Case mix for nursing home payment: Resource utilization groups, version II

by Don P. Schneider, Brant E. Fries, William J. Foley, Marilyn Desmond, and William J. Gormley

A study of 3,427 nursing home residents in New York State, measuring both resources used and resident characteristics, was used to develop a resident classification system for payment purposes. The system balances clinical, statistical, and administrative criteria, making it useful both for the New York State Medicaid payment system and for quality of care and facility management.

Introduction

The majority of systems currently employed to pay for institutional long-term care do not explicitly recognize the differences between residents in the amount of resources that are utilized in their care. Without such recognition, payment systems produce undesired incentives for providers to care for the least disabled residents. Nursing homes are implicitly encouraged to admit light-care residents and to delay discharge of those with the lowest care needs to home- or community-based care. Moreover, facilities that serve those residents requiring more intensive care either are reimbursed for their costs or are paid an average amount based on their cost experience pooled with the costs of other nursing homes.

Both cost-based and flat rate (one price for all residents) systems are inequitable for an individual nursing home. Cost-based reimbursement encourages nursing homes to expend money with the full expectation that whether such expenditures are effective, efficient, or even justifiable in the care of residents, they will become part of future cost bases and thereby increase future reimbursements. Under cost-based reimbursement, efficient management systems can be replaced by strategies to increase costs up to an imposed ceiling; those facilities that do this best, and have done so for years, will fare the best regardless of the types of residents for whom they care. On the other hand, flat-rate systems pay all nursing homes in a class or statewide a constant per diem rate, without considering their resident populations.

In the development of a new Medicaid payment system for New York State nursing homes, we developed a methodology that would directly recognize differences between residents in the quantity and thereby the cost of the care they received. Such a "case-mix" resident classification system would remove the incentive to avoid admitting heavy-care residents. In fact, it permits design of payment systems that are even-handed, appropriately paying

for care of all residents at a level commensurate with the cost of their care.

A key step in the development of the payment system was the construction of a method to measure the case mix of nursing homes. We discuss here that effort; the issues of moving from a case-mix measurement system to a case-mix payment system are addressed in other work we have done (Schneider, Fries, Desmond et al., 1985). We strongly advocate separating consideration of the issues of case-mix measurement from those of case-mix payment. Although considerable attention is required to construct a measurement system applicable to case-mix payment, this work can be performed prior to payment design. Case-mix measurement is primarily a scientific issue: how to determine those characteristics that relate to the cost of care. On the other hand, payment design is more a political process, designed around the goals and realities of providers and payers. Confounding the two efforts complicates the process unnecessarily. Nevertheless, excellent case-mix payment systems are designed around excellent case-mix measurement systems. We thus began our development by considering methods to measure the resource use—the case mix—of residents in nursing homes.

Background

There have been two approaches to developing case mix for nursing homes, as there have been, although earlier, for hospitals (Fetter et al., 1980). In the earlier approach, a variety of characteristics about nursing homes were used to explain facility cost. These variables usually included facility size, ownership, level of care, payer mix, and so forth, and at times summary measures of facility case mix such as the percent of residents who are ambulatory. This resulted in a nominative relationship which was then used as a measure of the facility's case mix. An example of this was the system previously in place in New York, described by O'Donnell and Hannon (1983). Eleven such studies were reviewed by Birnbaum et al. (1981).

We have avoided the facility-related approach for a variety of reasons. First, such an approach does not recognize that it is the care needs of individual residents that are the major determinants of cost and thus does not provide inherent and explicit incentives to accept heavier-care residents. The variables given in the previous paragraph, for example of facility payer

mix as a predictor of cost and thereby a case-mix surrogate, are representative of this type of problem. Second, many of the measures that might be used are those cost-related items that might well be controlled. Thus, their inclusion in a rate-setting formula is a sophisticated, indirect return to cost reimbursement rather than a prospective rate-setting approach. Third, nursing homes would still have to cope with differences in the intensity of care required by their residents, and these differences over time or across facilities would not be recognized outside of those correlated with facility characteristics.

The second generic approach to define case mix solves these problems by directly examining the residents in a facility. The logic for this approach is based on the industrial engineering concept of the "product" of a firm. If the types of residents in a facility or the level of their needs can be specified, then a facility's case mix is the amalgamation of these measures for all its residents. The best known application of this approach is the diagnosis-related groups (DRG's) system for acute care hospitals.

In long-term care there have been several resident-level case-mix measure and resident classification systems suggested. Some of the work we considered included systems developed by the Battelle Human Affairs Research Centers (McCaffree, Winn, and Bennett, 1976; McCaffree, Baker, and Perrin, 1979; Cavaiola and Young, 1980), and the State of Maryland (Deane and Cella, 1981), and the earlier version of the resource utilization groups (Fries and Cooney, 1985). Examples of these approaches are well documented in the report to the Health Care Financing Administration by Stassen and Bishop (1983). Other systems, for example those used for reimbursement determination in Illinois, Ohio, and West Virginia, define case mix at the resident level, but principally utilized services in this determination; we discuss later our concerns with service-based approaches. During or subsequent to the development work described here, other systems have been developed by Weissert et al. (1983), Cameron (1985), the Minnesota Department of Health (1986), Morris et al. (1986), and Arling et al. (1987).

Conceptual approach

The goal of the current work was to develop a resident classification system consisting of resident groups, defined by resident characteristics, which would be predictive of nursing home resource use and usable in a payment system. We denote the system derived "resource utilization groups, version II," or RUG-II. Each part of this goal statement had important implications for the development effort.

First, we sought a "classification" or "grouping" system; an alternate approach would lead to an "index" or "rating" system in which a number representing relative resource use is assigned to each

resident directly.[1] Advantages exist to each approach on conceptual as well as technical grounds. Index (rating) systems can make conceptual sense if they add the times spent caring for particular resident conditions. Such systems are most often derived through regression or discriminant analysis. Grouping systems conceptually follow categorization approaches often seen in biology and medicine. They are derived through analysis of variance and allied statistical techniques. We have selected this latter approach because of its adeptness in handling interactions, i.e., the synergistic effects of two or more resident conditions on resource use. For example, we believe that a single condition may not lead to high resource use, but may in combination with a second.

Following the grouping approach, one can envision a range of possible systems which on the one extreme classifies every resident into a separate group and on the other has all residents in the same group. Both extremes are inappropriate: the former, although recognizing the differences always seen between any pair of residents, provides no understanding of resource use; the latter, similar to most current reimbursement systems, provides little or no differentiation of residents. Most current systems recognize only the difference between intermediate care facility (ICF) and skilled nursing facility (SNF) residents. We sought a system between these two extremes, identifying clusters of residents who have relatively homogeneous resource use. Clearly, with a limited number of groups there remains substantial differences between residents within each group, but the groups are relatively homogeneous and different from each other.

Second, we considered only systems that employed *resident* characteristics. Based on the work of others and our own observations, significant explanation of resource could be expected by recognizing the type of facility (SNF versus ICF), ownership (proprietary, voluntary, governmental), organization (freestanding versus hospital-based), Medicaid occupancy, and so forth (Bishop, Plough, and Willemain, 1980; Shaughnessy et al., 1985). Based on the logic discussed in the previous section, all of these possible defining conditions were considered inappropriate: they did not refer to residents' characteristics, but rather to where residents were and the services they were provided.

