506 (admitted in last 7 days) AND v 102506 (all)
y	y	xp1bb	BIPAP/CPAP - Post Admit	0.80	183	0	1	98.6	1.4									
	y	xp1cb	Hyperbaric Oxygen -Post Admit	--	183	0	1	100.0									All ratings were 0	
y	y	xp1db	Infection Isolation -Post Admit	0.56	183	0	1	98.1	1.9									
y	y	xp1ea	IV Meds- Prior Admit	0.77	52	0	1	85.6	14.4									Versions 042706 to 102506
y	y	xp1eb	IV Med - Post Admit	0.65	182	0	1	95.1	4.9									Combined: v042706-090506 (admitted in last 7 days) AND v 102506 (all)
y	y	xp1fb	Wound Therapy -Post Admit	0.49	181	0	1	98.9	1.1									
y	y	xp1ga	Oxygen Therapy- Prior Admit	0.74	51	0	1	87.3	12.7									Versions 042706 to 102506
y	y	xp1gb	Oxygen Therapy -Post Admit	0.56	180	0	1	83.3	16.7									Combined: v042706-090506 (admitted in last 7 days) AND v 102506 (all)
	y	xp1hb	PET Scan -Post Admit	--	181	0	1	99.7	0.3								1 rater used only 1 value	
	y	xp1ia	Suctioning - Prior Admit	--	51	0	1	98.0	2.0								1 rater used only 1 value	Versions 042706 to 102506

Shading indicates rating prevalence >90%

1: If number of rating categories = 2 then a simple kappa is reported. If categories >2 AND are ordered, then the weighted kappa is reported (using Fleiss-Cohen weights)

Kappa >0.4	Use for case mix	Variable		Kappa 1	N res	Min	Max	Coding Frequency (%)									Notes	IRR on versions
		Name	Label					0	1	2	3	4	5	6	7	9		
y	y	xp1ib	Suctioning -Post Admit	0.91	182	0	1	97.0	3.0									Combined: v042706-090506 (admitted in last 7 days) AND v 102506 (all)
y	y	xp1ja	Tracheostomy Care - Prior Admit	1.00	52	0	1	98.1	1.9									Versions 042706 to 102506
y	y	xp1jb	Tracheostomy Care -Post Admit	0.91	182	0	1	97.0	3.0									Combined: v042706-090506 (admitted in last 7 days) AND v 102506 (all)
	y	xp1ka	Transfusion- Prior Admit	--	52	0	1	98.1	1.9								1 rater used only 1 value	Versions 042706 to 102506
	y	xp1kb	Transfusion - Post Admit	--	182	0	1	100.0									All ratings were 0	Combined: v042706-090506 (admitted in last 7 days) AND v 102506 (all)
	y	xp1la	Ventilator- Prior Admit	--	53	0	1	99.1	0.9								1 rater used only 1 value	Versions 042706 to 102506
y	y	xp1lb	Vent/Respirator - Post Admit	0.80	181	0	1	98.6	1.4									Combined: v042706-090506 (admitted in last 7 days) AND v 102506 (all)
Sec XQ: Discharge Potential & Overall Status																		
y		xq1a	Discharge: Interest in return to community	0.59	170	0	1	64.1	35.9									Versions 090506 to 102506
y		xq1b1	Discharge preferred location: home	0.60	168	0	1	78.3	21.7									Versions 090506 to 102506

Shading indicates rating prevalence >90%

1: If number of rating categories = 2 then a simple kappa is reported. If categories >2 AND are ordered, then the weighted kappa is reported (using Fleiss-Cohen weights)

Kappa >0.4	Use for case mix	Variable		Kappa [1]	N res	Min	Max	Coding Frequency (%)									Notes	IRR on versions
		Name	Label					0	1	2	3	4	5	6	7	9		
y		xq1b2	Discharge preferred location: Rel/Friend	0.44	163	0	1	83.7	16.3									Versions 090506 to 102506
		xq1b3	Discharge preferred location: Grp Home	0.27	162	0	1	89.2	10.8									Versions 090506 to 102506
		xq1b4	Discharge preferred location: Other	-0.05	156	0	1	95.5	4.5									Versions 090506 to 102506
y		xq1d (xq1c_v4)	Discharge: Discuss return to community	0.60	171	0	1	73.4	26.6									Versions 090506 to 102506
		xq1e (xq1d_v4)	Discharge: Permission	0.26	170	0	1	85.6	14.4									Versions 090506 to 102506
		xq1e1	Discharge Info: Resident	0.24	170	0	1	45.3	54.7									Versions 090506 to 102506
		xq1e2	Discharge Info: Rel/Friend	-0.02	169	0	1	84.4	15.6									Versions 090506 to 102506
		xq1e3	Discharge Info: Legal Guardian	--	168	0	1	98.2	1.8								1 rater used only 1 value	Versions 090506 to 102506
		xq1e4	Discharge Info: Other	0.25	161	0	1	70.8	29.2									Versions 090506 to 102506

Shading indicates rating prevalence >90%

1: If number of rating categories = 2 then a simple kappa is reported. If categories >2 AND are ordered, then the weighted kappa is reported (using Fleiss-Cohen weights)

