

# Comparing case-mix systems for nursing home payment

by Brant E. Fries

*Case-mix systems for nursing homes use resident characteristics to predict the relative use of resources. Seven systems are compared in structure, accuracy in explaining resource use, group homogeneity, and ability to identify residents receiving heavy care. Resource utilization groups, version II (RUG-II), was almost uniformly the best system, although management minutes*

*and the Minnesota case-mix system were also highly effective. Relative weights for case-mix groups were sensitive to cost differences and should be recomputed for new applications. Multiple criteria should be used in choosing a case-mix system, including consideration of inherent incentives and how residents' characteristics are defined.*

## Introduction

A current trend in Federal and State payment for nursing home care has been toward systems in which the differences among residents in resource use are recognized explicitly. In such systems, some portion of the payment rate is associated with measures of the resources used in caring for a facility's residents, the facility's case mix. Operationally, residents are assessed and a case-mix measurement system is used to determine the relative measure of resources used in each facility. This relative measure is employed to determine rates either directly (e.g., by scaling payment for nursing services) or indirectly (e.g., by determining rate ceilings or peer groups of facilities). The mechanics of translating measures of case mix into rates are complex, especially considering the incentives and disincentives embodied in any payment system (Schneider et al., 1985; Liu et al., 1986; Holahan and Sulvetta, 1987). We focus here, however, on the more technical problem of deriving appropriate measures that predict resource use by nursing home residents. The issues in this latter focus are not fully separable from those in the former, in that the choice of a case-mix measurement system will in itself create incentives and disincentives when applied for payment. Recognizing this linkage, selecting an appropriate case-mix system is the precursor to developing an effective payment system. I sought to evaluate and compare several nursing home case-mix systems, to understand their commonalities and differences, and to examine their ability to predict accurately actual resource use.

## Background and approach

Interest in case-mix measurement systems for nursing homes began in earnest in the mid-1970s with work at the Battelle Human Affairs Centers (McCaffree, Winn, and Bennett, 1976; McCaffree, Baker, and Perrin, 1979; and Winn, 1975) and The Johns Hopkins University (Cavaiola and Young, 1980, and Flagle et al., 1977). Since then, numerous case-mix systems have been developed. The criteria of some have been published: Deane and Cella,

1981 (in use in Maryland); Weissert et al., 1983; Fries and Cooney, 1985; Cameron, 1985; Morris et al., 1987; Arling et al., 1987; Arling, Zimmerman, and Updike, 1989; and Schneider et al., 1988 (in use in New York). Several other case-mix systems, unpublished, have been implemented as part of Medicaid State nursing home payment systems, such as those in Ohio, Illinois, West Virginia, and Minnesota. In all of these systems, resident characteristics are evaluated directly in determining case mix. Although considerable success has been achieved in understanding which institutional characteristics are predictive of costs for use in policy analysis (e.g., Birnbaum et al., 1981; Liu et al., 1986; and Sulvetta and Holahan, 1986), these appear to be less appropriate for setting payment than are patient characteristics. The residents for whom a facility cares rather than the characteristics of the facility should determine case mix.

Despite the plethora of case-mix systems available, little work has been done to contrast their operation and evaluate their ability to predict resources. A major effort in developing these systems is the customized collection of necessary data. Along with the effort to assess resident characteristics, this involves the substantial cost and difficulty of measuring actual resource use. Thus, it is not surprising that developers utilize each scarce, expensive data point to the fullest in the derivation of a system and, in reporting a new system's abilities, utilize the same data from which the system was originally derived. In this study, I compare several alternative resident classification systems to understand how they operate when applied to new populations without adaptive adjustments to improve their variance explanation (a measure of how well the system explains differences among residents in resource use) or other properties—that is, applied "out of the box."

## Methodology

The two major objectives of this study are to evaluate the structural similarities and dissimilarities of several major case-mix systems based on resident classification and to contrast each system's ability to explain resource utilization across several data sets. In this section, I discuss the systems evaluated, the choice of data sets for the analysis, and the methodologies employed to address these objectives.

Reprint requests: Brant E. Fries, Ph.D., Institute of Gerontology, The University of Michigan, 300 North Ingalls, Ann Arbor, Michigan 48109-2007.

## Selection of case-mix systems

Out of a number of potential candidates, the case-mix systems evaluated were those currently in use or under consideration for nursing facility payment. The several systems chosen fell into four groups:

- Maryland case-mix reimbursement system (classification and index).
- Minnesota case-mix reimbursement system.
- Resource utilization groups (RUGs)—three systems.
- Management minutes.

The first two systems are currently in use by their respective State Medicaid programs. The Maryland system is considered here as two different systems: one including only four major groups, the second an index representing the reimbursement provided based on these groups plus specified services. The original resource utilization groups system (RUG-I) was employed until 1986 by the Veterans Administration for resource allocation (akin to reimbursement). The second version of resource utilization groups (RUG-II) was developed for the New York State Medicaid program, is in use by the Department of Veterans Affairs and in a Texas Medicaid payment demonstration, and is under consideration for use in several other States. The third allied system, resource utilization groups—title 18 (RUG-T18), was derived for Medicare residents and is being used in a Federal payment demonstration. The management minutes system was developed for use in internal management of nursing homes. It provides the organizing structure for management information systems of the Hillhaven Corporation and the National Health Corporation, and it is being modified for use in the Massachusetts payment system. These systems are described briefly in the following section.

These seven case-mix systems represent the two basic constructs employed in predicting resource utilization. Most of the systems classify residents into homogeneous groups that are mutually exclusive and exhaustive of all residents found in nursing facilities. Associated with each group is a case-mix index (CMI), which represents, at least relatively, the time or cost of the average resident in the group. For classification systems, I therefore evaluated separately the wisdom of using the recommended groups of residents and the accuracy of the CMI for each group. For the remaining two index systems—the Maryland index and management minutes—weights were given to a large number of resident characteristics and an index value for each resident was determined directly. For these latter systems, I evaluated only the appropriateness of the index values.

Aside from the many experimental systems cited earlier, several other notable systems were not included. For example, prior to 1985, the reimbursement system in New York State was based primarily on facility characteristics (O'Donnell and Hannan, 1983). In the resident classification methodologies embedded in the Illinois, West Virginia, and Ohio Medicaid nursing home payment systems, services are used as the primary indicators of resource use; thus, resource use is recorded rather than predicted. Although several of the systems considered in this project also rely on the provision of

service as a resident characteristic, such variables are in more secondary roles. In a later section, I consider in greater depth the relationship of these issues to the systems evaluated.

## Description of case-mix systems

The several case-mix systems employed in this analysis are described briefly here. Additional description can be found in the works cited in the descriptions or in Grimaldi and Jazwiecki (1987).

### Maryland system

The State of Maryland Medical Assistance Nursing Home Reimbursement System can be used both as a patient classification and an index system. The classification system, used in Maryland since 1983 (*Maryland Register*, 1982), is based on the work of Deane and Cella (1981). The basic system consists of four groups of residents, differentiated by the number of activities of daily living (ADLs)—bathing, dressing, mobility, continence, and feeding—in which the resident is dependent. Residents with up to two dependencies form the light-care group, those with three or four form the moderate-care group, and those with a full five dependencies form the heavy-care group. A fourth group, heavy special care, is then identified as those in the heavy-care group who, in addition to five ADL dependencies, receive care for Stage III or IV decubiti, tube feeding, or 24-hour turning and positioning.<sup>1</sup> This entire system is denoted "Maryland groups." I employed as CMIs the Maryland reimbursement level for each group for fiscal year 1987.

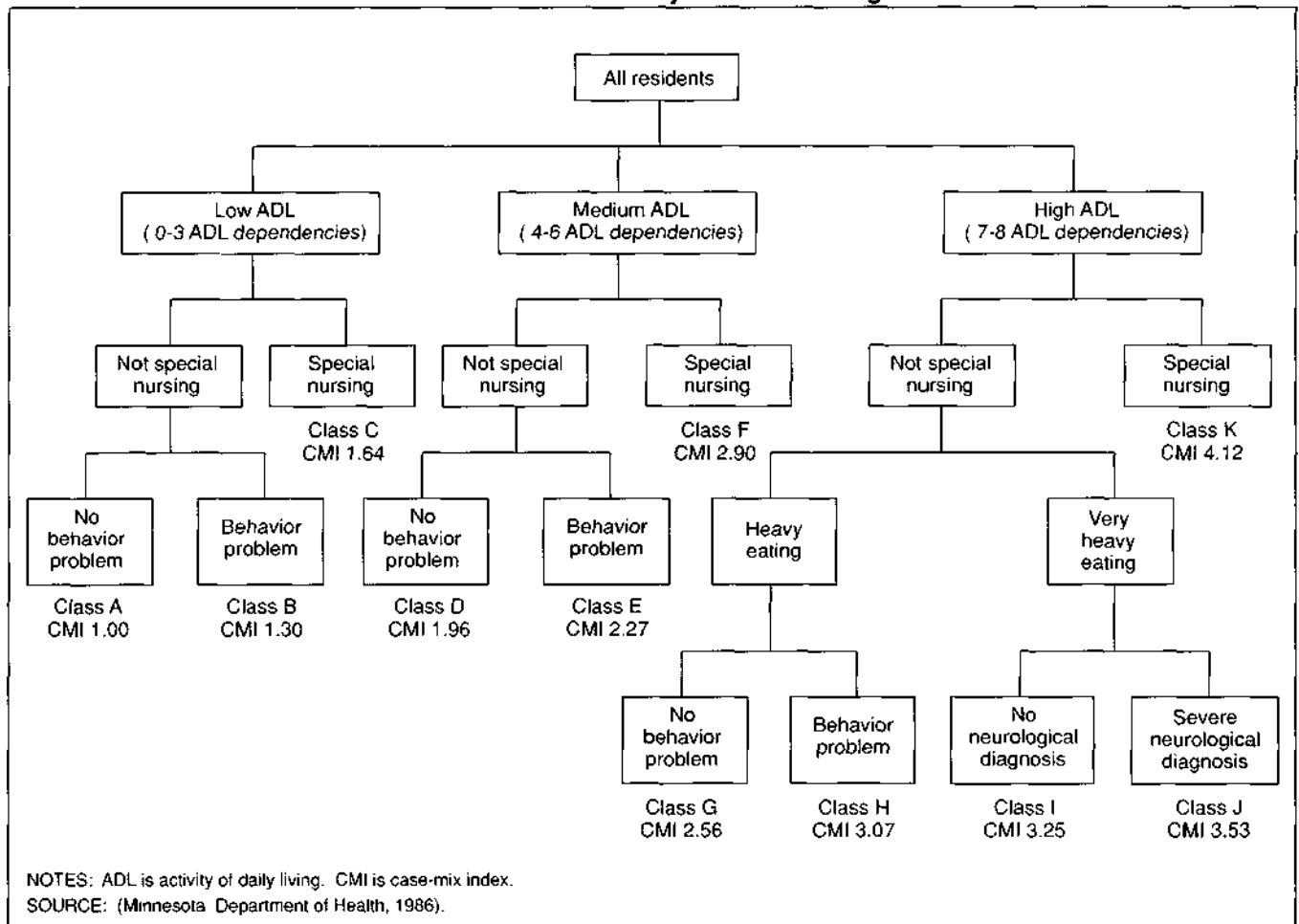
In our modeling, an index system was derived as well, based on the computations involved in setting rates in Maryland. This system included flat payments for up to 10 additional services: decubitus ulcer care, tube feeding, turning and positioning, single injections daily, multiple injections daily, oxygen or aerosol therapy, intravenous (IV) care, restraints or protective devices, ostomy care, and suctioning or tracheostomy care. The Maryland index was computed by adding the 1987 payment level for these services to the CMI for the Maryland group.