Third, we wished to predict actual resource use, at the level of the individual resident. A potential model for this development, the diagnosis-related groups (DRG's), used for case-mix measurement in the acute care system and the basis for Medicare's prospective payment system predict resource use for an entire episode of care. However, for long-term care, lengths of stay—and thereby episode costs—are extremely variable, with similar residents institutionalized for days, weeks, or multiple years. Although paying for an episode would be technically feasible and perhaps

---

[1]It should be noted that the RUG-II system utilizes an (activities of daily living) index, but only to define patient groups; thus, this system remains one of the "grouping" type.

appropriate for financing at the aggregated level of a large population, practically, and at the level of an individual nursing facility, this high variability makes it necessary to consider per diem payment systems. The effort here, therefore, focuses on explaining those costs specifically associated with the care of a resident in a nursing home for an individual day.

Finally, operational implementation in a payment system is feasible only if the case-mix measure is based on characteristics that are valid, measurable, and reliable, and that can be audited. For example, we found that a resident's functional capability to bathe was highly associated with a facility's practice patterns: In some, residents taking a bath would have active staff assistance regardless of their functional capability. It follows that bathing was not a valid resident characteristic, even when it could be measured reliably.

The overall goal led to a set of specific criteria for system development. These criteria form three major categories: statistical, clinical, and administrative.

Statistical criteria

The major statistical criterion used was a measure of how many of the differences in resource use between nursing home residents could be explained by the resident groups; this proportion is denoted "variance explanation." For example, the pilot RUG system (Fries and Cooney, 1985), consisting of 9 groups, achieved a variance explanation of 37.8 percent—37.8 percent of the differences in the dependent variable were explained by this system—an initial target for the current development. For comparison purposes, the DRG system, when applied to all patients in acute care hospitals, has a variance reduction of approximately 30 percent (Cretin and Worthman, 1986), although we caution against the direct comparison of these numbers. Other statistical criteria included measures of the homogeneity of the final groups, as measured by their coefficient of variation (the group standard deviation divided by the group mean) and measures of the differences between group means.

Clinical criteria

A system that makes clinical as well as statistical sense is more likely to be accepted and used in clinical environments. We defined "clinical sense" to mean that residents within a category had clinical affinity. In our development effort, we were assisted by a technical panel including clinicians (physicians, nurses, therapists, etc.) as well as administrators, researchers, and health care policy analysts. The charge to this panel was to help design possible taxonomies of residents and new combinations of variables to be tried and to assist in understanding analytic results.

Initially, we expected that this panel would assist solely in providing a clinical reality to the derived system. We anticipated that their suggestions, provided interactively with our analyses, would lead

to an admixture of new possible approaches and significant numbers of invalid hypotheses. Although these both did occur, we subsequently found that this clinical reality provided us with new insight that led to a superior system on statistical grounds as well. This result by itself is important for future studies, emphasizing the need to combine statistical and clinical criteria.

Administrative criteria and incentives

Classification systems often are not appropriate for the particular context in which they might be employed. For payment applications, often the problems involve the provision of inappropriate provider incentives. One such is an incentive to "game" the system: to develop for a resident a characteristic at little cost to the provider that will result in increased case-mix payment. An example can be drawn from the earlier version of the RUG's (Fries and Cooney, 1985), a system developed for retrospective analysis of the differences between nursing homes' costs. This system used the monitoring of intake and output of fluid to differentiate the two highest care groups. In addition to making clinical sense, fluid monitoring acted as a surrogate (later we use the term "indicator" for such a variable) for the bundle of needs of the sickliest nursing home residents. However, for prospective payment, a system based in part on I/O monitoring would provide a significant incentive for facilities to game by placing relatively ill residents on fluid monitoring, especially because monitoring would cause little harm to the resident.

Some variables that can be gamed provide appropriate incentives, such as for the provision of rehabilitation therapies. In the system we developed, rehabilitation services were explicitly included to provide incentives for their expanded use. Other services such as bowel and bladder training and psychiatric care are required to support the claims of particular resident characteristics (incontinence and behavioral problems, respectively). The choice of which variables to include or exclude became a major issue in the derivation process.

The incentive issues of utilizing measures of service as indicators in a case-mix system are complex. Where ever possible, we attempt to use measures of the need for a service rather than the provision of the service itself. However, in many cases one cannot avoid such variables, as need cannot be well predicted. The most obvious example is the use of surgical procedures in the DRG system. On the other hand, uncontrolled use can lead to heavily service-based systems such as those employed for nursing home payment in Illinois, Ohio, and West Virginia, with significant potential for gaming. Neither choice is right or wrong.

We suggest that wherever possible one should avoid service variables unless they meet at least some of the following criteria:

• The extra cost of providing the service offsets a significant portion of the increased payment.

- The service has the potential of serious negative consequences to a resident; again, this makes it unlikely to be gamed.
- There are no other ways to predict the need for this service.

The stronger the effect of service variables in a system, the greater the need to assure the appropriate provision of these services outside of the associated payment system, for example by utilization review.

The application of a resident classification system for prospective payment thus places additional requirements on the choice of variables. In addition to being relatively immune to inappropriate gaming, the variables must lend themselves to being measured validly and reliably, and effective auditing must be possible. Also, the variables must provide incentives for appropriate care, where possible.

Methods

Data collection

The derivation of the resident classification system necessitated data collection on a sample of nursing home residents to obtain objective measures of resource use and a broad spectrum of resident characteristics.

The major methodological difficulty in data collection was the measurement of the resources—primarily staff time—used in the care of residents. We rejected an observer time-and-motion study in favor of one employing self-reporting by staff, which would permit efficient data collection on a large number of residents. This required a variety of techniques to assure accurate measurement. Staff members—registered nurses (RN's), licensed practical nurses (LPN's), and aides and orderlies—kept personal records of all time spent with all residents under their care or on more general "unit activities." All time spent with or for a resident was associated with that resident, including time spent on actual care or monitoring, charting, discussing the resident in a meeting or on the telephone with a physician, and so forth. Only the time itself was collected; we did not collect either the number or duration of individual tasks, as others have done (e.g., Cameron, 1985). All other "non-resident-centered" time was assigned to major activities in support of the ward, including meetings, generic charting, restocking supplies, ward maintenance, or for breaks or meals (Table 1).

The major problem we addressed was to assure that all time was allocated. The total time available by the practitioner should be approximately equal to the time attributed to residents plus unit activities. At least at the end of the shift, all staff members reconciled the time documented with the duration of the shift; the two had to be within 30 minutes. Any anomalies were addressed in a structured manner considering activities performed, residents cared for, or special events, with the assistance of senior staff from the facility. We

believe that this method of data collection is both feasible and effective in developing accurate measurement of resource use in health facilities.

As our time measurement required staff to account for all of their time, we enrolled entire resident care units in our data collection, with all full-time staff participating during a 24-hour collection period. For staff who rarely had contact with residents in the unit, including those involved in physical, occupational, or recreational therapy (PT, OT, RT), speech, or inhalation therapy, psychotherapy, audiology, or social services, we collected treatment time via a week-long log for each resident. Of these services, only the first three—PT, OT, and RT—along with social services were provided to more than a very few residents. The others, provided to less than 0.4 percent of the residents, were eventually not considered in the analysis, although it does not follow that either the resident dependencies requiring such services or the payment for such services are omitted from the derived system.

We also collected ancillary charges, including pharmacy, physician services, central services supplies, and laboratory tests for a sample of residents. Of these, only pharmacy represented significant costs (4 percent of all reported direct resident care costs) and was tested in the analysis. Again, although the system was derived without these costs, many of them can be expected to be related to other resident service needs and, as such, are implicitly included.