Figure 7-4. Stage 3/4 Pressure ulcer length measured by the two assessors

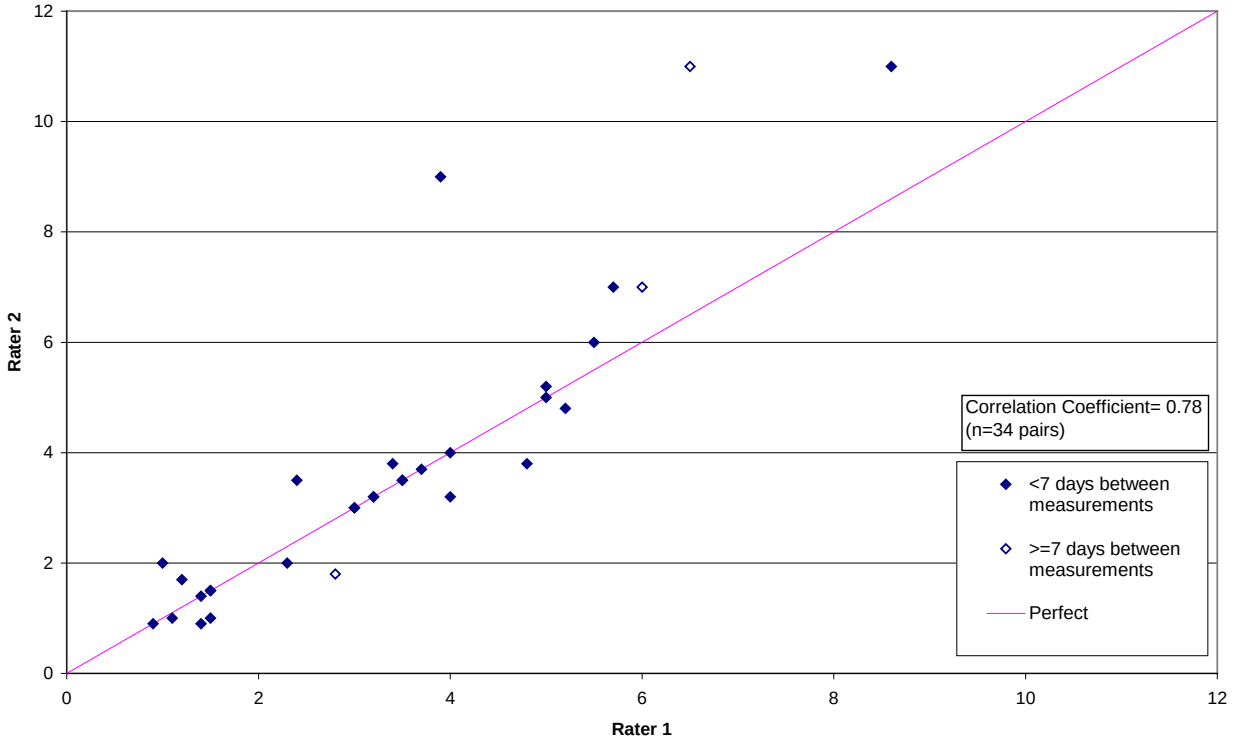


Figure 7-5. Stage 3/ 4 Pressure ulcer width measured by the two assessors

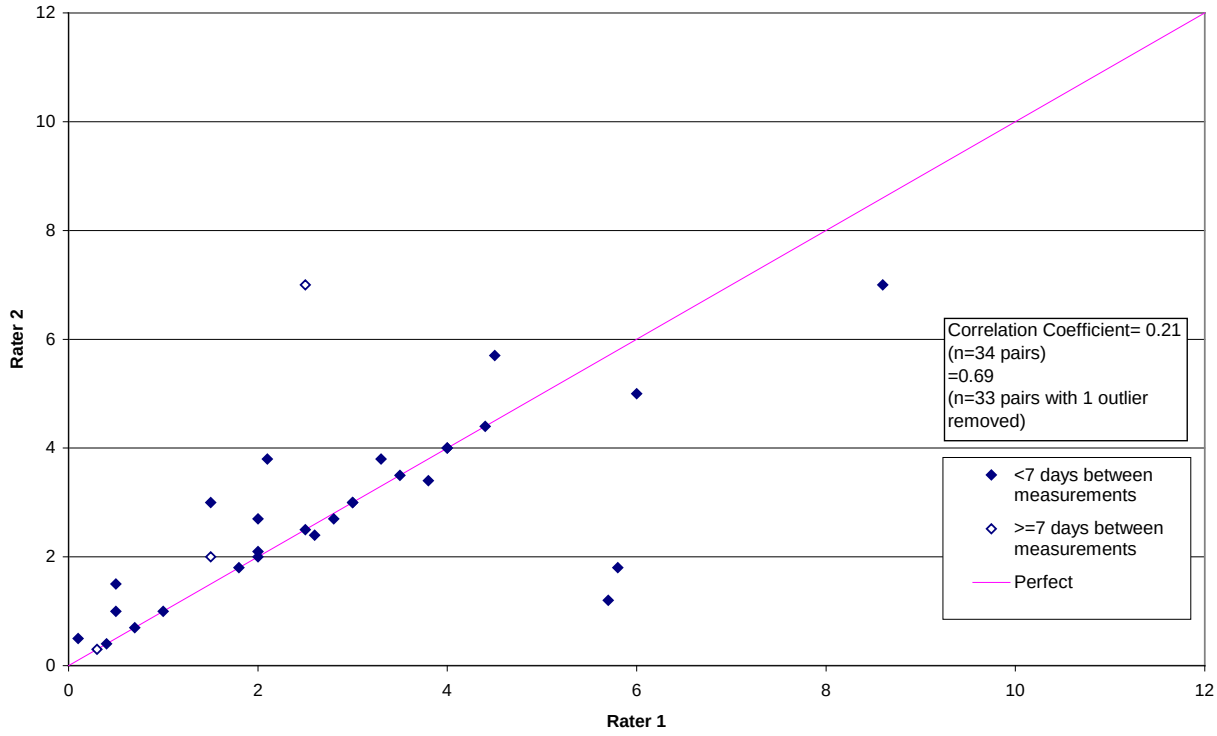