### Minnesota system

The Minnesota Department of Health Quality Assurance and Review uses the Minnesota case-mix system for setting payments to nursing homes and certified boarding care homes (Minnesota Department of Health, 1986). This system was developed under contract by Lewin and Associates in 1985. As with the RUG systems, cluster analysis was used to derive groups of residents with similar resource consumption. A total of 11 groups are specified, as shown in Figure 1. The first major split is based on the number (0-3, 4-6, 7-8) of ADL dependencies from a group of eight—bathing,

<sup>1</sup>The system was recently revised to change the criteria for heavy special care, dropping the condition of 24-hour turning and positioning and adding the provision of intravenous fluids and communicable disease care.

**Figure 1**  
**Minnesota classification case-mix system for nursing home residents**



dressing, grooming, eating, mobility, transferring (getting in or out of a bed or chair), toileting, and walking. After this split, subgroups are defined for residents requiring special nursing. Special-nursing patients receive at least one treatment—tube feeding, oxygen and respiratory therapy, ostomies and catheters, dressings, skin care, hyperalimentation or Hickman catheter, IV fluids or medications, blood transfusions, drainage tubes, symptom control for the terminally ill, or isolation precautions—or clinical monitoring every shift. Residents not requiring special nursing yet in the first two ADL groups are differentiated if they have frequently one or more behavior problems, ranging from disorientation and wandering to being physically abusive. These splits produce a total of six groups for those residents with up to six ADL dependencies. For those in the highest ADL group, a special-nursing group is identified first; other residents eventually are classified in one of four groups. Those denoted “very heavy” require partial or full assistance in eating and are split into two groups according to the presence or absence of a neurological diagnosis (including diseases of the neurological system, head and spine fractures, and neoplasms of the brain or spine). The “heavy” group, those without eating difficulties, are split, as with the other, less

ADL-dependent groups, by the presence or absence of a behavioral problem. The resultant 11 groups are designated A-K. For each group, a CMI is specified, with the CMI for the lowest group, A, set at 1.00.

### Management minutes

The management minutes system was developed by William Thoms and is employed by the Hillhaven Corporation as part of their patient accounting system (Hillhaven Corporation, 1985) and by the National Health Corporation. In this system, an index is constructed for each resident based on a spectrum of resident characteristics, each with a specified weight. The index values are proposed as actual nursing times, so total management minutes for a resident are expected to correspond directly to staffing needs. The system and weights were derived by expert opinion rather than statistical analysis. First, time values for each of 18 items are added: dispensing medications and charting (18 minutes provided to everyone); skilled observation (15); assistance (18) or total care (20) in personal hygiene; assistance in dressing (30), mobility (32), or eating (20); feeding (45), tube feeding (90), IV feeding (90), or tube and IV feeding (135); indwelling catheter (20),

incontinence (48), bowel incontinence with catheter (18), bladder training (50), or bowel training (18); and positioning (36) and decubitus prevention (10). To this total are added, as appropriate, 10 minutes for each "skilled procedure" from the following list: special skin care, decubitus care, IV, oxygen therapy, tracheostomy care, ostomy care, intake and output fluid monitoring, sterile dressings, suctioning, drug regulation, multiple injections, irrigation or special catheter care, inhalation therapy, pain control, splint assistance, ventilator, transfusion, chemotherapy, and radiation therapy. If a resident requires special attention (isolation; special attention for being noisy, obese, spastic, or rigid), then the cumulative score is increased by 10 percent. Finally, 30 minutes are added for any resident who receives restorative nursing care, including any of the following: patient or family education, range-of-motion exercises, meal training, ambulation training, reality orientation, ADL training, heat treatment, and speech training. Special rules make some of the categories exclusive or inclusive of others.

### Resource utilization groups, version I

The pilot version of the resource utilization groups system (RUG-I) was derived using cluster analysis to explain differences in a subjective estimation of the relative time spent caring for a resident (Fries and Cooney, 1985). Only four variables are employed in defining the nine categories: the ADLs for dressing, ambulation, and feeding and whether fluid intake and output is monitored, the last differentiating the two highest groups. As with all of the grouping systems, each resident is assigned to a single group and a case-mix index is specified for each group.

### Resource utilization groups, version II

The RUG-II system was a major redevelopment of the RUG-I system, part of the development of a case-mix Medicaid payment system for New York nursing homes. A wide variety of criteria were used in deriving the groups. In addition to the measures traditionally used to explain resource use, these criteria included clinical rationality; homogeneity of groups; incorporation of incentives; and choice of characteristics that could be assessed validly, reliably, and through audits (Schneider et al., 1988).

The RUG-II system is a combination of two major constructed variables. The first is a typology of residents, a hierarchy ranked from high to low resource use, that classifies residents into five groups:

- Heavy rehabilitation (R).
- Special care (S): with serious medical problems—e.g., comatose, with Stage IV decubiti, or quadriplegic.
- Clinically complex (C): requiring significant medical treatment—e.g., chemotherapy, transfusions, or treatment for dehydration—or requiring a physician visit at least once a week.
- Severe behavioral (B): manifesting severe physical aggression, regression, verbal abuse, or hallucinations.

- Reduced physical function (P): all residents who do not fall into any of the first four groups.

Residents are classified into the first hierarchical group for which they qualify, as this classifies them by their most serious and resource-intensive problems.

The second component of the RUG-II system is an ADL index, in which the scores for three individual ADL variables—toileting, eating, and transferring—are combined. Each of the three ADLs is scaled from one (requiring, at most, supervision) to three (incontinent but toileted, hand fed, continuous physical assistance of two persons, or bedfast). A level four exists for classifying residents who are fed parenterally or by tube. Values for the three ADLs are then summed to an ADL index, ranging from 3 to 10.

The 16 RUG-II classifications are formed by first locating residents within the clinical hierarchy, then classifying them in individual RUGs within the hierarchy category using ranges of the ADL index. The two-letter RUG-II group name represents first the hierarchy letter, then a letter from A to E denoting a range (different for each hierarchy group) of the ADL index. Thus, RA would be the most functional heavy-rehabilitation residents.

Two case-mix indexes were derived for the RUG-II system. The nursing index is based solely on nursing time; the total index is based on nursing plus therapy (physical, occupational, and recreational therapy and social work). Each is normalized so that the average resident in the sample has a value of 1.0.

### Resource utilization groups, title 18

An adaptation of the RUG-II system was found to be the best resident classification system for Medicare residents in skilled nursing facilities when evaluated on a variety of criteria similar to those used to develop the RUG-II system (Fries et al., 1989). With 20 groups, the RUG-T18 system has the same basic structures as the RUG-II system—a typology of resident types and an ADL index—although there are some new elements as well. The criteria used for inclusion in the five-group hierarchy of residents are the same as those for RUG-II, except that the goals of both maintenance and restoration are recognized for rehabilitation residents. In the RUG-T18 system, heavy-rehabilitation residents are split into one of three groups according to the number (one or fewer, two, or three) of the three types of rehabilitation therapies (physical, occupational, and speech) they receive for an average of more than 30 minutes per day. These three rehabilitation groups are then further split, as are the other hierarchy groups, by the RUG-II ADL index. Case-mix indexes are derived for all except three groups, in which few or no residents are classified.

### Selection of data sets

In this study, I utilized three existing data sets, chosen for their inclusion of directly obtained measures of resource use and a spectrum of resident characteristics

broad enough to permit me to model the several alternative classification systems.

### **New York State Patient Assessment Instrument**

This data set contains information on 3,427 residents in 52 New York State nursing homes. Collected in 1984, these data were used in the derivation of the RUG-II system (Schneider et al., 1988). Represented in the sample are residents of both skilled nursing facilities (SNFs) and intermediate care facilities (ICFs). Heavy-care residents were oversampled by choosing units with severe cases in the sampled facilities.

### **Patient Assessment Instrument for Medicare**

This data set contains information on 2,312 residents (3,269 observations) in 38 SNFs in California, Florida, Illinois, New York, and Pennsylvania (Fries et al., 1989). Facilities and units were chosen that had high percentages of Medicare residents. In this study, I utilized the entire data set, including 2,564 assessments of 1,684 Medicare residents and 705 assessments of 628 non-Medicare residents in the same nursing units. Medicare residents and, to a lesser extent, non-Medicare residents in the sample were significantly different from the average SNF or ICF resident, having more medical complications and fewer long-term care problems such as contractures.

### **Texas Client Assessment and Research Evaluation**

As part of the development of a Medicaid payment system, data were collected on 1,997 nursing home residents in Texas (Texas Department of Human Services, 1986). The 1,906 observations deemed usable in the Texas project were employed in our analyses. Texas nursing homes, with almost all residents at the ICF level of care, have somewhat lower staffing than those in New York.

A total of 8,602 observations from the three data sets were available for analysis. To my knowledge, this is the largest nursing home data set of its type ever assembled.

In all three sets, data on actual time spent by staff in caring for residents were collected with the same protocols. Full-time staff (nurses and aides) self-reported their time during either two or all three of a day's shifts, with a variety of checks performed to assure that these measurements were accurate. Other staff and therapy times were self-reported over a day- or week-long period. Time spent on activities not centered on patients (e.g., inservice training, restocking shelves, and routine documentation) was collected and then computer-allocated to individual residents. The measures of resource use derived for our analysis included nursing time; total time (nursing plus therapies), when available; and costs (wage-weighted nursing or total times, with weights specific to the category of caregiver, for example, registered nurses, licensed practical nurses, and aides).

### **Philosophy of analysis**

A major design issue in this project was the extent to which I should remove outliers or correct

irregularities, which were identified either through data analysis or through information from personnel in the projects developing the data. I decided not to make any adjustments other than to correct errors (i.e., 91 observations dropped from the Texas data set), as I was concerned that the choice of outliers might bias results for individual systems. This led to my overall philosophy of modeling the several systems "out of the box," without adaptive adjustments to improve them. One implication of this approach is that the results obtained may indicate variance explanation of resource use substantially lower than that reported by the systems' developers and lower than the variance explanation that these same systems might achieve when applied to large populations, in which outliers have relatively less impact.

### **Crosswalks**

Each of the three data sets was collected under a different protocol and, more importantly, with different assessment questions and possible responses to those questions. Thus, it was impossible to merge the sets under common definitions of data elements. Equally important, each of the case-mix systems required particular data elements that were missing from certain data sets or, when available, required response categories that did not match those available. A major effort in this study was the evaluation of each pair of data set and system, to develop "crosswalks" for the most appropriate assignment of a case-mix group or index to each resident. Experts who either developed or managed each system, along with those who collected each data set, reviewed the crosswalks. Again, in congruity with the decision on the treatment of outliers, I ran systems as they were specified rather than attempting to fit them to a particular data set (for example, to maximize variance explanation). Although fitting the systems might have led to a more appropriate picture of how a system would operate if adapted for use in a given State, it also might have unfairly biased the comparative results.