Residents were assessed by using an instrument that listed a broad spectrum of characteristics. The major sections included administrative and resident demographics (gender, age, admission date, primary reason for admission, etc.); medical conditions (diagnoses, hospital DRG, level and number of decubiti, vision, hearing, edema, severe pain, contractures, etc.); activities of daily living (ADL's)—a resident's functional ability in bathing, eating, mobility, transferring from bed to chair, toileting, dressing and personal hygiene—and continence;

Table 1

Percent distribution of reported staff time in skilled nursing facilities

Time classification	Registered nurses	Licensed practical nurses	Aides
Total	100	100	100
Resident centered	38	48	58
Unit activities	62	52	42
Meetings and discussion	7	3	7
Documentation and auditing	21	16	[1]4
Administration	12	6	([2])
Ward maintenance, supplies, and routine activities	8	15	17
Off unit	6	4	4
Meals and breaks	8	8	10

[1]For aides, this category represented writing notes and reports.
[2]Not applicable.

psycho-social and behavioral conditions (regression, physical aggression, verbally abusive, withdrawn, delusions, etc.); services and conditions (therapies, restraints, medications, etc.); and availability of social support. In total, 195 items were included. The instrument was carefully designed to reduce ambiguity in assessment, including specifying time periods and frequency (e.g., ADL's were those applicable 60 percent of the time for the previous 2 weeks) and providing instructions and definitions on the form itself. The instrument itself was reviewed by more than 150 professionals and pretested and checked for reliability in pilot nursing homes.

The sample of New York State nursing homes chosen for data collection was selected to assure that all types of residents were seen, rather than to attempt to randomize over all types of nursing homes. In particular, we sought to oversample heavy-care residents. These residents, relatively rarer in the population, needed to be obtained in numbers sufficient for statistical recognition in separate payment categories. All facilities chosen were screened to meet quality of care standards. A system based on care patterns in substandard nursing homes would have little use. As well, facilities in the top or bottom 10 percentiles of staffing were excluded. The sample of homes was stratified by geographic location, sponsorship, hospital-based versus freestanding, size, case-mix intensity (as measured by the 1983 New York State reimbursement methodology), and single level status (ICF or SNF) versus combined. From one to three units of each selected facility—and thereby all residents in these units—were involved. In total, we collected data on 3,427 residents in 52 facilities.

We developed a variety of methodologies to address potential problems of a decentralized data collection effort. These included: recruiting facility staff as reviewers of data quality; involving a limited number of assessors; training of all data collectors and data quality reviewers; pretesting all methods in nursing homes immediately before actual data collection; careful timing of all data collection and review; internal reviewing of all data and reassessment of 10 percent of the residents; returning computerized edits rapidly to facilities for resolution; on-site project staff reviewing of all data and computer-generated edits; and closely communicating via a "hot-line" with all collection, review, and other project personnel. Based on problems noted when the data were reviewed, two nursing homes were asked to recollect their data, which they subsequently did. Of the final data received, 22 residents were subsequently excluded because of unusual events, such as a resident not being in the unit the day the staff time was collected, part of data was still missing, etc. (Schneider et al., 1987).

Analysis

The principal statistical technique we employed in the derivation of a classification was AUTOGRP (Mills et al., 1976), an interactive implementation of

AID (automatic interactions detection) (Morgan and Sonquist, 1963) and used in the original pilot RUG study and in the formation of the DRG's. In AID clustering, the full set of data points is recursively partitioned into subgroups by a set of splits, each split based on the values of a particular independent variable, the resident characteristics. The primary statistical criterion is that partitions be chosen to maximize the prediction of the dependent variable, resource cost (Fetter et al., 1980).

In addition to AUTOGRP, we used statistical tools including discriminant and outlier analysis to identify alternative constructs that might be tested. Some variables that had good predictability, however, were not usable for case mix determination as they did not meet our administrative criteria or provided poor incentives. In some cases, we were able to find alternate variables that provided equivalent variance explanation. These methods were augmented by the suggestions derived interactively with the project's technical panel on new variables, alternative strategies for initial splits, and alternative interpretation of results.

Several dependent variables were used in the analysis. Initially, analysis of reported direct times by staff and the simple sum of these times led to the observation that residents with certain selected but not highly prevalent conditions received a much different mix of resources than their cohorts. The differences were often in the use of skilled nursing services, for example as provided by registered nurses. These findings were used to identify clinically homogeneous types of residents and proved useful in developing the clinical sense of the system.

Two composite dependent variables were used for the majority of the analysis: wage-weighted nursing staff time and wage-weighted times of nursing plus other professional staff. To develop staff costs, for each resident we first summed the times for three shifts for RN's, LPN's, and aides. Non-resident-centered times represented a significant portion of the time spent by all staff, ranging from 62 percent for RN's to 42 percent for aides (Table 1). These unit times were allocated to residents either proportionally to their staff time (for documentation and meetings) or equally to all unit residents (for ward maintenance, breaks, and meals). Times were wage-weighted using constants representing the facility-specific relative wage rate for each type of staff, then summed across types of staff. (We later found that virtually identical results would have been obtained by using statewide average wages.) The second dependent variable augmented the first with PT, OT, RT, and social service times, again wage-weighted.

Results

The variables that, by themselves, were most effective in explaining differences in resource use in the sample population are given in Table 2. In conformity with almost every study of nursing home resource use, the ADL's provided some of the highest

Table 2

Single variables with highest percent variance[1] explanation and number of groups

Variable	Percent explanation	Number of groups
Toileting[2]	32.4	4
Transfer[2]	31.6	6
Dressing[2]	31.5	4
Mobility[2]	29.5	4
Eating[2]	27.9	5
Personal hygiene[2]	26.8	4
Bed rails	23.6	3
Bladder control[2]	23.3	4
Bathing[2]	21.1	3
Bowel control[2]	19.9	3
Primary impediment to discharge	17.0	3
Motivation	15.8	4
Decubitus level	15.2	3
Frequency of medical doctor review of resident	15.0	3
Frequency of medical doctor assessment of resident	13.9	3
Diet	13.8	3
Refusal to care for self	13.0	2
Learning ability	11.2	3
Expressive in communication	11.1	3
Receptive in communication	10.2	3
Number of decubiti	8.5	2

[1]All other variables had variance explanation of less than 8.5 percent.
[2]Activities of daily living variable.

variance explanations. Alternately expressed, if one seeks a single characteristic of nursing home residents to understand the level of resources involved in their care, an ADL measure, and almost any one, would be the best choice. All nine ADL measures included in the data collection (from toileting to bowel control, excluding bed rails) were virtually exclusively the most explanatory variables (Table 2).

When we followed an algorithm of maximizing variance explanation in each choice of a variable, we began by differentiating residents on a single ADL, then taking another ADL to form secondary splits, then a third. Such a strategy led to many groups with substantially decreasing marginal variance explanation at each level; after three splits we observed almost no additional variance explanation. Thus, the ADL's are highly redundant. Once we knew three ADL's we knew as much regarding cost variation as if we knew all of them. Second, which three ADL's we picked did not demonstrably change the total variance reduction. Given such interchangeability, we were able to choose the ADL's that were the best on nonstatistical criteria, for example, those providing incentives for better care. Finally, we found that significant information was obtained by recognizing intermediate levels for ADL's rather than using them only dichotomously, for example, as in the Maryland case-mix system.