Figure 7-6. Stage 3/ 4 Pressure ulcer depth measured by the two assessors

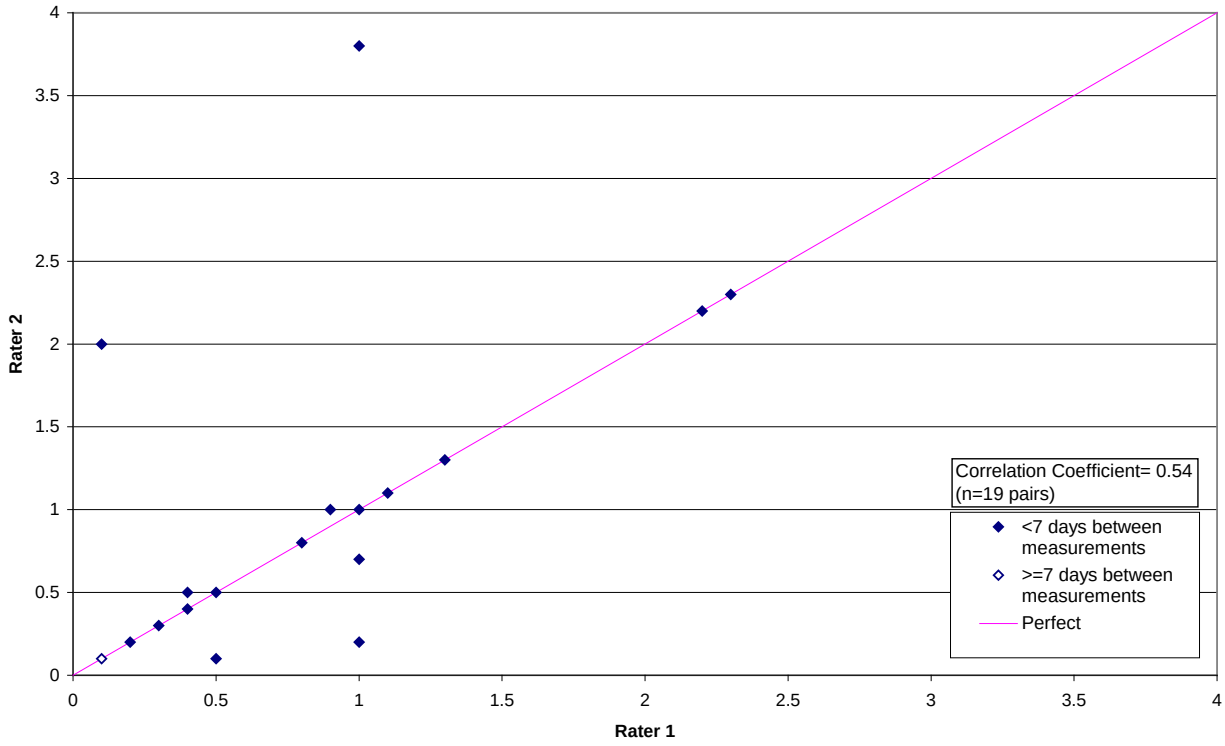


Figure 7-7. Stage 2 Pressure ulcer length measured by the two assessors

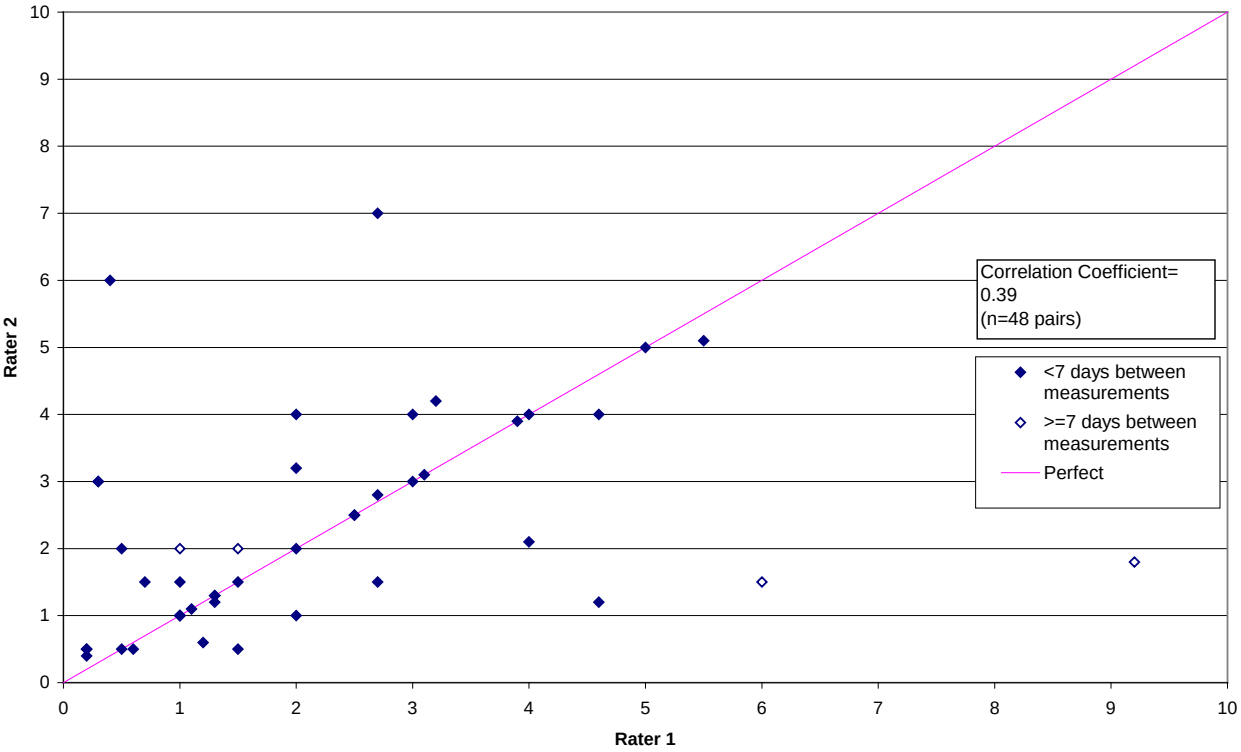