An example of the problems faced is the use of diagnoses in the RUG-II system. Residents are assigned to particular groups based on their primary diagnosis, for example, hemiplegia. Consider the situation of a resident with hemiplegia listed as a secondary diagnosis. For this study, should the resident be classified into a higher group based on the reporting of hemiplegia as a secondary diagnosis? If this resident resided in a facility under a case-mix payment system based on RUG-II, which provides additional revenues for caring for hemiplegic residents (but only when hemiplegia is indicated as a primary diagnosis), the facility would attempt to report the hemiplegia as the primary diagnosis, an example of "assessment creep." Such actions would be best modeled by searching for any mention of hemiplegia in a list of diagnoses. By looking at only primary diagnoses in the analysis, I underestimate case mix. However, in practice, such "creep" would need to be controlled by regulations requiring that the primary diagnosis be the major focus of care planning. Such a situation would be more closely approximated by ignoring the secondary condition of hemiplegia. Although an argument could be mounted for each option, in balance, the latter option was deemed more accurate.

The crosswalks are a critical part of the methodology. The decisions made in developing the crosswalks may, by themselves, have had a significant effect on the relative performance of the systems modeled. My experience, however, has been the reverse. In allied projects, the lack of congruent items only slightly affected the performance of case-mix systems, as errors were automatically corrected by the intrinsic intercorrelations among data elements.

## Analysis

The goal of the data analysis was to evaluate the performance of each system and the accuracy of published case-mix weights for each group (for classification systems) or parameters used in predicting resource use (for index systems).

In initial analysis, I considered the structure of the several systems. Then I evaluated the performance of each of the case-mix systems on each of the three data sets. Performance was measured in part by variance explanation, how well a system explained differences seen in actual resource use by individual residents. For index systems (the Maryland index and management minutes), variance explanation was measured employing linear regression. For the others (the classification systems), I used both analysis of variance based on the classification groups and a linear regression on the constructed index (i.e., the case-mix index, or weight, for each category published by the system developer). The dependent variables were nursing time and cost, and total time and cost (for nursing plus therapy staff). I considered as well logarithmic transforms of these variables.

In addition, residual analysis provided insight into types of residents for whom resource use was poorly predicted and the need for data transformation. Other methods of assessing performance included analyses of the homogeneity of individual groups, as measured by the coefficient of variation, and analyses of the variance explanation for subgroups of residents—in particular, the heaviest care residents, those whose nursing time or cost puts them in the upper quartile of the patient population.

A second analysis was concentrated on the case-mix indexes for the five classification systems. I assessed their appropriateness, especially for data sets other than that from which each was derived. Because of the differences in definitions of the CMIs, I normalized each system on each data set so that the average resident had a CMI of 1.00. Similarly, I normalized the staffing and cost values for each resident. These operations permitted me to compare different CMIs (e.g., those based on nursing or total staff time) and costs without regard to the unit of measure.

## Results

In this section, I present an overview of the data sets, followed by the results of the comparative analyses.

## Major characteristics of data sets

There are significant differences in resident characteristics across the three data sets, differences that follow expected patterns of care. For example, in the New York State data set, long-stay residents in ICFs and SNFs contribute to the higher percentage of residents who have contractures and are unable to understand well. Such patients are less common among the short-stay, heavy-rehabilitation SNF residents in the Medicare data set. The total number of minutes spent by all nursing staff was somewhat lower for Texas (91 minutes for the average resident) than for New York (101 minutes), although some of this difference may result from oversampling of heavy-care patients. For the Medicare sample, data were collected over only two shifts (day and evening); staff times for a full 24-hour day would be larger yet. Nevertheless, as I found in previous research, night shift times are highly correlated with the sum of day plus evening shifts. Therefore, using data from only two shifts, relative resource use for residents is still maintained. The average in this sample of 108 minutes for two shifts should be inflated to 136 minutes per day for comparison purposes, as discussed later.

## Structure of case-mix systems

In an initial comparison of the several case-mix systems, I considered their structure alone, without use of data or statistical analysis. I considered both the resident characteristics employed in a system and the overall structure of how these characteristics are combined in a system.

### Resident characteristics

The variables employed in all resource measurement systems fall into three major groups: ADLs, medical conditions and services, and mental and behavioral conditions.

*Activities of daily living*—Since the original pioneering work of Katz et al. (1963), ADLs have formed the cornerstone of nursing home resident classification. All the resource measurement systems considered here incorporate them in a major role. I used the ADL index of the RUG-II system as the benchmark for my comparisons. Variation exists across systems in which and how many ADLs are used and the number of levels recognized for each. From three ADLs (RUG-I, RUG-II, and RUG-T18 systems) to eight (Minnesota) are employed.

In developing the RUG-II system, colleagues and I found that the increased variability in resource use explained by using more than three ADLs was marginal at best. We also found that the choice of which three ADLs to use should be made based on criteria other than their statistical explanation of resource use. One criterion is whether their inclusive provides appropriate incentives for nursing homes and whether nursing homes can develop a system characteristic for residents at little cost

to themselves but resulting in increased case-mix payment—that is, whether the system can be “gamed” (Schneider et al., 1985). In the present study, I examine the performance of systems that utilize more than three ADLs as another test of the hypothesis that these additional ADLs are redundant.

An issue of potentially greater importance than the number of ADLs used is the number of levels of functionality recognized for each. Again, the earlier research demonstrated that considerable variance explanation resulted from recognizing the intermediate levels of ADL functioning, not just the dichotomy of independent versus dependent. All of the non-RUG systems use dichotomous ADL measures, although, in the Minnesota system, intermediate levels for eating are recognized for the residents most heavily dependent in ADLs. The RUG-II and RUG-T18 systems both employ an ADL index incorporating three levels for toileting and transfer and four levels for eating. Again, the tests performed here have the potential to validate whether distinguishing intermediate levels of ADLs is useful.

*Medical conditions and services*—Several medical conditions and services are seen in each system considered here except RUG-I, which utilizes only fluid monitoring. The lack of these variables, none of which is probably correlated with fluid monitoring, likely restricts the RUG-I system practically to applications involving less severely disabled residents.

The provision of service should be distinguished from the medical conditions that prompts service. Whenever possible, it is preferable to employ the underlying condition rather than the service in a case-mix system, to avoid gaming. Services are most appropriately involved in case-mix systems when they are costly, serious, or invasive to the resident and linked directly to the presence of a medical condition. On these grounds, services such as resident or family education, range-of-motion exercises, and reality orientation, which are included as conditions for special attention in the management minutes, may be inappropriate for payment application. I do not suggest that these services are insignificant in time or unimportant for the resident but that better indicators of resource use should be found. On the other hand, in Fries et al. (1989), we defended the direct identification of rehabilitation services, given their high cost and the lack of alternative methods to predict accurately the need for these services. In any case, the inclusion of services will increase variance explanation insofar as services are acknowledged both in the resource measure to be explained and in the explanatory variables measuring case mix. This emphasizes the need to evaluate case-mix systems using criteria beyond solely variance explanation.

*Behavioral conditions*—Three of the systems considered explicitly include behavioral conditions in a major role: the RUG-II, RUG-T18, and Minnesota systems. The Maryland system includes such conditions only if they lead to restraint, and management minutes includes “noisy” only as a criterion for the special-attention category. In the RUG-II and RUG-T18 systems, which are based on empirical results, behavioral conditions are considered only for residents who neither have severe medical conditions nor receive heavy rehabilitation services. The developers of the Minnesota

system similarly concluded that behavioral characteristics would be useful only in differentiating residents who do not have medical conditions represented in the special-nursing category. Whereas the RUG-II and RUG-T18 systems use only behavior characteristics—aggression, regression, hallucinations, and verbal abuse—the Minnesota and management minutes systems recognize mental conditions explicitly. The Minnesota system identifies mental conditions such as disorientation; the management minutes system gives less recognition to mental status, only including reality orientation as an element of restorative nursing care.

### Classification versus index systems

Classification systems, based on grouping residents with specified characteristics, are the result of the application of a variety of clustering methodologies. I compare this type of system with the index system, which combines weights for particular resident characteristics to form a single numerical value for each resident. The weights traditionally are determined by expert opinion, direct time studies, or using multiple regression (e.g., Cavaiola and Young, 1980). In the comparison here, I consider both the applications and the constructive methodology of these two genres.

On the one hand, a classification system based on clustering would appear to have several advantages over an index system:

- Variables potentially useful for predicting resource differences are highly correlated: One condition often is comorbid with others. Linear models such as regression and discriminant analysis may work poorly under such conditions of multicollinearity and always are awkward for handling interactions. Many of these interrelations will not be identified through expert opinion. However, it is in these very interactions that significance often lies. For example, changes in ADL functionality may be associated differently with resource use in residents with significant medical problems than in those without such problems. Index systems may sacrifice accuracy and reality by not recognizing such interactions.
- Any categorical variable of several levels describing a resident characteristic must be quantified in order to be used in most multivariate techniques, such as regression or discriminant analysis. One approach to such quantification leads to an untenably large number of new variables, with a variable representing all but one of the levels that each of the original variables can take. In a second approach, values for a single variable are assigned, but these values have little relationship to the relative impact of that level. Both solutions are awkward. As a result of these methodological problems, many developers of index systems either decide a priori to differentiate levels not associated with different resource usage or revert to more simplistic dichotomies (e.g., for the ADLs) to reduce complexity. For classification systems, clustering permits categorical variables to remain in their original form, using them for partitions (splits) only when improvement (reduced variance) can be achieved.



- Grouping residents may be more widely acceptable than individual levels for each resident, as this fits the paradigm of many biological and medical classification systems.
- The involvement of human intuition and clinical judgment can be a major advantage of clustering, as it can produce a system that permits users to visualize the residents in a group. For example, for the RUG-II system, ADLs provided the best initial splits. However, it was a combination of first a clinical split—by type of resident—followed by a split on ADLs that resulted in the system that had the best variance reduction. Myopically following only statistical rationale or automated procedures such as stepwise multiple regression would have resulted in a system that had poorer total variance explanation. Index systems based on expert opinion retain this type of intuition but do not support it with statistical analysis.

On the other hand, the additive index model also has significant appeal, for it follows our understanding that managing particular resident conditions directly adds to care time and resources. Index systems provide a single number that may be highly sensitive to changes in individual resident characteristics. This sensitivity has both advantages and disadvantages for payment systems, for small real changes may be correctly identified or may simply be daily fluctuations. A payment system can be made less sensitive, however, by using ranges of index values or averages over the period of a month.

## Prediction of resource utilization

A major goal of most case-mix systems is to predict actual resource use. These predictions were evaluated here, applying each system out of the box, as described earlier. It follows that my tests are more stringent than those of the developers, especially as the systems are applied to new data sets as validation. On the other hand, this analysis needs to be performed with care, as two of these data sets have a bias. The New York State data set was used in the derivation of the RUG-II system, so certainly this system can be expected to perform well in explaining resource consumption. The Medicare data base, in addition to being the source of the RUG-T18 system, contains data on both Medicare and non-Medicare residents on units that care for large numbers of Medicare residents. It thus provides only a partial test of resource measurement systems; that is, a test principally on a particular type of resident. Therefore, the use of the Texas data is especially important in evaluating the RUG-II and related RUG-T18 systems. Otherwise, my methods should provide unbiased comparison of the systems' accuracies.