Activities of daily living measure

These results encouraged the development of a summary measure of functionality that would compactly represent the influence of all of the ADL's. We chose three ADL's that were less easily gamed,

that were less affected by facility practice, and that provided incentives for quality of care. For example, eating is relatively difficult to game, given the time pressures on all staff at meal times. On the other hand, bathing was excluded because of the link, mentioned earlier, between bathing and facility practice. Toileting provided the opportunity to encourage better care by recognizing the additional resources needed to toilet rather than diaper an incontinent resident. The three ADL's chosen, toileting, eating, and transferring, were clinically judged to measure small and large motor skills as well as, implicitly, mental and physical functioning.

On the basis of variance explanation and clinical judgment, we determined that three levels would suffice for toileting and transfer with a fourth level (nasogastric tube feeding or parenteral feeding) for eating (Table 3). Considerable analysis of alternate formulation demonstrated that combining these three recoded ADL variables by simply adding their level scores resulted in an index that had a variance reduction of 35.4 percent, close to a computed optimal theoretic limit of 36.6 percent. The

Table 3

Construction of the resource utilization groups, version II (RUG-II) activities of daily living (ADL) index

Activity of daily living	Resident functioning level	ADL score
Toileting	Independent or minimal supervision and/or physical assistance	1
	Continuous supervision and/or physical assistance or total assistance or incontinent, does not use toilet	2
	Incontinent, taken to toilet on a regular schedule	3
Eating	Independent or minimal supervision and/or physical assistance	1
	Continuous supervision and/or physical assistance	2
	Hand fed	3
	Tube or parenteral feeding	4
Transfer[1]	Independent or minimal supervision and/or physical assistance	1
	Continuous supervision or continuous physical assistance of one person	2
	Continuous physical assistance of two persons or bedfast	3

Construction of RUG-II ADL index:

The ADL index (range 3 to 10) is formed by adding the scores for the three ADL variables above:

ADL index = Toileting + Eating + Transfer

For example, a resident who needs total assistance toileting, continuous supervision eating, and the physical assistance of one person while transferring would have an ADL index of 6.

[1]Bed, chair, standing.

construction of the RUG-II ADL index is outlined in Table 3.

In addition to the RUG-II ADL index, we reviewed other existing ADL systems. Three were selected for AUTOGRP analysis: ADL index (Katz, 1963), Maryland long-term care reimbursement index (Deane and Cella, 1981), and the Hebrew Rehabilitation Center for the Aged (HRCA) functional assessment tool (Hebrew Rehabilitation Center for the Aged, 1987). These systems produced variance reductions ranging from 29 percent to 31 percent, less than using some ADL's singularly. In the case of the Katz and Maryland scales, the use of dependent and independent categories, without intermediate levels, resulted in most of the residents being classified in the dependent category and in little variance being explained. The HRCA functional assessment tool used additional factors relating to mental and behavioral problems, and diagnoses. This construct, using these additional factors, had a variance explanation below that achieved employing only the ADL factors.

Using the ADL's and the RUG-II ADL index, and following a myopic policy of choosing at each step that variable with the highest variance reduction, led to classification systems with a total variance explanation of, at most, 40 percent. Most importantly, the resulting system was relatively clinically barren in that it was based totally on ADL characteristics. (When ADL's were used following this strategy, no non-ADL variable added significant variance explanation.) Interestingly, alternative approaches would prove superior.

Hierarchy of resident types

Simultaneous to these ADL development efforts, we labored with the technical panel to formulate a typology of nursing home residents. Two facts spurred our efforts. First, we believed that a case-mix system based strictly on ADL factors would not meet widespread clinical acceptance. Second, we observed in our analysis that the ADL measures alone were not capable of measuring the amounts of care by different types of staff needed by residents. We sought to create a typology that could be clinically meaningful and would facilitate communication and analysis of data among clinicians, policymakers, and health care administrators.

These major categories of residents were designed to capture broadly the essence of the types of residents cared for in SNF's and ICF's. The principal factor used to define these categories was clinical affinity. By this, we mean the residents in a category were to have much in common, although amount or severity of care were not used in the definition. Thus, we did not attempt clinically homogeneous groups at this level, but rather in the fuller system.

The major categories we developed were:

Heavy rehabilitation—Residents who receive heavy, daily physical or occupational therapy services with a treatment goal of restoration of functioning.

Special care—Very heavy-care residents, often with very low functional level, who have particular serious conditions such as coma or quadriplegia which primarily determine their care.

Clinically complex—Residents with particular medical or skilled nursing problems (e.g., hemiplegia, dehydration, terminally ill) or otherwise extensive medical needs. As a group, these residents are not as functionally impaired in ADL as the special care group.

Severe behavioral problems—Residents with high frequency and severe level of one or more of four behavioral problems: physical aggression, regression, verbal abuse, or hallucinations.

Reduced physical functions—This group includes all residents who do not qualify for any of the above four hierarchy groups; they are principally characterized by reduced levels of activities of daily living functioning.

The criteria that place residents into any of these groups are shown in Table 4. It should be noted that in addition to the classification criteria noted in Table 4, additional criteria are necessary to employ the system for payment purposes. Such additional criteria are designed to prevent gaming and to specify the documentation necessary to facilitate auditing.

This system of major categories is termed a "hierarchy" as each category was ranked by cost. Thus, residents could qualify for more than one hierarchic category but would be identified in the most resource-intensive one. Residents who had no characteristics to place them in the four highest groups were those in nursing homes mainly for reduced physical functions, the fifth, residual, and largest group.

Within each hierarchic category, more detailed subcategories were developed based on the results of cluster analysis and clinical insight. Almost uniformly, the ADL index provided subgroups with the best explanation of variance.

The principal motivation for the development of the hierarchy was to foster clinical richness and usefulness of the system across possible applications. Interestingly, the use of the hierarchy, followed by the ADL index, explained 53 percent of the variation in cost, compared with approximately 40 percent achieved when we followed a pattern of using sequentially the variables that provided the greatest variance reduction.

The increase in variance explanation was not a result of a high degree of variance explained by the factors used in the hierarchy. In fact, none of the hierarchic factors were among the top 25 factors in variance explanation. The entire hierarchy by itself explained only 17 percent of the cost variation. Thus the statistical strength of the hierarchy lies in its role of splitting the resident population into clinically similar groups, which then allows the ADL scale to categorize more effectively the resulting hierarchic groups. This interaction is an important attribute of the system.

Table 4
Criteria[1] for classifying residents into resource utilization groups, version II groups

Hierarchic groups	Criteria
Heavy rehabilitation	Either physical or occupation therapy time in excess of an average of 30 minutes per day, 5 days a week with a treatment goal of restoration of functioning.

Special care	Any of the following conditions: .
	     Comatose
	     Nasogastric feeding
	     Parenteral feeding
	     Quadriplegia
	     Multiple sclerosis
	     Stage 4 decubiti
	     Suctioning
	 and ADL index of 5 or more

Clinically complex	Any of the following treatments:
	     Oxygen therapy
	     Wound/lesion care
	     Chemotherapy
	     Transfusion
	 or any of the following diagnoses:
	     Cerebral palsy
	     Urinary tract infection
	     Hemiplegia
	 or any of the following conditions:
	     Dehydration
	     Internal bleeding
	     Terminally ill
	     Stasis ulcer
	 or one or more physician visits per week
	 or residents in the special care category with ADL index of 3 or 4

Severe behavioral problem	Any of the following problems at the severe level:
	     Physical aggression
	     Verbal abuse
	     Regressive behavior
	     Hallucinations

Reduced physical functions	Those not classified in any of the above.