Figure 7-8. Stage 2 Pressure ulcer width measured by the two assessors

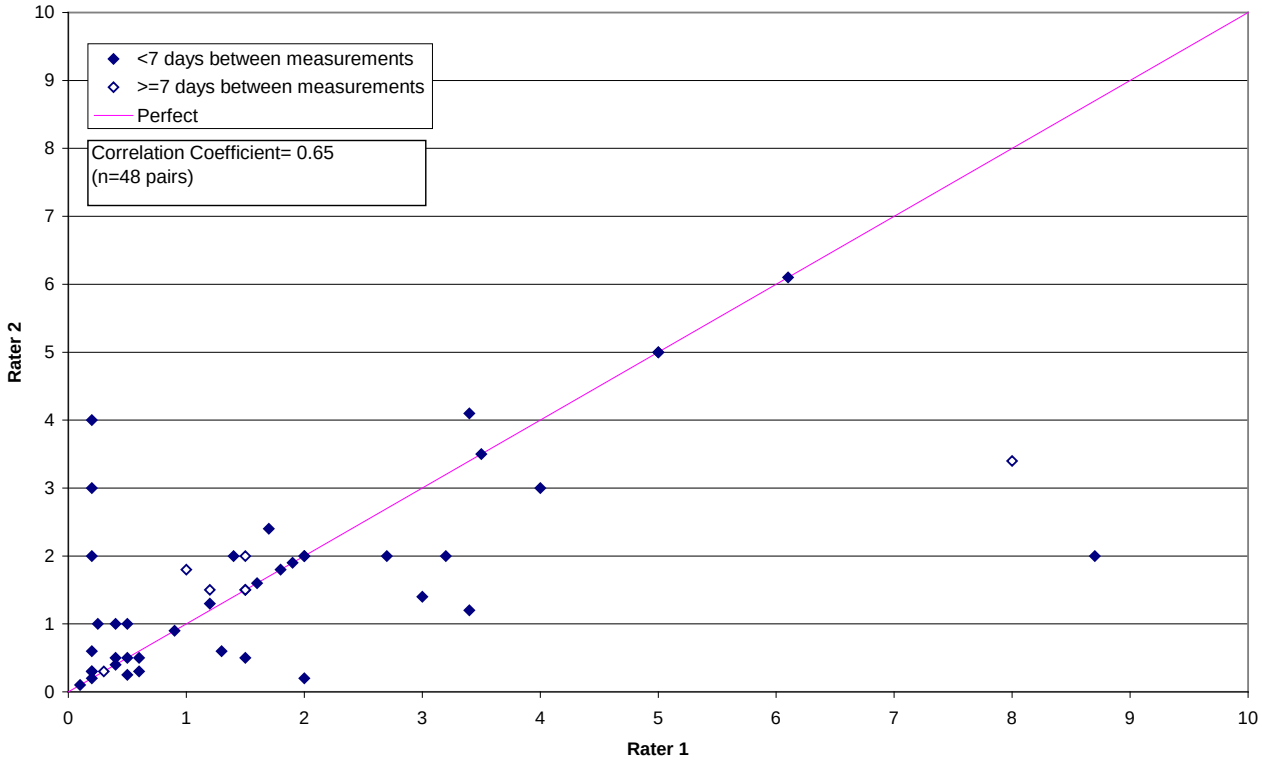


Figure 7-9. Pressure ulcer length for resident with 1 ulcer measured by the two assessors