After each classification system was used to assign each resident in each data set to groups, two types of analyses were performed to measure the prediction of actual resource use. Analysis of variance (ANOVA) was used to determine the variance explanation of the groups, as measured by the  $R^2$ . (Although not reported here, the same results in all analyses were found for adjusted  $R^2$  statistics.) Each group was considered only as a category, and its own group weight was therefore computed to fit

best the given data. This evaluation therefore provides an upper bound for the variance explanation possible for a given classification system.

For the index systems (management minutes and Maryland index), an index value was computed for each resident in each data set. A similar assignment was performed for the classification systems, using the developer-specified CMI for each group. (For the RUG-II and RUG-T18 systems, CMIs were available that described both nursing and total costs.) For each data-system pair, the  $R^2$  was computed for a linear regression, with the computed index value as the independent variable. This method implicitly incorporates a linear adjustment (i.e., multiplicative scaling and the addition of a constant) to achieve the best fit. We can recognize whether a system correctly estimates the ratio of resource use for different types of residents. The test is not affected if, because of staffing standards or practice patterns intrinsic in a given data set, the system routinely overestimates or underestimates actual resource use.

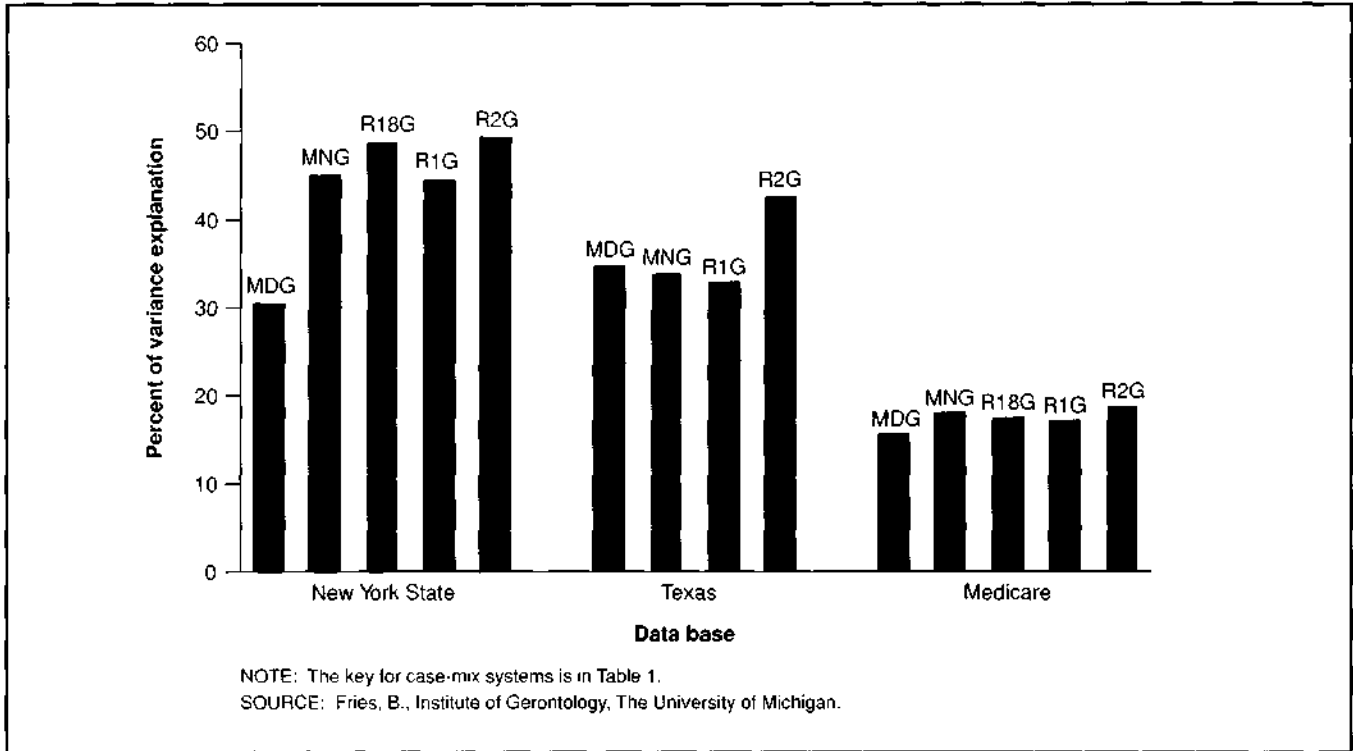
These analyses were performed for all systems and data sets, employing nursing time and cost and, when available, total time and cost as the dependent variables. The major results for nursing and total time, and total cost are summarized in Figures 2-7. (The same basic results were obtained for logarithmic transformations of the time- and cost-dependent variables.) In Figures 2-5, the variance explanation for a spectrum of systems across the three data sets is displayed. In Figures 6 and 7 (total cost), the same information is shown for only two data sets, as the Texas set did not have sufficient information to compute non-nursing costs. As discussed earlier, classification systems are considered both as grouping systems (using ANOVA) and for their numerical CMI values (using regression); this distinction is indicated in the labels on relevant bars. (Although the  $R^2$  statistics of the regression and ANOVAs technically cannot be statistically tested, I maintain that they provide insight into the relative merits of the systems compared.)

Overall, variance explanations of nursing costs were highest for the New York data, intermediate for the Texas data, and lowest for the Medicare data. The lower explanation for Medicare appears to result from substitution of rehabilitation for nursing services in the care of post-acute rehabilitation residents: If residents receive heavy rehabilitation, they simply are not on nursing units to receive nursing care (Fries et al., 1989). In fact, only a system that explicitly recognizes rehabilitation is able to have anything other than truly minimal variance reduction for total costs of Medicare residents. The RUG-II system validates reasonably well, and the RUG-T18 system, derived using Medicare data, does significantly better (Figures 6 and 7), principally because of its recognition of rehabilitation costs.

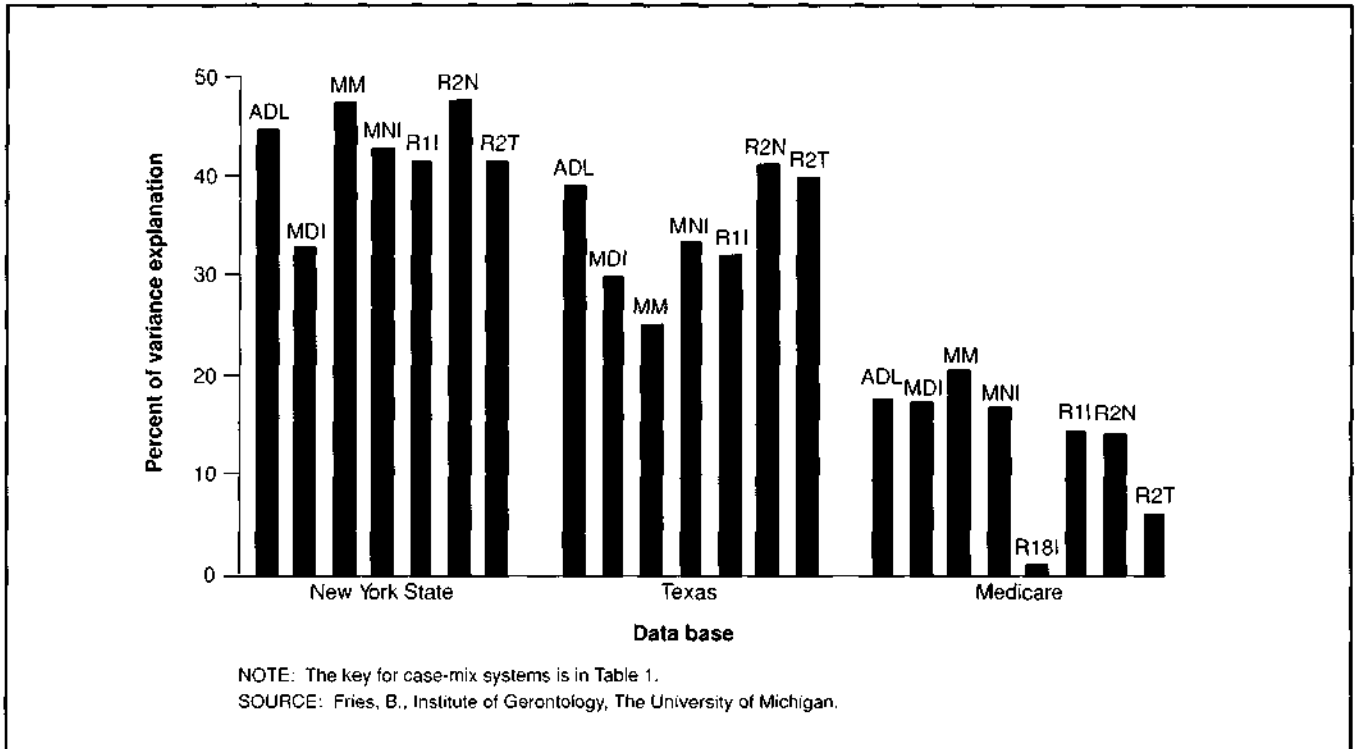
It is tempting to hypothesize that the intermediate values of variance explanation in Texas result from the lower staffing levels in that State, with more homogeneous care given to all residents and, thus, a poorer relationship between need for and provision of care. However, the coefficient of variation for nursing time is larger for the Texas data than for New York data, indicating less rather than more homogeneity. Clearly, the relationship among staffing, provision of rehabilitation,