[1]In addition to these classification criteria, additional criteria are used in the New York State Medicaid payment system to prevent gaming, provide guidelines for documentation, and assist in auditing.

Resource utilization groups classification system

The RUG-II classification system integrates the two basic approaches described previously. Groups are determined by first locating residents within the clinical hierarchy, then placing them into individual RUG's within that category by using ranges of the ADL index. In total, the system consists of 16 mutually exclusive groups, shown in Figure 1. The first two RUG-II groups consist of residents in the heavy rehabilitation category. The first RUG-II group, denoted "RA" (heavy rehabilitation-A), contains all residents who receive significant rehabilitation services (with a restorative goal) and a RUG-II ADL index value of either 3 or 4. A second group, denoted "RB" (heavy rehabilitation-B),

consists of the remainder of the residents in the heavy rehabilitation category, those with ADL index scores from 5 to 10. The other groups are similarly specified. The 16 RUG-II groups and their abbreviations are shown in Table 5. A resident falls into one and only one group.

It is important to understand that the RUG's system employs only a limited number of characteristics to classify residents. Although clinical description of a resident would require a fuller appraisal, with the correlations inherent between characteristics, the presence of a single attribute of a resident is an "indicator" of a spectrum of others. For example, if residents are independent in the three ADL's involved in the index, they are unlikely to require significant amounts of extra resources even if they are not totally independent in the remaining ADL's such as bathing, mobility, personal hygiene, are mentally confused in the nursing home environment, etc. Thus, this type of system provides a parsimonious approach to resident classification. In total, 12 individual items need to be assessed to classify residents into RUG-II groups.

In New York State nursing homes, the vast majority of residents, almost 70 percent, fall into the hierarchic category of reduced physical functions (Figure 2). Here, two RUG-II groups (PA and PC) each represent more than 25 percent of the residents; the former residents are most often seen in ICF's, the latter most frequently in SNF's.

The RUG-II ADL index had high explanatory power for the reduced physical functions group. This high explanatory power in the most populous set of groups confirms and explains the phenomenon of high variance explanation of ADL's for all nursing home residents seen in all other studies. The other 11 RUG-II groups provide the special characteristics that identify much smaller percentages of residents, but ones that need to be identified within a payment system for their extremely high costs of care.

Case-mix index

The case-mix index (CMI) for a group in a classification system is the relative weight of resource use of this group to some base resource use level. The base level for scaling the resource use of the RUG-II groups is arbitrarily set at the average resource use across all groups. Thus, the case-mix index for a group represented the relative cost of caring for the average resident in that group, compared with the average resident in the population.

We chose to develop a single case-mix index representing the relative cost of nursing plus other professional services (physical, occupational and recreational therapy, and social service) for each of the RUG-II groups. The analysis adjusted the costs for other professional services so that they were computed on the same basis as nursing services, including the time of assistants and aides working with these professionals and the time each professional spent in activities other than direct care
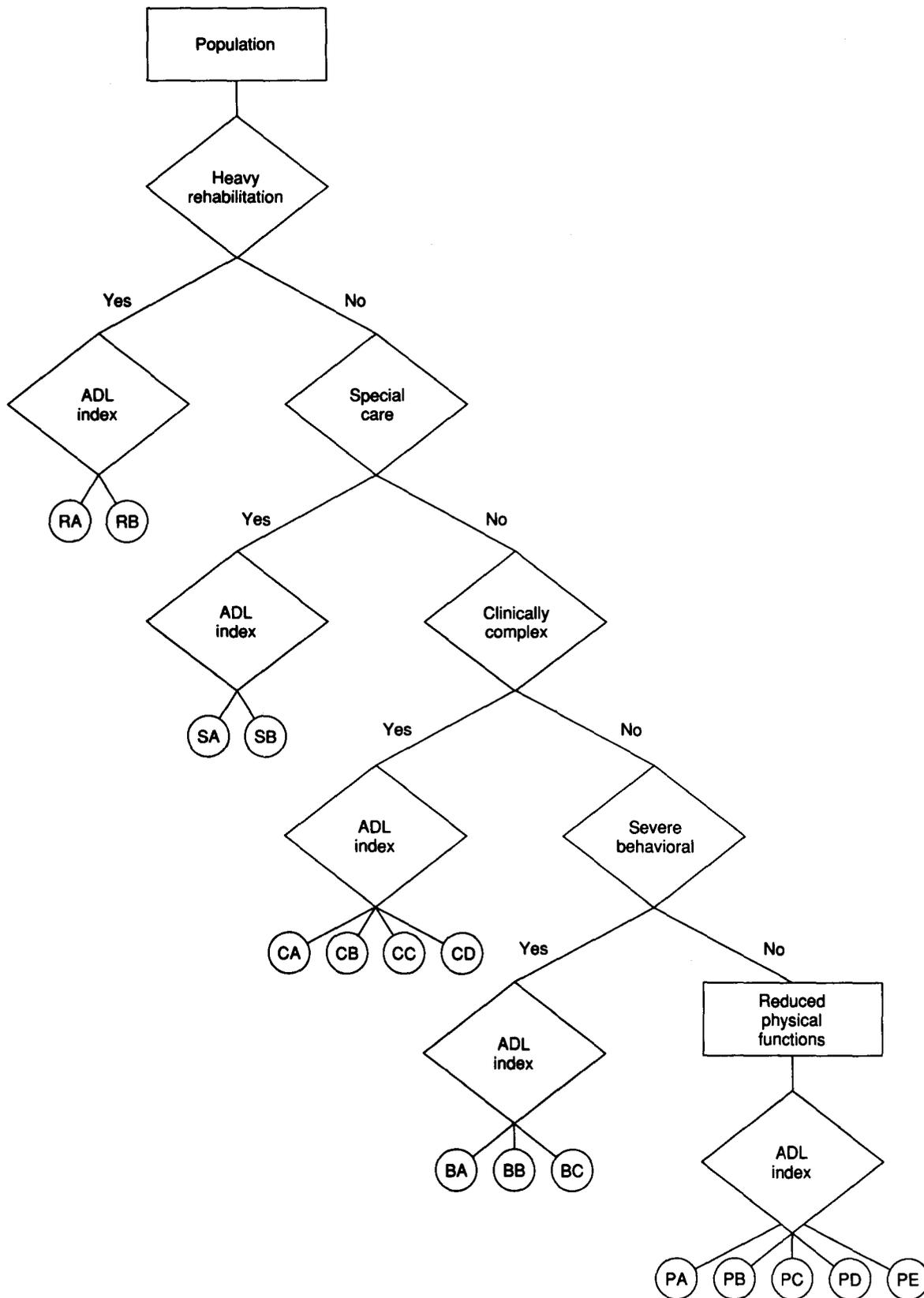
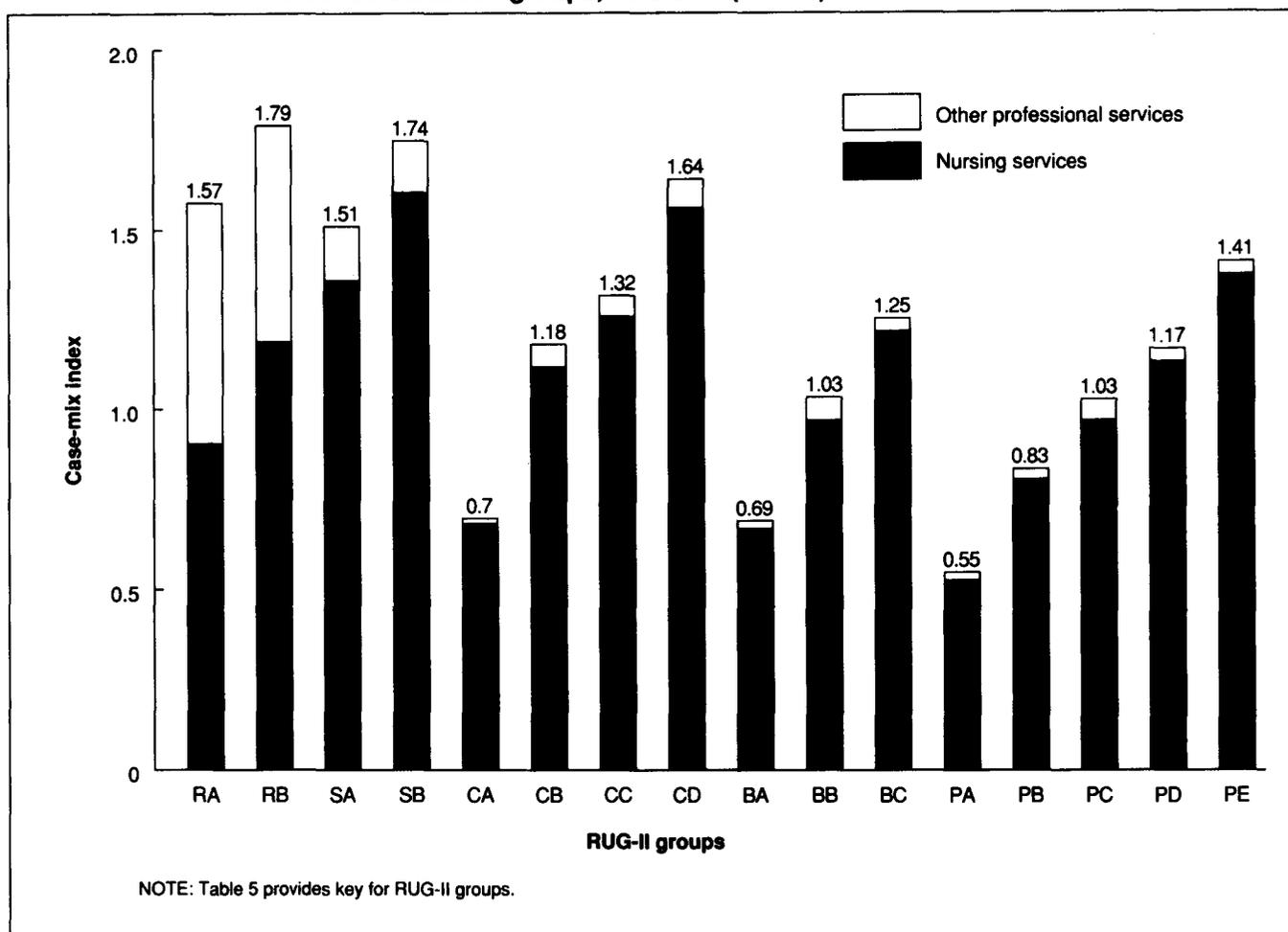NOTE: Table 4 provides classification criteria, and Table 5 provides the key for RUG-II groups.