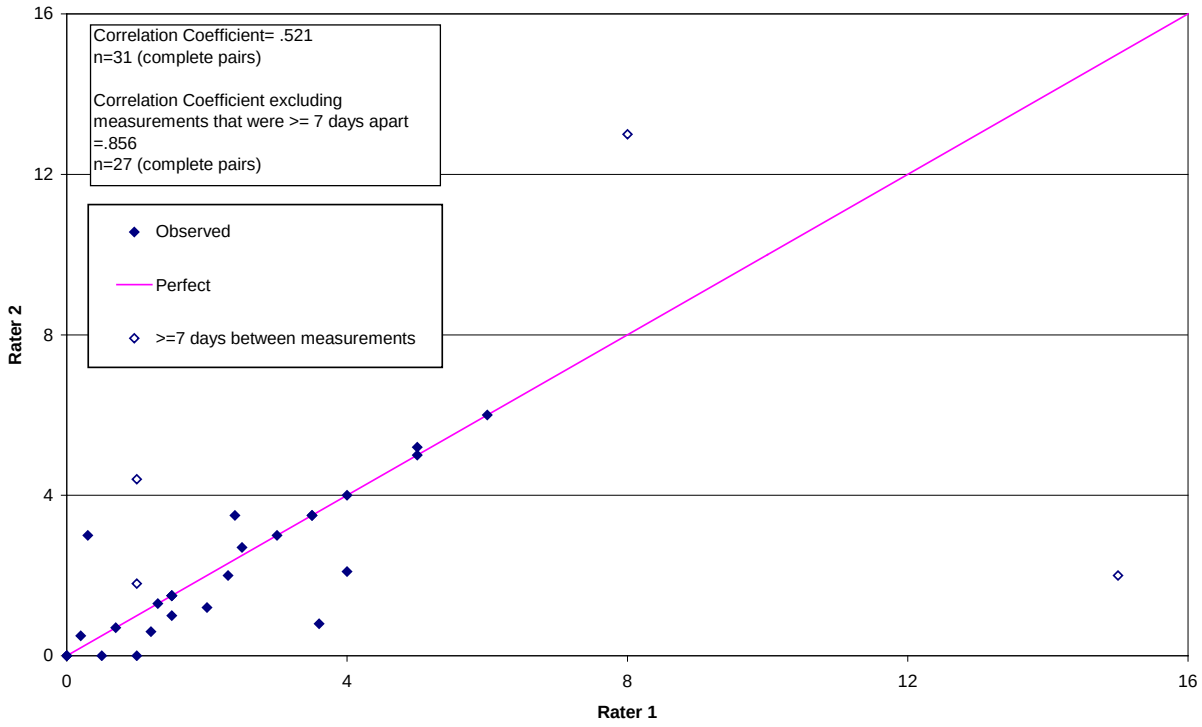


Figure 7-10. Pressure ulcer width for resident with 1 ulcer measured by the two assessors