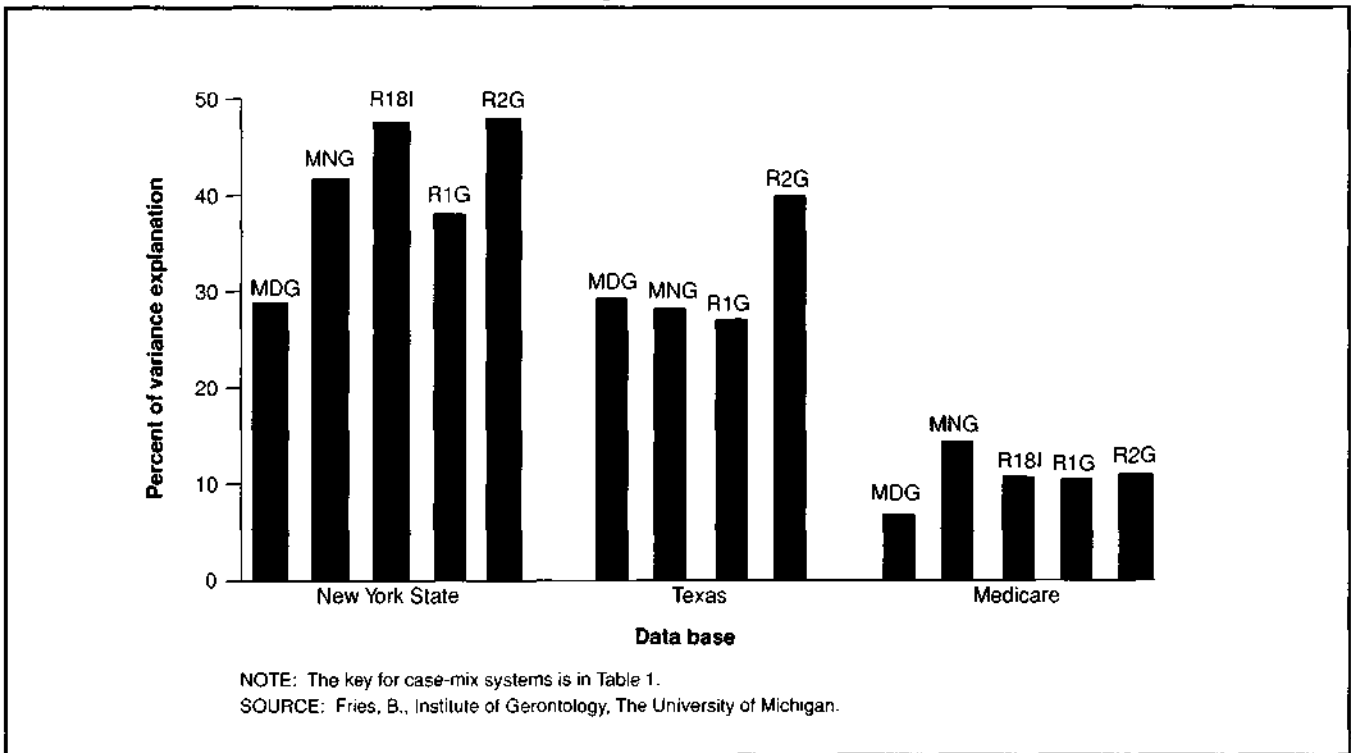
**Figure 2**  
**Percent of variance explanation of nursing time for selected classification case-mix systems for nursing homes, by data base**



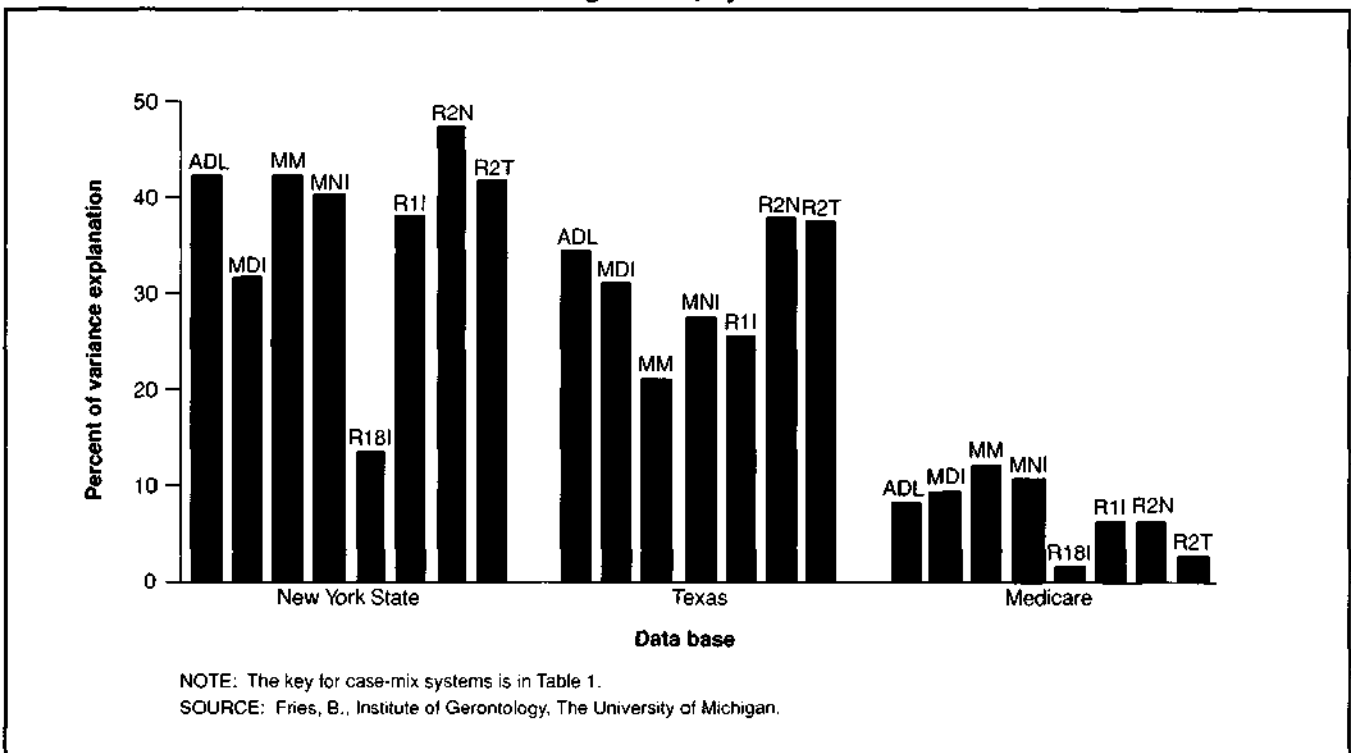
**Figure 3**  
**Percent of variance explanation of nursing time for selected index case-mix systems for nursing homes, by data base**



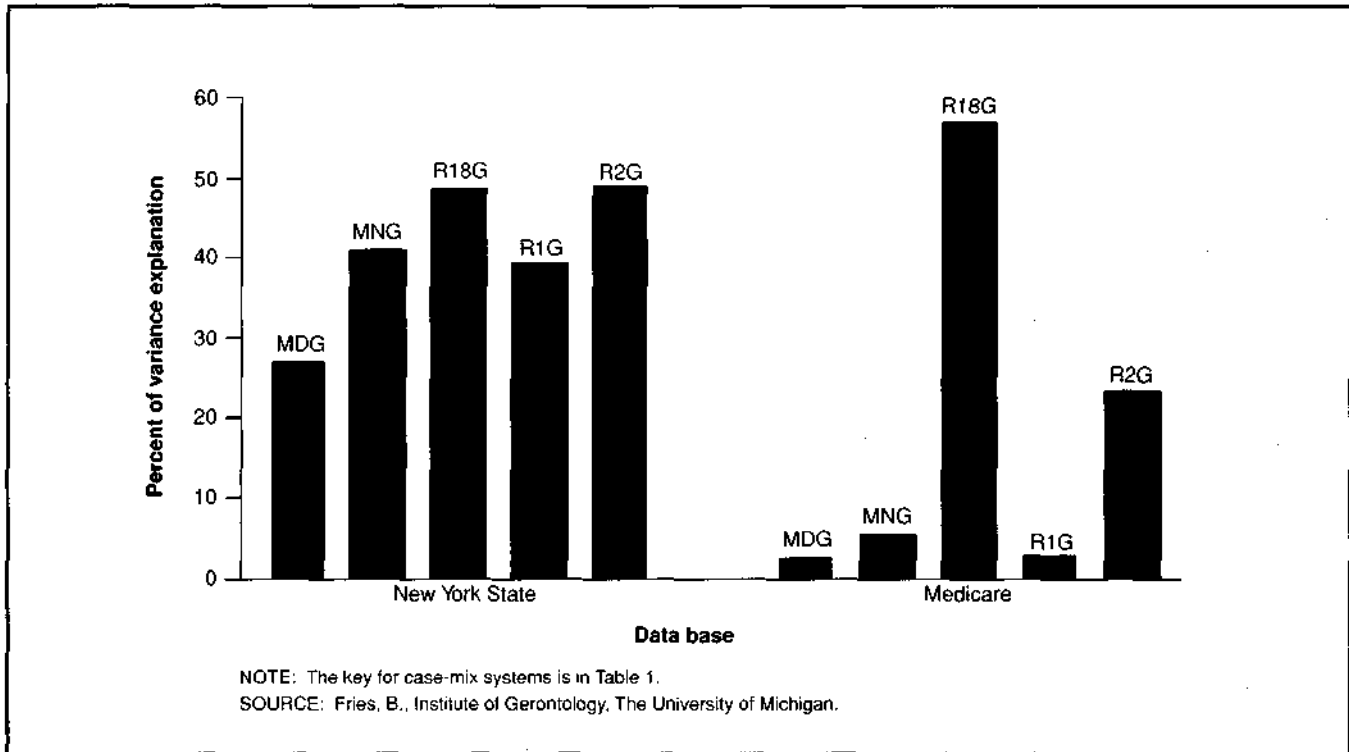
**Figure 4**  
**Percent of variance explanation of nursing cost for selected classification case-mix systems for nursing homes, by data base**



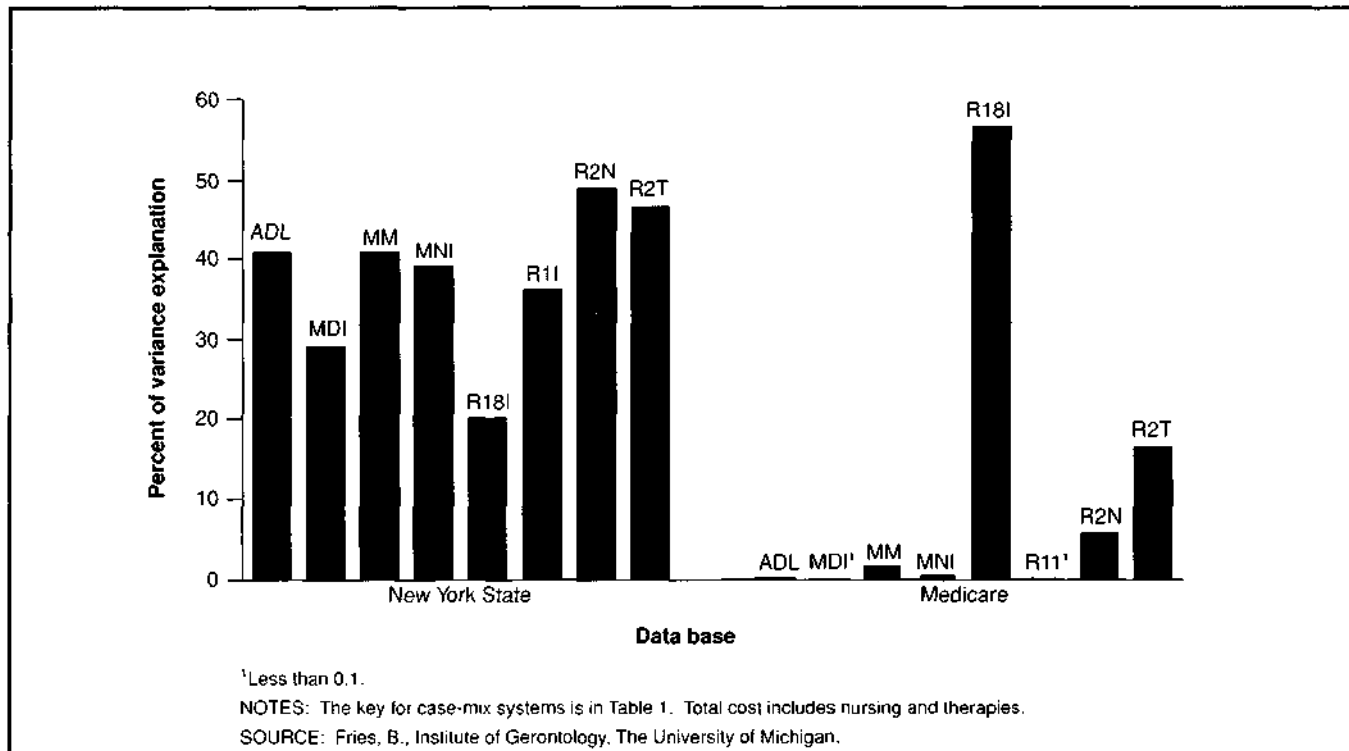
**Figure 5**  
**Percent of variance explanation of nursing cost for selected index case-mix systems for nursing homes, by data base**



**Figure 6**  
**Percent of variance explanation of total cost for selected classification case-mix systems for nursing homes in New York State and Medicare data bases**



**Figure 7**  
**Percent of variance explanation of total cost for selected index case-mix systems for nursing homes in New York State and Medicare data bases**



nursing time, and nursing cost is much more complex than may be initially expected.