Table 5

Resource utilization groups, version II resident categories, by code, hierarchic group, and activities of daily living (ADL) index, with coefficient of variation

Category	Code	Hierarchic group	ADL index	Coefficient of variation
Heavy rehabilitation-A	RA	Heavy rehabilitation	3–4	.25
Heavy rehabilitation-B	RB	Heavy rehabilitation	5–10	.31
Special care-A	SA	Special care	5–7	.33
Special care-B	SB	Special care	8–10	.30
Clinically complex-A	CA	Clinically complex	3	.48
Clinically complex-B	CB	Clinically complex	4–6	.29
Clinically complex-C	CC	Clinically complex	7–8	.29
Clinically complex-D	CD	Clinically complex	9	.24
Severe behavioral-A	BA	Severe behavioral	3	.49
Severe behavioral-B	BB	Severe behavioral	4–7	.27
Severe behavioral-C	BC	Severe behavioral	8–9	.23
Reduced physical-A	PA	Reduced physical functions	3	.44
Reduced physical-B	PB	Reduced physical functions	4	.34
Reduced physical-C	PC	Reduced physical functions	5–7	.27
Reduced physical-D	PD	Reduced physical functions	8	.24
Reduced physical-E	PE	Reduced physical functions	9	.21

Figure 2

Distribution of New York State patients in the resource utilization groups, version II (RUG-II) system: April 1987

NOTE: Table 5 provides key for RUG-II groups.
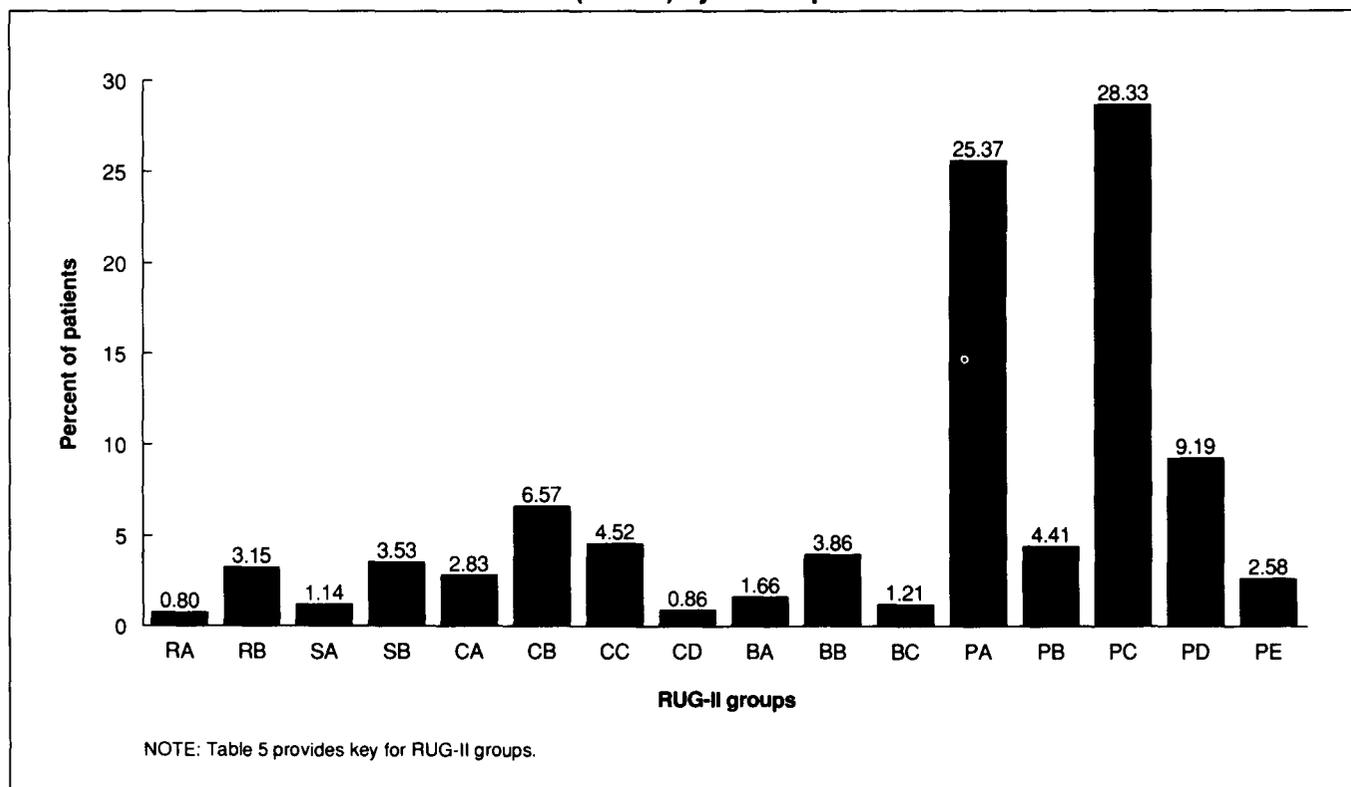
Figure 3
Resource utilization groups, version II (RUG-II) case-mix indexes

NOTE: Table 5 provides key for RUG-II groups.

of the resident, such as for administrative duties, training of staff, etc.