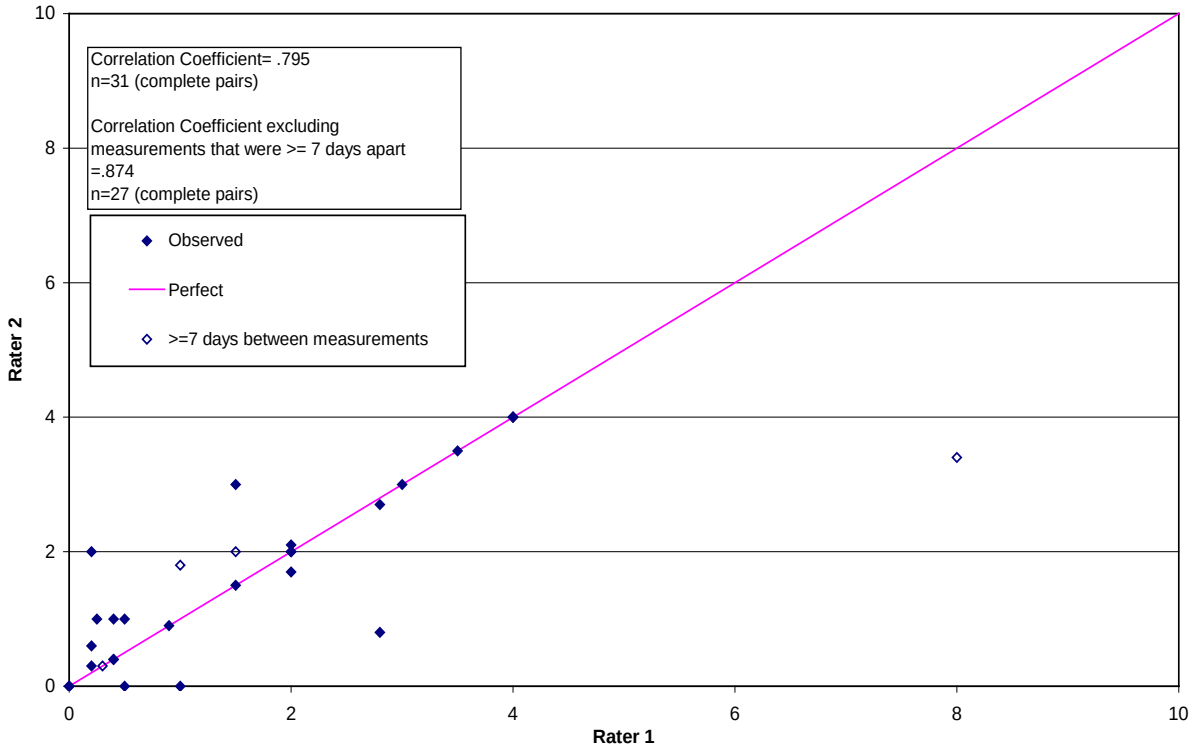


Figure 7-11. Pressure ulcer depth for resident with 1 ulcer measured by the two assessors

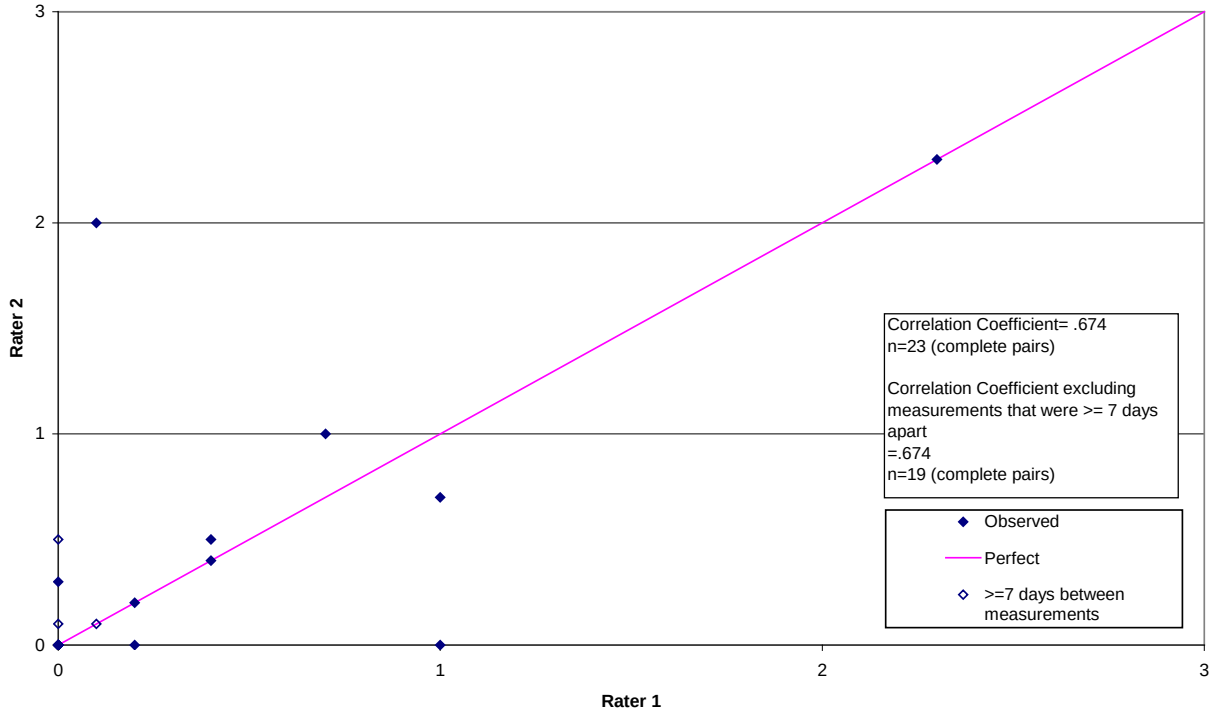
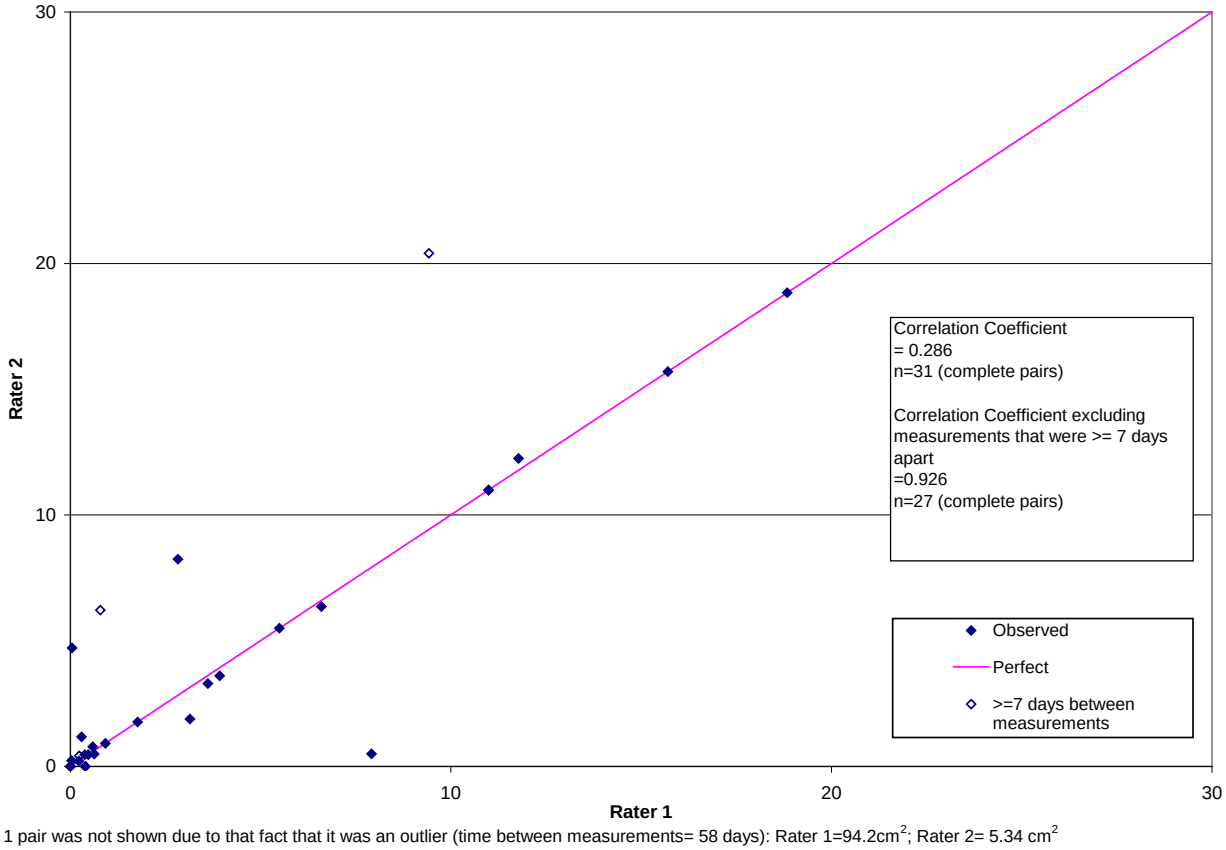