A second general phenomenon was the decreased variance explanation of almost every system as we move sequentially through nursing time, nursing cost, total time, and total cost. Except for the RUG-II and RUG-T18 systems, all the systems tested were designed to explain nursing times. Thus, they were not structured to cope with the substitution effect between rehabilitation and nursing times seen, for example, in the Medicare sample.

I focus attention first on the variance explanation achievable by each classification system in defining groups and by the two index systems. (As I describe later, adjustments in the CMIs used to weight individual groups in a classification system often significantly affected the variance explanation.) For calibration as well as for a measure of the importance of the ADLs by themselves, I included results for the RUG-II ADL index, which is treated as a grouping system (i.e., a group for each level from 3 to 10). The codes for the systems tested, used in Figures 2-8, can be found in Table 1.

Except for two specific cases, for each of the three dependent variables in each of the three data sets, the RUG-II or RUG-T18 systems achieved the highest explanations—in many cases, by substantial margins. For the New York State data, the RUG-II system achieved 49.1 percent for nursing time (Figure 2) and only slightly less (47.9 percent and 48.8 percent, respectively) for the two available cost measures (Figures 4 and 6). However, it was on this data set that RUG-II was derived. For the Texas data, the RUG-II explanations were lower, but almost 10 percent higher than all others (except the ADL index, part of the RUG-II system) for both resource measures considered. On the Medicare data base, management minutes and Minnesota performed well in explaining nursing time and nursing cost, with the RUG-II and RUG-T18 systems (the latter developed on this data set) only slightly less effective (Figures 2-5).

**Table 1**

**Case-mix classification and index systems used in this analysis**

Code	System
ADL	Activity of daily living index used in resource utilization groups, version II
MDG	Maryland groups
MDI	Maryland index
MM	Management minutes
MNG	Minnesota groups
MNI	Minnesota index
R1G	Resource utilization groups, version I, groups
R1I	Resource utilization groups, version I, index
R2G	Resource utilization groups, version II, groups
R2N	Resource utilization groups, version II, index using nursing case-mix indexes
R2T	Resource utilization groups, version II, index using total case-mix indexes
R18G	Resource utilization groups—title 18, groups
R18I	Resource utilization groups—title 18, index using total case-mix indexes

SOURCE: Fries, B., Institute of Gerontology, The University of Michigan.

However, in this same data base, RUG-II excelled (23.3 percent) and RUG-T18 excelled even more (57.0 percent) compared with other systems in explaining total cost, mostly because of the ability of these systems to identify rehabilitation costs (Figures 6 and 7). Again, comparisons based on the Texas data, when available, are the least biased, as these data were not involved in the derivations of any of the systems considered.

It is useful to compare these variance explanations with those achieved by the ADL index. This index performed as well as and often better than many of the more complex systems. As a single example, for the Texas data, the ADL index not only was superior to all systems except RUG-II but also had variance explanation at least 5 percent, and in some cases as much as 13 percent, greater. One can conclude that these several systems do not employ adeptly the information contained in the ADLs. Also, the additional ADLs present in the Maryland, Minnesota, and management minutes systems, at least in the manner in which they are used, provide little to no improvement in variance explanation.

I next address the performance of the classification systems with the designed case-mix weights: Minnesota (MNI), RUG-II (R2N for nursing CMIs, R2T for total CMIs), and RUG-T18 (R18T). The variance explanation of the constructed index will perform no better than that obtained by ANOVA, which, in effect, fits optimal weights for the given data set. (There is no anomaly in the Maryland index, which utilizes more information about the resident than was used to form the groups.) The loss of explanatory power for many of the comparisons is slight, only a few percentage points. The largest inaccuracies occur, not surprisingly, when the indexes computed to allow for total (i.e., nursing plus other professional therapy) times or costs are applied only to nursing times or costs, or the reverse. This finding was confirmed in the analysis of residuals. For most classification systems, the CMIs appear to be reasonably accurate for predicting nursing times. This issue is discussed again later in an evaluation of each system's weights for individual categories.

In addition to comparing variance explanation, classification systems can be evaluated by comparing the homogeneity of the resource measure within classification groups. The coefficients of variation (CVs) for classification systems, data sets, and dependent variables are summarized in Table 2 by a simple average of the CVs. The results for this summary statistic mirror the results derived from distribution or highest and lowest values. For each combination of data and dependent variable, the RUG-II system proved superior (i.e., had a lower average CV) to the RUG-I, Minnesota, and Maryland systems. For example, using the Texas data and examining nursing times, it can be seen that the average CV for the 16 RUG-II groups was 0.47, compared with at least 0.53 for the other three systems.

**Accuracy for heavy-care residents**

One purpose of case-mix systems is to identify heavy-care residents and target increased payment for them.

**Table 2**

**Average coefficient of variation for measures of resource use by nursing home residents, by data set and case-mix system**

Data set and case-mix system	Nursing		Total	
	Time	Cost	Time	Cost
<b>New York State data set</b>	Coefficient of variation			
RUG-I	0.43	0.37	—	0.37
RUG-II	0.41	0.35	—	0.34
Minnesota	0.46	0.39	—	0.39
Maryland	0.46	0.39	—	0.39
<b>Texas data set</b>				
RUG-I	0.53	0.59	—	—
RUG-II	0.47	0.51	—	—
Minnesota	0.55	0.65	—	—
Maryland	0.53	0.60	—	—
<b>Medicare data set</b>				
RUG-I	0.44	0.47	0.79	0.54
RUG-II	0.41	0.43	0.59	0.45
Minnesota	0.43	0.43	0.77	0.52
Maryland	0.47	0.49	0.83	0.57

NOTES: RUG-I is resource utilization groups, version I. RUG-II is resource utilization groups, version II. Total time and cost include nursing and therapies.

SOURCE: Fries, B., Institute of Gerontology, The University of Michigan.

Aside from providing more equitable payment for facilities, this encourages the admission of heavy-care residents, who are often in hospital backlogs and are the most difficult patients to place. Measures of ADL functionality alone, and systems based primarily on them, explain resource use for the vast majority of nursing facility residents. However, although they appear to work well for the entire spectrum of residents, such systems may perform poorly for a small but vital segment of the population for which ADLs are inadequate. The standard general management “80-20” rule of thumb is close to being exactly applicable here: Most of the residents are well represented by a small portion of a system, and we need the rest of the classification system’s design to deal with the remaining few residents. In particular, in the RUG-II system in use in New York State, more than 70 percent of the residents fall into the reduced-physical-functioning group, which is subdivided into five groups using only ADL score. The remainder of the groups (four of the five major categories, including 11 RUG-II groups) are used to classify the balance of the residents (Schneider et al., 1988).

In a first test of this criterion, I examined the residuals of the fit between CMI and resource use. There was some indication of underrepresentation of the heavy-care resident, but concrete evidence was scarce. A superior overall test was to focus on the heaviest care resident in each sample. I fitted each system to the top quartile of residents for the selected dependent variable, using the same ANOVA and regression methods described earlier. Results for the variance explanation of nursing cost in the top quartile of residents for nursing cost in the three data sets, which are representative of my overall findings, are shown in Figure 8. The RUG-II system is clearly and uniformly superior, often achieving close to 50-percent

higher variance explanation than the second best system. For example, for the Medicare data, variance explanation was 38 percent for RUG-II groups versus 24 percent for RUG-I groups and approximately 21 percent for management minutes and Maryland groups. There was no uniform second-best system.

**Appropriateness of case-mix indexes**

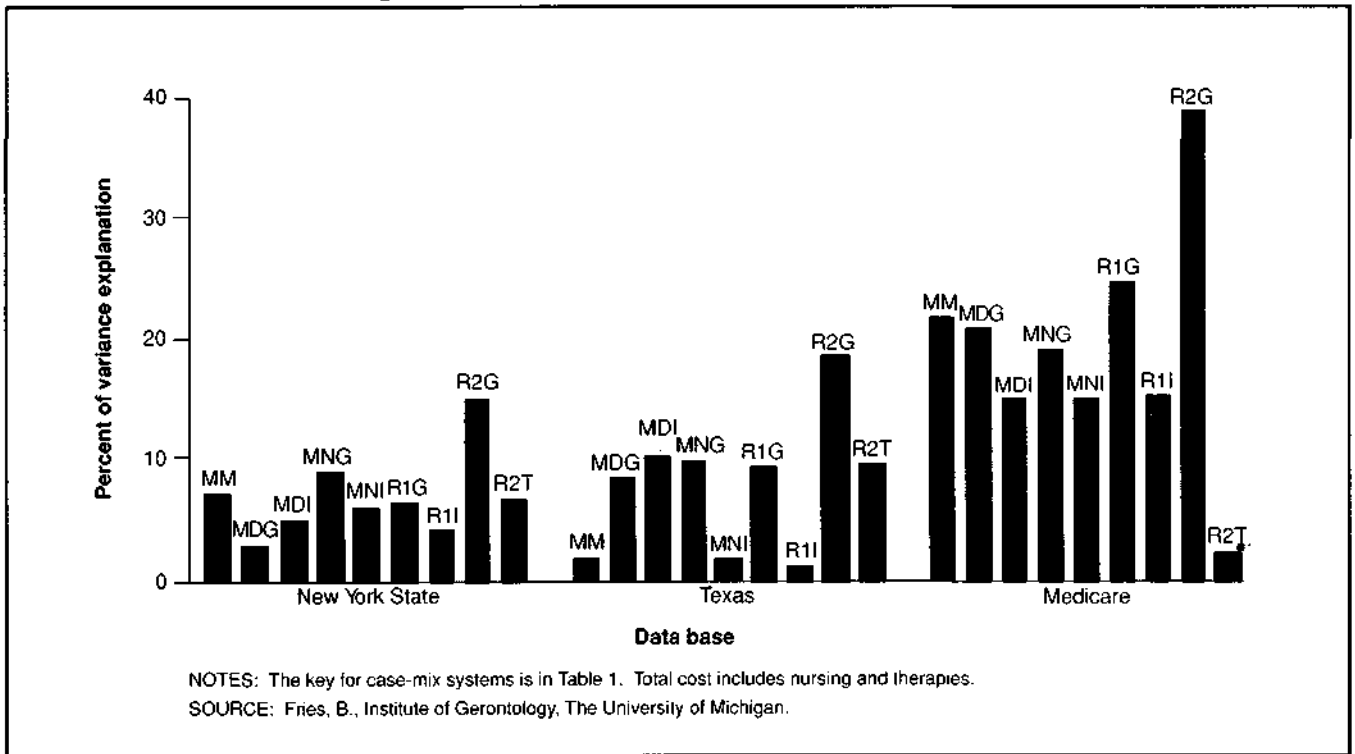
All payment systems that utilize case-mix measures employ them in relative rather than absolute terms. Typically, the average case mix for all residents is anchored at the average payment, and payments for residents or facilities with case mix higher or lower than this average value are adjusted according to the CMI. It follows that the accuracy of a classification system across the spectrum of residents can be evaluated by comparing the relative CMIs with the relative resource use. For equivalence and ease of display, both the CMI and relative resource measure can be normalized to an average value of 1.00 across all residents. Consider, for example, the application of the Minnesota system to New York data. The Minnesota CMI, which ranges from 1.00 for the lowest group, A, to 4.12 for the most resource-intensive group, K, was normalized so that the average normalized CMI was 1.00. For the New York data, the unadjusted Minnesota CMI was computed at 3.11, and each group’s CMI was divided by this constant to accomplish the normalization. Similarly, nursing time, nursing cost, and total cost were normalized by employing the mean value of each. The groups were then sorted by increasing CMI and the two variables plotted against each other (Figure 9).