The case-mix indexes for the RUG-II classifications are presented in Figure 3. The order of groups in this figure follows that of Figure 1. The height of each bar represents the composite CMI, and the components represent, separately, nursing and other professional services. The majority of other professional services are seen, appropriately and by construction, for residents in the two rehabilitation groups (RA and RB). However, there are therapies provided to residents in each of the other groups as well, so payment for these groups would include reimbursement for rehabilitation care.

The highest-cost group in the RUG-II system (RB) requires 79 percent more resources than the average-cost resident, the lowest-cost group (PA) only 55 percent of the average. Thus, we see more than a 3-to-1 range of costs for residents that under many payment systems are considered equivalent. Each of the five bundles of bars represents a level of the hierarchy and the downward trend of relative resource cost as we descend the hierarchy. Overall, the observable "saw-tooth" pattern represents the increasing cost within hierarchic categories associated with increased ADL dependencies. For the same ADL

index value, the case-mix index is lower as the hierarchy is descended.[2]

Understanding explanation of variance

The variability in total cost explained by the RUG-II system was 53 percent. Although this exceeds the results reported for any other long-term care classification system and compares favorably with acute care systems such as the DRG's, it remains a question: How good is this? Alternately stated, what causes the remaining unexplained variance? To investigate if factors beyond resident-specific information explained additional variance, we incorporated facility type (SNF, ICF), sponsorship (voluntary, proprietary, public), region (seven geographic regions in New York State), and size of facility (Schlenker and Shaughnessy, 1984; Schlenker, Shaughnessy, and Yslas, 1983; Shaughnessy et al., 1985; Sulvetta and Holahan, 1986; Liu and Mossey, 1980; Bishop, 1980; Caswell and Cleverley, 1983; Meiners, 1982). Each of these factors increased our

---

[2]The one exception to this is residents with ADL index 9 in the BC group who had a higher case-mix index than those in the PE group, a rare exception and likely an anomaly.

understanding of cost variations, with type and sponsorship the most dominant. Together, these factors increased the variance explanation to approximately 65 percent. No other factors were found to increase it further.

To understand the remaining one-third of variance, it should be noted that the variance explanation achieved by any classification system is clearly much lower than 100 percent. The statistic quoted in this research is a computation at the resident-specific level. The additional variations that are not explainable at the classification level include:

- By resident: Each resident has many characteristics beyond those used in the RUG-II system, clinical differences that can and should be recognized in their appropriate care. Thus, there is variability that is unique to each resident.
- By day: The staff time measurements were performed on a single day. For the same resident, the amount of care varies from day to day. For example, if residents get a bath every 3 days, one-third of the residents will use significant time getting a bath, two-thirds will not.
- By facility: Different nursing homes care for the same resident with differing levels and types of staffing.

When accumulated to the level of a unit or beyond that to a facility of any size, the variance explanation of total resource use increases to close to 100 percent.

Discussion

We believe that our overall approach led to a system that is superior both in its capability to differentiate types of residents and in its applicability to nursing home payment.

Although developed in New York, the robustness and versatility of the RUG-II system has been demonstrated in other States and in care settings outside of nursing homes. In the State of Texas, a similar nursing home study validated both the system and the case-mix indexes (Coleman, 1988). Similar studies have been performed in the Veterans' Administration (Schaeffer, 1986), which is currently using the RUG-II system nationwide for resource allocation.

The long-term care systems of Texas and New York may be representative of the two extremes of a scale. Texas has one of the lower staffing levels per resident and New York has one of the highest. Despite this wide difference, the RUG-II classification was discriminatory between residents with high resource needs and use and those with low resource needs and use in both States. The case-mix indices (a relative-cost scale) computed from data in each State were approximately equal for each category even though the number of staff hours represented by the index was significantly different.

The robustness of the RUG-II classification has also been demonstrated in two other studies we have performed, which apply the system both to a national sample of Medicare SNF residents (Fries, Schneider, and Foley, 1987) and to New York State home care residents (Foley et al., 1986).

These observations reinforce the view that resident classification is primarily a scientific issue, based on the intrinsic relationships of resident characteristics to needs for care that transcend differences in reimbursement and care patterns.

The RUG-II system is currently used in setting the Medicaid payment rate for all nursing homes in New York State. Although a discussion of the use of the RUG-II measurements in the payment system is beyond the scope of this article, it has many other uses. Two are discussed briefly in the following sections.

Monitoring quality of care

A case-mix measurement system provides a powerful data system that can be used to assist in monitoring the quality of care in a nursing home. Resident assessment or billing systems contain data that can be used to track automatically outcomes for the residents. We have found that the RUG-II system and the individual items that define it provide useful outcome measures of quality of care. A facility, while providing documentation for payment purposes, at the same time provides data usable by a quality assurance system to monitor outcome performance on both a resident-specific and an aggregate level. Such dual use puts nursing homes at risk if they over-report case-mix criteria (at the risk of indicating inappropriate care) or they under-report quality problems (at the risk of losing reimbursement). The development of a monitoring system can include many types of review: aggregate case-mix index changes at the facility level; longitudinal outcomes at the case-mix group level; longitudinal outcomes at the resident-specific level; and incidence of poor outcomes.

Case-mix measurement can thus address the basic objectives of a quality assurance system to find poor quality care and can make this process more efficient through computerized automation. Specifically, a case-mix data base can be used to generate by computer a variety of reports employable within a quality assurance system, including incidence of specific outcomes (e.g., decubitus ulcers, tube feeding); comparisons of a facility with others, or a longitudinal view of a single facility across time; lists of targeted residents to be reviewed during the on-site quality assurance review; or menu of activities to focus reviewers on potential problem areas in the facility.

The type of system described above is currently being developed at Rensselaer Polytechnic Institute and the New York State Department of Health under a HCFA grant. The resulting system, termed the New York Quality Assurance System (NYQAS), was implemented in Fall 1988.

Nursing home management

A case-mix measurement system enables new management opportunities for nursing homes. In States with case-mix payment systems, nursing home administrators can now change both their costs and their revenues within the structure of the system and can relate costs to revenues. Case mix creates these opportunities, but it also engenders new needs in the provider community. One of these needs is for management information that can be used to predict revenues and costs under the case-mix system. Another need is for increased management skills to use the information available and to recognize the opportunities and pitfalls that a case-mix system presents to a facility.