Figure 7-12. Pressure ulcer area for those with 1 pressure ulcer measured by the two assessors



1 pair was not shown due to that fact that it was an outlier (time between measurements= 58 days): Rater 1=94.2cm^2; Rater 2= 5.34 cm^2
*equation used in area of ellipse= (π*length*width)/4

** Note that measurements of 0.00 were possible because, even though the MDS indicated an ulcer (which was not necessarily a Stage 2 or higher pressure ulcer), no Stage 2 or higher pressure ulcer was present for measurement.

7.3 Distributions of Addendum Items

During STRIVE data collection, several Addendum[125] items were collected for study participants in addition to MDS items. Several of these are MDS 3.0 items, while others were from the interRAI LTCF, or OASIS-BI (Center for Health Services

[125] Final STRIVE Addendum (Version 6): STRIVE Addendum020107B.pdf
The 6 Addendum version dates for STRIVE data collection are:
1 - April 27, 2006
2 - June 9, 2006
3 - August 7, 2006
4 - September 5, 2006
5 - October 25, 2006
6 - February 1, 2007

Research, UCHSC, Denver, CO) items. Frequency distributions of all items are shown in Appendix C.

7.4 Public Release Files

At the completion of RUG-IV development and before the Final Rule for RUG-IV was published in August 2009, a documented STRIVE public data set was made available by CMS for public review. Included in the release were all of the variables required to recreate the RUG-III and RUG-IV results in the Notice of Proposed Rulemaking for RUG-IV in May 2009.

After the Final Rule was published, a revised STRIVE public data set with documentation was created for public release. This data set included all of the variables needed to create RUG-III and the Final Rule version of RUG-IV results.

The STRIVE Project public data set with documentation includes the following five items:

1) STRIVE Project Documentation for Public Release Feb 2010.xls
This Excel file includes 2 tabs:

- Data Dictionary. Includes the list of 100 STRIVE staff role time variable names, labels, type, lengths and comments. Variables include per diem reconciled minutes for nursing staff roles, per diem minutes for therapy staff roles, per diem Non-Resident Staff Time (direct and proportional) for staff roles, wage weights for staff roles, RUG-III and RUG-IV classification outcomes, STRIVE sample case weights, and case number.
- SAS Code. SAS code used for obtaining mean outcomes shown in file rug4_66grp_rawmin_24feb_2010.xls. This code can be revised by substituting each of the remaining 5 RUG classification variables listed in the data dictionary to obtain mean outcomes shown in 5 additional files:
  rug4_57grp_rawmin_24feb_2010.xls,
  rug4_48grp_rawmin_24feb_2010.xls,
  rug3_53grp_rawmin_24feb_20101.xls,
  rug3_44grp_rawmin_24feb_20101.xls,
  rug3_34grp_rawmin_24feb_2010.xls

2) strive_021210.sas7bdat
This SAS data set includes the 100 STRIVE variables defined in “STRIVE Project Documentation for Public Release Feb 2010.xls” (see #1).