A perfect relationship between the CMI and measures of resource cost would be represented by a straight line with positive slope, describing a relationship between increasing case mix and increasing actual resource use. Such a relationship could never be achieved because of the intrinsic variability in resource use even by the same residents (which especially affects groups with small numbers of residents) and the complex relationships among payment, staffing, resident needs, and the provision of service. On the other hand, bias in the original data used to compute the indexes or failure to identify a type of resident or characteristic will result in a deviation from a linear relationship between the normalized resource cost and CMI. The deviations from a fitted line were evaluated statistically at the 1-percent level, taking into account the variability intrinsic in each group. Although I display only two examples (Figures 9 and 10), I discuss here the results for three of the systems—RUG-I, RUG-II, and Minnesota—on the three data sets. The fourth classification system (Maryland) had only four groups and was therefore difficult to analyze.

In Figure 9, based on New York data, it can be seen that Minnesota Group F, those who have 4-6 ADL dependencies and require special nursing services, use significantly more resources than are attributed to them by the Minnesota CMI. Alternatively stated, the CMI for Group F is too low. Similarly, although to a lesser degree, Group I is undervalued by the CMI, and Groups E and H are overvalued. These findings hold for the three

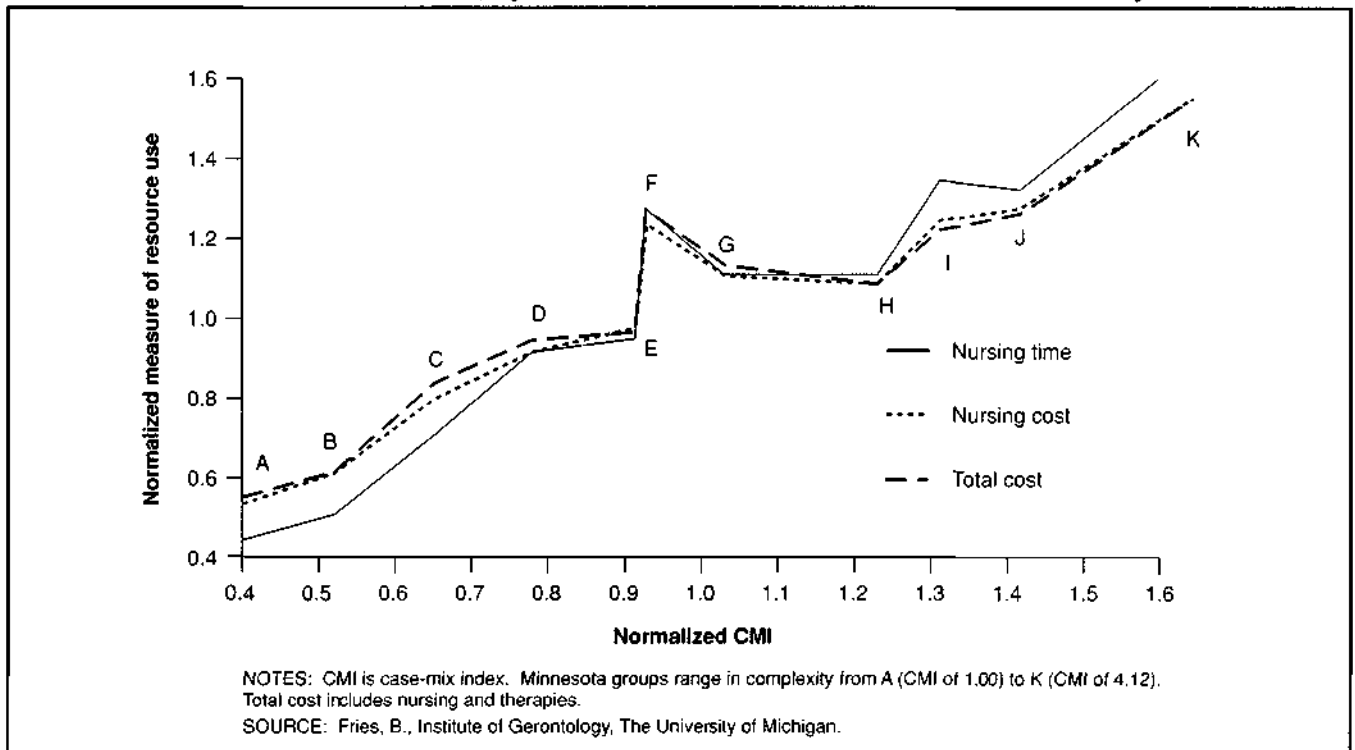
**Figure 8**

**Percent of variance explanation of nursing cost for nursing home residents in the top quartile of nursing costs for selected case-mix systems, by data bases**



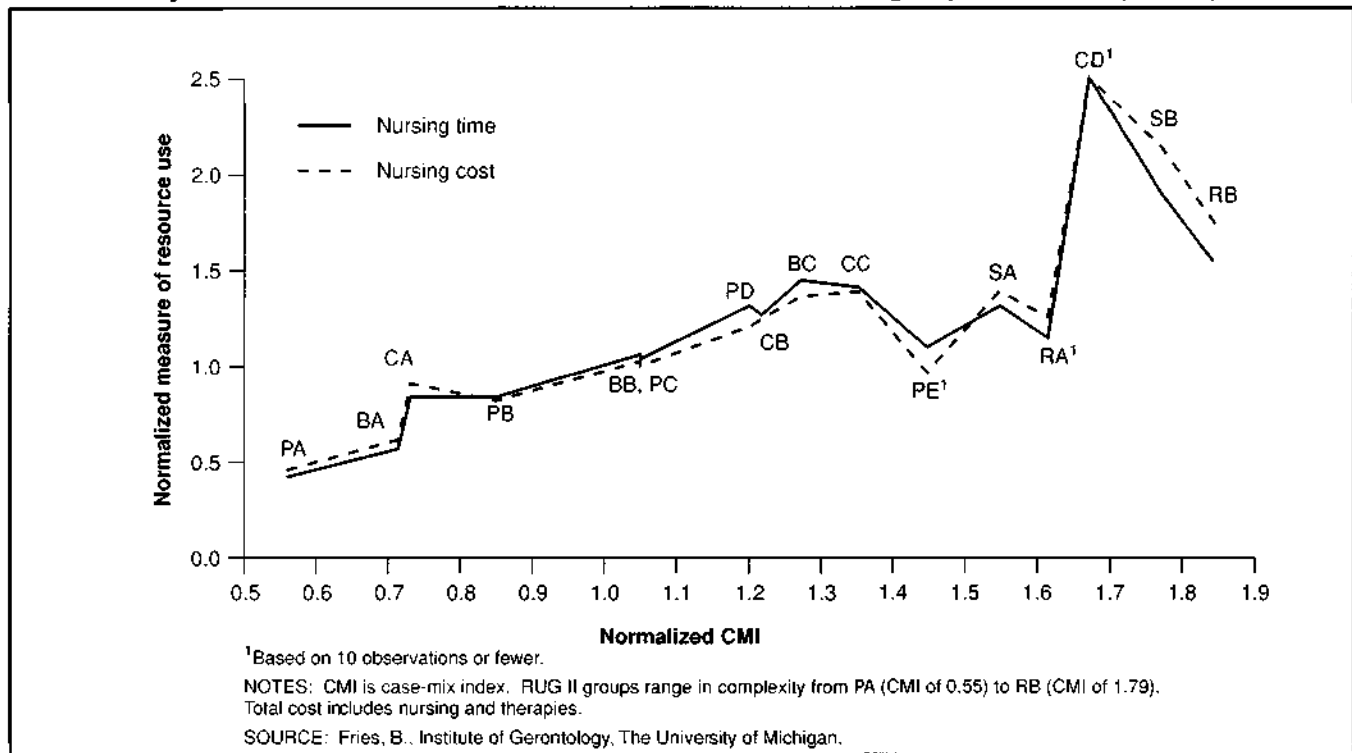
**Figure 9**

**Normalized measures of nursing time, nursing cost, and total cost for nursing home residents in New York State data base, by normalized case-mix index used in Minnesota system**



**Figure 10**

**Normalized measures of nursing time and nursing cost for nursing home residents in Texas data base, by normalized case-mix index used in resource utilization groups, version II (RUG-II)**



measures of resources (the three lines in the figure representing nursing time, nursing cost, and total cost) and, although not displayed here, across the three data sets.

The other two classification systems have similar discrepancies. In the RUG-I system, two of the moderately heavy-care groups have CMIs that appear to be too high (not shown). For the RUG-II system applied to Texas data (Figure 10), all but one of the groups on the right—those with the highest CMIs—are so rare as to prevent any conclusion from being drawn. However, the resources used by the clinically complex-A (CA) group are higher than predicted by the CMI. When used with the Medicare data, the strength of the relationship between CMI and resource use is poor, again a result of the complex relationship between rehabilitation and nursing time. Given the different structure of this part of the industry, for the RUG-II system (or its sister RUG-T18 system) to be used for Medicare residents, CMIs computed on Medicare data would have to be employed.

In general, the relationships between CMIs and normalized resource measures were superior for all classifications when evaluated utilizing the New York State and Texas data sets. The Medicare data displayed less coherent relationships. In part, this resulted from the focus on heavier (SNF) care among Medicare residents, resulting in small sample sizes in certain groups. Another factor is the strong negative interaction between rehabilitation and nursing, discussed earlier. I also found that, in cases of good concordance of CMIs with actual times, the concordance with costs was substantially

poorer, although the only difference between measures of times and costs was the cost weighting applied to each type of resource. It follows that users of classification systems (or, indeed, any of the resource measurement systems) for payment purposes, even if they accept a particular system without modification, should rederive CMIs using application-specific relative wage rates for different types of staff. These calculations are easily performed without extensive research effort. Wage rates would be multiplied by available values of resource use for registered nurses, aides, etc., to develop the most appropriate specification of CMIs.