Even without case-mix-based payment, a case-mix measurement system creates increased awareness by facility management of the types of residents under their care. Management can now relate their facility's capacity to provide service (such as rehabilitation therapies) to the cost for this capacity and the revenues that this capacity can produce. Case mix provides a better match between the revenues and the costs of caring for a given resident mix. This match provides an opportunity for management to increase the quality of care in a nursing home by using revenues to augment the number and skill-mix of staff giving care.

Conclusions

The development of the RUG-II system was guided by clinical, statistical, and administrative criteria. An important conclusion of the research was that synergy existed between these criteria. The addition of clinical insight significantly improves the variance explanation of the resulting classification system beyond that developed through statistical methods alone.

The analysis supports the finding of previous research that ADL's are major explanatory factors in resource use. However, a classification based on ADL's would clinically be relatively barren and thus would likely not find clinical acceptance. The RUG-II system utilizes a clinical hierarchy of five types of residents (heavy rehabilitation, special care, clinically complex, severe behavioral problems, and reduced physical functions). A second component is a summarization of the ADL's into an index based on toileting, eating, and transfer. Together, these produce a system of 16 relatively homogeneous groups of nursing home residents.

The RUG-II system was developed for and is currently being used to pay all New York State nursing homes under Medicaid, as well as being employed by the Veterans' Administration.

The development of methods to measure and predict the resources used for caring for different types of nursing home residents provides the metric needed to address a variety of opportunities at both the governmental and provider level. Some caution is needed as well. Case-mix measurement has been promoted by some as solving the plethora of problems identifiable in long-term care. It provides solutions to perhaps only a few, but in doing so, makes needed progress in a complicated arena.

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.

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.

Bishop, C. E.: Nursing home cost studies and reimbursement issues. *Health Care Financing Review*. Vol. 1, No. 4. HCFA Pub. No. 03045. Office of Research, Demonstrations, and Statistics, Health Care Financing Administration. Washington. U.S. Government Printing Office, Spring 1980.

Bishop, C. E., Plough, A. L., and Willemain, T. R.: Nursing home levels of care: Problems and alternatives. *Health Care Financing Review*. Vol. 2, No. 2. HCFA Pub. No. 03068. Office of Research, Demonstrations, and Statistics, Health Care Financing Administration. Washington. U.S. Government Printing Office, Fall 1980.

Cameron, J. M.: Case-mix and resource use in long-term care. *Medical Care* 23(4):296-309, Apr. 1985.

Caswell, R. J., and Cleverley, W. O.: Cost analysis of the Ohio nursing home industry. *Health Services Research* 18(3):359-391, Fall 1983.

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.

Coleman, P.: Director, Texas Long Term Care Case Mix Reimbursement Project, Texas Department of Human Resources. Personal communication. Austin, Tex. 1988.

Cretin, S., and Worthman, L. G.: *Alternative Systems for Case Mix Classification in Health Care Financing*. Cooperative Agreement No. 99-C-98489/9-03. Prepared for Health Care Financing Administration. Santa Monica, Calif., Rand Corporation, Sept. 1986.

Deane, R. T., and Cella, M. A.: New Concepts in Nursing Home Reimbursement. Working paper. Applied Management Sciences. Silver Spring, Md. 1981.

Fetter, R. B., Shin, Y., Freeman, J. L., et al.: Case mix definition by diagnosis-related groups. *Medical Care* 18(2)Supp.:1-53, Feb. 1980.

Foley, W. J., Schneider, D., Dowling, M., et al.: *Development of a Survey, Case Mix Measurement System, and Assessment Instrument to Rationalize the Long Term Care Home Care System.* Contract No. 12,000 C 001558. Final report prepared for Office of Health Systems Management, New York State Department of Health. Troy, N.Y., Rensselaer Polytechnic Institute, 1986.

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. J., et al.: *A Case Mix Measure for Medicare Skilled Nursing Facility Patients.* Cooperative Agreement No. 18-C-98581/2-01 with Rensselaer Polytechnic Institute. Final report prepared for Health Care Financing Administration. Ann Arbor, Mich., The University of Michigan, 1987.

Hebrew Rehabilitation Center for the Aged: The HRCA Functional Assessment Tool. Unpublished working paper. Boston. Jan. 1987.

Katz, S., 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* 195(2):914-919, Sept. 21, 1963.

Liu, K., and Mossey, J.: The role of payment source in differentiating nursing home residents, services, and payments. *Health Care Financing Review.* Vol. 2, No. 1. HCFA Pub. No. 03054. Office of Research, Demonstrations, and Statistics, Health Care Financing Administration. Washington. U.S. Government Printing Office, Summer, 1980.

McCaffree, K. M., Winn, S., and Bennett, C. A.: *Final Report of Cost Data Reporting System for Nursing Home Care.* Seattle, Wash., Battelle Human Affairs Research Centers, 1976.

McCaffree, K. M., Baker, J., and Perrin, E. B.: *Long Term Care Case Mix Employee Time and Costs.* Seattle, Wash., Battelle Human Affairs Research Centers, 1979.

Meiners, M. R.: An econometric analysis of the major determinants of nursing home costs in the United States. *Social Science and Medicine* 16:887-898, 1982.

Mills, R., Fetter, R. B., Riedel, D. C., Averill, R., et al.: AUTOGRP: An interactive computer system for the analysis of health care data. *Medical Care* 14(7):603-615, July 1976.

Minnesota Department of Health: Facility Manual for Completing Case Mix Requests for Classification. Unpublished manual. Minneapolis. May 1986.

Morgan, J. N., and Sonquist, J. A.: Problems in the analysis of survey data and a proposal. *Journal of the American Statistical Association* 58(302):415-434, June 1963.

Morris, J. N., Sherwood, S., May, M. I., et al.: FRED: An Innovative Approach to Nursing Home Level-of-Care Assignments. Working paper. Hebrew Rehabilitation Center for Aged. Boston. 1986.

O'Donnell, J. F., and Hannan, E.: An analysis of case-intensity-adjusted nursing staff patterns in New York State skilled nursing facilities. *Medical Care* 21 (1):82-91, Jan. 1983.

Schaeffer, P.: Veterans' Administration Central Office. Personal communication. Washington, D.C. 1986.

Schlenker, R. E., and Shaughnessy, P. W.: Case mix, quality, and cost relationships in Colorado nursing homes. *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.

Schlenker, R., Shaughnessy, P., and Yslas, I.: The effect of case mix and quality on cost differences between hospital-based and freestanding nursing homes. *Inquiry* 20(4): 361-368, Winter 1983.

Schneider, D., Fries, B. E., Desmond, W., 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., Mar. 1985.

Schneider, D., Desmond, M., Anderman, S., et al.: *Resource Utilization Group—II: Case Mix Reimbursement System for Long Term Care.* Contract No. 11-C-98325/2-03 with New York State Department of Health. Draft final report prepared for Health Care Financing Administration. Troy, N.Y., Rensselaer Polytechnic Institute, 1987.

Shaughnessy, P. W., Kramer, A. M., Schlenker, R. E., and Polesovsky, M. B.: Nursing home case-mix differences between Medicare and non-Medicare and between hospital-based and freestanding patients. *Inquiry* 22(2):162-177, Summer 1985.

Stassen, M., and Bishop, C. E.: *Incorporating Case Mix in Prospective Reimbursement for Skilled Nursing Facility Care Under Medicare: Critical Review of Relevant Research.* Contract No. 18-P-97038/1-05. Prepared for Health Care Financing Administration. Waltham, Mass., Center for Health Policy Analysis and Research, Brandeis University, Mar. 1983.

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.

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.