3) strive_110609.csv
This is a comma-delimited data file of the 100 STRIVE variables defined in “STRIVE Project Documentation for Public Release Feb 2010.xls” (#1). It contains exactly the same observations and data as in the SAS data set (#2, above), but in a format that may be more accessible for non-SAS users.

4) Files of Adjusted Raw Minutes for State Use

These six files contain the times by staff roles that states may use in creating state specific CMIs. The six files are:

rug3_34_adjstd_raw_min_for_state_use_20100224.xls
rug3_44_adjstd_raw_min_for_state_use_20100224.xls
rug3_53_adjstd_raw_min_for_state_use_20100224.xls
rug4_48_adjstd_raw_min_for_state_use_20100224.xls
rug4_57_adjstd_raw_min_for_state_use_20100224.xls
rug4_66_adjstd_raw_min_for_state_use_20100224.xls

5) Memo: STRIVE_Data_Release_Memo_April 2010.doc

STRIVE Public Data Set:

(The STRIVE public data release included the set of variables to create RUG-III and RUG-IV results.[126] Because STRIVE was a real-time collection of staff-time utilization, the RUG calculations based upon therapy time use STM variables. It is important to note several STRIVE algorithms used for calculating RUG classifications:

MDS 2.0 Item Disease Diagnosis

Resolved disease indicator variables used in RUG-IV are turned on based on both the MDS 2.0 disease checklist and appropriate ICD-9 codes from the I3 ICD-9 code list. For example, resolved item I1y_dx (Parkinson's Disease indicator) is turned on if MDS 2.0 item I1Y is checked (value of 1 in an MDS record) or the I3 ICD9 code list includes 332 or 332.1, the codes for Parkinson's.

RUG-III only uses MDS disease checklist variables, such as I1A, I1R, and I1S, and not ICD-9 codes.

MDS 2.0 Resolved Items for RUG-III

---

[126]Prior to public release several data issues were approved by CMS. These include the following:

1 Scramble order of observations
2 Identify observations only by a case number (1 to 9721)
3 Omit Category qualifier variables (e.g., n_spec1_iv)
4 Omit Numerical (and sortable) number for RUG-IV groups
5 Include single high outlier dropped in calculating CMIs
6 Include nursing staff times for RN, LPN, and aide
7 Omit Other staff times not used in analyses(e.g.., administration)
8 Include per diem therapy adjusted times
9 Assign RUG-IV code "AAA" to 14 observations with missing data
10 Omit Stratum (hospital-based, etc.)
11 Omit State identifier
12 Omit CMIs for RUG-IV
13 Omit "Medicare" Identifier
14 Name for therapy: "adjusted"
15 Provide wage weights

After STRIVE data were released (04-20-2009), CMS requested an addendum to be released that included #13 above: Medicare Part_A was released 05-29-2009.

K5A (parenteral/IV), P1AC (IV med Tx), P1AI (suctioning Tx), P1AJ (tracheostomy), P1AL (vent Tx) MDS 2.0 items are resolved to account for under reporting: Count if STRIVE MDS or RPT MDS indicates "yes."

RUG-III Rehabilitation Therapy Classification Based on STRIVE STM

MDS 2.0 rehabilitation therapy items P1BAA (speech days), P1BAB (speech minutes), P1BBA (occupational therapy days), P1BBB (occupational therapy minutes), P1BCA (physical therapy days), and P1BCB (physical therapy minutes) are resolved variables set to STRIVE STM values. Concurrent therapy is not allocated.

MDS 2.0 Resolved Items for RUG-IV

K5A (parenteral/IV feeding), P1AC (IV medications), P1AI (suctioning), P1AJ (tracheostomy care), P1AL (ventilator) MDS items are resolved to account for under reporting: Count if STRIVE MDS or RPT MDS indicates "yes." For residents with days of stay <=7 and who have a STRIVE MDS, then resolved variables use STRIVE Addendum post admission data XP1 Items for K5a (IV medications), P1AG (oxygen therapy), P1aI (suctioning), P1AJ (tracheostomy care), P1AK (transfusion), and P1AL (ventilator).[127]

RUG-IV Rehabilitation Therapy Classification Based on STRIVE STM

Rehabilitation therapy items P1BAA (speech days), P1BAB (speech minutes), P1BBA (occupational therapy days), P1BBB (occupational therapy minutes), P1BCA (physical therapy days), and P1BCB (physical therapy minutes) are resolved variables: Day values are STRIVE STM values, and Weekly minute values are STM values deflated to account for allocation of concurrent therapy times.

[127] The STRIVE Addendum form is provided in the STRIVE Phase I report in Appendix A-4.