One measure of the overall appropriateness of CMIs and the systems as a whole is how well they operate on a large population. To test this, I compared predictions of staff time across the three samples. First, each system was used to develop its own average CMI for each population. These average values are displayed in Table 3. Using the average CMI values computed for the New York data as a reference, we see that the management minutes system indicates that case mix in the Texas sample is 9 percent higher than New York's (203 compared with 186, as a percentage of the latter). Similarly, average values of nursing time were computed for each of the three data sets and compared with the New York values. The measured nursing time for Texas was lower than that for New York by 10 percent. The Medicare time, measured at 108 minutes for only the day and evening shifts, was 7 percent higher than the New York time. Assuming that the distribution of time across three shifts is the same as it was in the New York data, the Medicare time should be inflated by 27 percent.



resulting in a total 24-hour shift of 136 minutes, or an increase of 35 percent over the New York time (Table 3).

Only the RUG-II system predicted the decreased staff time for Texas, although it underestimated how much lower the Texas time was. (The Minnesota system also predicted less time for Texas, but the difference from the New York CMI was not statistically significant.) For the Medicare sample, all systems except Minnesota predicted the higher case mix. (Again, the RUG-I system's predicted increase was not statistically significant.) However, only the RUG-II system predicted the magnitude of this increase.

Although the collection of data on nursing time was performed relatively similarly in each data set, some concerns about the interpretation of these results remain. An argument can be made that, with the large and relatively similar resident populations in the New York and Texas data sets, any measurement of case mix should be relatively constant, with any aberrations in actual resource use being derived from exogenous variation in staffing as a result of State reimbursement levels, staffing decisions, or facility practice. In fact, the results are likely a combination of all of these factors, including case-mix differences resulting in different levels of actual resource use. Thus, in comparing the New York and Texas data, we may well be seeing differences in staffing without true differences in case mix. However, in comparing the New York and Texas data with the Medicare populations, we expect different types of residents and commensurate variations in resource use.

## Conclusions

Overall, the RUG-II system ranked the best of the systems tested. Its variance explanation was only slightly better than that of other systems, but it was consistent across data sources as well as across the comparison criteria employed. Only slightly less attractive are the management minutes and Minnesota systems, which

appear to be well designed and rate well on the multiple criteria evaluated. The management minutes system needs to be adjusted to better predict actual minutes, and some modification could improve its ability to identify heavy-care residents. Similarly, the Minnesota system would be strengthened if its groups were adjusted to better identify heavy-care residents and if the CMIs for particular groups were adjusted. On the other hand, some systems performed more poorly, including the Maryland and original RUG-I system. Both have since been updated. For all systems, recomputation of CMIs would improve their appropriateness in a future application.

The variety of analyses performed in this study underlie a basic assumption that the choice of a resource measurement system needs to be made with a variety of criteria, not only on the variance explanation most often touted in the literature. For example, a system might explain reasonably well the majority of residents, who primarily need maintenance care (perhaps by efficiently employing measures of ADL). However, such a system might seriously fail to recognize heavy-care residents or those with particular needs. When used in a payment system, such deficiencies, although not substantially affecting measures of variance explanation, may serve as disincentives for nursing homes to care for these residents. We suggest that a battery of tests, both qualitative and quantitative, need to be applied in evaluating competitive systems. As well, systems should be evaluated for their intrinsic incentives and disincentives (Smits, 1984; Schneider et al., 1988). The system that meets the goals for one State may need to be modified for another State's goals. Such a selection also needs to be based on the context in which the system will be employed. I have focused here on systems to be used for payment, wherein parsimony of variables is advantageous. Thus, individual characteristics are indicators and stand in for other variables that are highly correlated. For more clinical applications, a wider breadth of characteristics would be more appropriate.

**Table 3**  
Average case-mix index, by data set and case-mix system

Case-mix system or measure	Data set					
	New York average value	Texas		Medicare		
		Average value	Ratio <sup>1</sup>	Average value	Ratio <sup>1</sup>	
<b>Case-mix system CMI</b>		CMI	Percent	CMI	Percent	
Management minutes	186.40	**202.52	8.6	**213.50	14.5	
Maryland index	19.50	**22.79	16.9	**24.17	23.9	
Minnesota	2.46	2.44	-0.8	**2.29	-6.9	
RUG-I	3.11	*3.16	1.6	3.15	1.3	
RUG-II using nursing CMI	1.00	**0.97	-3.0	**1.26	26.0	
RUG-II using total CMI	1.02	**0.98	-3.9	**1.42	39.2	
<b>Nursing time</b>		Average time	Percent	Average time	Percent	
Measured time (in minutes)	100.93	**90.94	-9.9	**107.63	6.6	
Adjusted nursing time for 24 hours (in minutes)	100.93	**90.94	-9.9	**136.24	34.9	

\*Significantly different from New York State value at the  $p = 0.05$  level.

\*\*Significantly different from New York State value at the  $p = 0.01$  level.

<sup>1</sup>Difference between value for data set and for New York as a percentage of New York value.

NOTES: CMI is case-mix index. RUG-I is resource utilization groups, version I. RUG-II is resource utilization groups, version II.

SOURCE: Fries, B., Institute of Gerontology, The University of Michigan.

## Acknowledgments

The author would like to acknowledge the assistance of Jennifer Elston, who labored to perform all the crosswalks and more than a statistically significant portion of the computer analysis; the patience of Project Officer Dana Burley of the Health Care Financing Administration (HCFA); the support of Elizabeth Cornelius, also of HCFA, and her unflagging enthusiasm for advancing the art of long-term care; and those from New York, Minnesota, Maryland, and the Hillhaven Corporation who helped assure the accuracy of our crosswalks. In addition, I wish to thank Don Schneider of Rensselaer Polytechnic Institute and Robert Godbout of the Texas Department of Human Services for their advice during this study and three anonymous reviewers for their helpful suggestions on an earlier manuscript.

## References

- Arling, G., Nordquist, R.H., Brant, B.A., and Capitman, J.A.: Nursing home case mix: Patient classification by nursing resource use. *Medical Care* 25(1):9-19, Jan. 1987.
- Arling, G., Zimmerman, D., and Updike, L.: Nursing home case mix in Wisconsin. *Medical Care* 27(2):164-181, Feb. 1989.
- Birnbaum, H., Bishop, C., Lee, A.J., and Jensen, G.: Why do nursing home costs vary? The determinants of nursing home costs. *Medical Care* 19(11):1095-1107, Nov. 1981.
- Cameron, J.M.: Case mix and resource use in long-term care. *Medical Care* 23(4):296-309, Apr. 1985.
- Cavaiola, L.J., and Young, J.P.: An integrated system for patient assessment and classification and nurse staff allocation for long term care facilities. *Health Services Research* 15(3):281-306, 1980.
- Deane, R.T., and Cella, M.A.: New Concepts in Nursing Home Reimbursement. Working paper. Applied Management Sciences. Silver Spring, Md. 1981.
- Flagle, C.D., Young, J.P., McKnight, E., et al.: *Health Services in Long-Term Care*. Grant No. 5-R18-HS01250. Prepared for Public Health Service. Baltimore. The Johns Hopkins University and the Hospital Association of New York State. Nov. 1977.
- Fries, B.E., and Cooney, L.M.: Resource utilization groups: A patient classification system for long-term care. *Medical Care* 23(2):110-122, Feb. 1985.
- Fries, B.E., Schneider, D., Foley, W., et al.: Case-mix classification of Medicare residents in skilled nursing facilities: Resource utilization groups (RUG-T18). *Medical Care* 27(9):843-858, Sept. 1989.
- Grimaldi, P.L., and Jazwiecki, T.: *Case-Mix Payment Systems*. Chicago. Pluribus Press, 1987.
- Hillhaven Corporation: *William Thoms' Patient Care Profile—User Notes*. Tacoma, Wash. 1985.
- Holahan, J., and Sulvetta, M.: Medicare Reimbursement Policy for Skilled Nursing Facility Care: Issues and Options. Working Paper 3267-02. The Urban Institute. Washington, D.C. July 1987.
- Katz, S., Form, A.B., Moskowitz, R.W., et al.: Studies of illness in the aged: The index of ADL: A standardized measure of biological and psychosocial function. *Journal of the American Medical Association* 185(12):94-99, Sept. 21, 1963.
- Liu, K., Wiener, J., Schieber, G., and Doty, P.: The feasibility of using case mix and prospective payment for Medicare skilled nursing facilities. *Inquiry* 23:365-370, Winter 1986.
- Maryland Register*: 9(20):2002-2005, Oct. 1, 1982.
- McCaffree, K., Baker, J., and Perrin, E.B.: *Long Term Care Case-Mix Compared to Direct Care Time and Costs*. Seattle, Wash. Battelle Human Affairs Research Center, 1979.
- McCaffree, K., Winn, S., and Bennett, C.A.: *Final Report of Cost Data Reporting System for Nursing Home Care*. Seattle, Wash. Battelle Human Affairs Research Center, 1976.
- Minnesota Department of Health: Facility Manual for Completing Case Mix Requests for Classification. Unpublished manual. Minneapolis. May 1986.
- Morris, J.N., Sherwood, S., May, M.I., et al.: FRED<sup>2</sup>: An innovative approach to nursing home level-of-care assignments. *Health Services Research* 22(1):118-138, Apr. 1987.
- O'Donnell, J.F., and Hannan, E.L.: An analysis of case-intensity-adjusted nursing staff patterns in New York State skilled nursing facilities. *Medical Care* 21(1):82-91, Jan. 1983.
- Schneider, D., Fries, B.E., Desmond, M., et al.: Reimbursement Issues and Alternatives. Working paper. New York State Department of Health, New York State Long Term Care Case Mix Reimbursement Project. Albany, N.Y. 1985.
- Schneider, D.P., Fries, B.E., Foley, W.J., et al.: Case mix for nursing home payment: Resource utilization groups, version II. *Health Care Financing Review*. 1988 Annual Supplement. HCFA Pub. No. 03275. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Dec. 1988.
- Smits, H.L.: Incentives in case-mix measures for long-term care. *Health Care Financing Review*. Vol. 6, No. 2. HCFA Pub. No. 03195. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Winter 1984.
- Sulvetta, M.B., and Holahan, J.: Cost and case-mix differences between hospital-based and freestanding nursing homes. *Health Care Financing Review*. Vol. 7, No. 3. HCFA Pub. No. 03222. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Spring 1986.
- Texas Department of Human Services: *Texas Nursing Home Case Mix Project*. Continuation application for award of Grant No. 11-C-98688/02. Prepared for Health Care Financing Administration. Austin, Tex. Dec. 1986.
- Weissert, W.G., Scanlon, W.J., Wan, T.T.H., and Skinner, D.E.: Care for the chronically ill: Nursing home incentive payment experiment. *Health Care Financing Review*. Vol. 5, No. 2. HCFA Pub. No. 03168. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Winter 1983.
- Winn, S.: Assessment of cost related characteristics and conditions of long-term care patients. *Inquiry* 12:344-353, Dec. 1975.